Simulationen und Randomisierungstests mit der Software TinkerPlots

-

Theoretische Werkzeuganalyse zur stochastischen Simulation und explorative Fallstudie zum statistischen Schließen mit Randomisierungstests

Dissertation

Zur Erlangung des Grades eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

Susanne Podworny

an der Fakultät für Elektrotechnik, Informatik und Mathematik
der Universität Paderborn

2018

Danksagung

Das Verfassen dieser Dissertation wäre ohne die Unterstützung vieler Personen nicht möglich gewesen. An dieser Stelle bedanke ich mich ganz herzlich bei allen, die mich in diesem Prozess begleitet und auf vielfältige Weise unterstützt haben.

Mein erster Dank gilt meinem Doktorvater Prof. Dr. Rolf Biehler, durch dessen intensive Anregungen, Diskussionen und Rückmeldungen die Arbeit zu dem geworden ist, was sie jetzt ist. Ich bedanke mich auch bei ihm für die gute Zusammenarbeit in Projekten, im Verfassen diverser Publikationen und für das Ermöglichen der Teilnahme an vielen internationalen Tagungen. All dies hat meinen Horizont in den vergangenen zehn Jahren wesentlich erweitert, wofür ich mich herzlich bedanke.

Prof. Dr. Joachim Engel danke ich für die gute Zusammenarbeit und Koordination im Projekt „ProCivicStat“, durch das ich viele interessante Diskussionen führen konnte und durch das ich viele europäische Orte besuchen durfte. Darüber hinaus bedanke ich mich bei ihm für die Übernahme des Zweitgutachtens und für die Bereitschaft, in der Promotionskommission mitzuwirken.

Der Arbeitsgruppe Biehler und der Fachgruppe Didaktik der Mathematik danke ich für die wunderbare kollegiale Atmosphäre.

Bei meinem Bürokollegen Dr. Daniel Frischemeier bedanke ich mich für unzählbare fachliche Diskussionen, Anregungen und Hilfen, aber auch für die hervorragende Zusammenarbeit in den letzten Jahren. Ohne diese Zusammenarbeit wären viele Publikationen, Vorträge und Lehrveranstaltungen nicht in dieser Form entstanden. Ich blicke zurück auf viele gute und fruchtbare Jahre mit ihm und hoffe, dass sich diese Zusammenarbeit auch in der Zukunft hält. Auf dass noch viele Taxifahrer kreisen werden…!

Ich bedanke mich bei Marie-Christine Scherer für das Erstellen der aufwändigen Transkripte und die fachlichen Diskussionen dazu und bei Jana Roske-Voß und Katrin Stolle für unzählige Stunden Korrekturlesens.

Meinen Kindern Benjamin, Jonathan und Theobald danke ich für all die Geduld mit „Mamas Arbeit“, ihr seid der Sonnenschein meines Lebens und gebt mir Grund, jeden Tag weiterzumachen.

Mein allergrößter Dank gilt meinem Ehemann Stefan Podworny. Ohne Dich, Deine Liebe und Unterstützung wäre diese Arbeit niemals fertig geworden und ich wäre nicht der Mensch, der ich heute bin. Ich liebe Dich.

Susanne

Zusammenfassung

Ausgangspunkt dieser Arbeit ist die weit verbreitete Ansicht, dass Lernprozesse durch computergestützte Simulationen profitieren. Einen einfachen Zugang zum Modellieren und Durchführen stochastischer Simulationen bietet die Software TinkerPlots (Konold & Miller 2017), die national und international eingesetzt wird. Aus diesem Grund wird im ersten Teil dieser Arbeit das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation untersucht, um Möglichkeiten und Grenzen der Software aufzuzeigen und didaktisch zu beleuchten.

Aufbauend auf dieser Werkzeuganalyse wurde die Software in Lehrveranstaltungen der Universität Paderborn zur Datenanalyse und zur Simulation breit eingesetzt, in den letzten Jahren vor allem in der mit dem Lehrpreis für den wissenschaftlichen Nachwuchs ausgezeichneten Veranstaltung „Modellieren, Größen, Daten und Zufall I“ für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung. Für diese Lehrveranstaltung wurde für das Sommersemester 2017 im Rahmen der vorliegenden Dissertation eine Lerneinheit zum statistischen Schließen mit Randomisierungstests nach dem Design-Based Research Ansatz entwickelt, um die zukünftigen Lehrer in die bis dahin weitgehend unbekannte Logik des statistischen Schließens einzuführen. Diese Lerneinheit wird im zweiten Teil der Arbeit vorgestellt zusammen mit einer dazu durchgeführten explorativen Fallstudie.

Die explorative Studie wurde im Anschluss an die implementierte Lerneinheit mit sechs Studierenden durchgeführt, um einen detailreichen Einblick in die kollaborativen Bearbeitungsprozesse Studierender zu bekommen, die einen Randomisierungstest mit TinkerPlots durchführen. Die Auswertung dieser Studie wird ausführlich ebenfalls im zweiten Teil der vorliegenden Arbeit dargelegt. Hieraus werden Faktoren für gelingende kollaborative Bearbeitungsprozesse abgeleitet und die genaue Nutzungsweise der Software TinkerPlots in diesen Prozessen untersucht. Aus diesen Erkenntnissen zu Vorstellungen und Schwierigkeiten beim Durchführen von Randomisierungstests mit TinkerPlots werden Hinweise für die Weiterentwicklung der Lerneinheit abgeleitet und allgemeine Empfehlungen für die Behandlung dieses Themas generiert.

Somit leistet die vorliegende Arbeit einen Beitrag zum Stand der mathematikdidaktischen Forschung erstens durch die Werkzeuganalyse der Software TinkerPlots und zweitens durch die Untersuchung einer Einführung in die Logik der Inferenzstatistik mit Randomisierungstests unter Verwendung der Software TinkerPlots.

Abstract

This work starts with the widespread view that learning processes benefit from computer-supported simulations. The software TinkerPlots (Konold & Miller 2017), which is used nationally and internationally, provides easy access to modeling and to performing stochastic simulations. For this reason, in order to show the possibilities and limitations of the software, the first part of this thesis examines how TinkerPlots supports stochastic simulations.

Based on this tool analysis, the software has been widely used in recent years in courses at the Paderborn University for data analysis and simulation, especially in the course „Modellieren, Größen, Daten und Zufall I“ (Modeling, Magnitudes, Data and Chance I). This is a course in mathematics education for primary school preservice teachers concentrating on basic data analysis, combinatorics, and probability. A learning trajectory on statistical inference with randomization tests was developed for this course using a design-based research approach. This short sequence of lessons – a completely new topic for most participants – was incorporated at the end of the course. The design of the learning trajectory will be presented in the second part of the thesis together with an exploratory case study.

After they experienced the new lessons, six of the preservice teachers participated in an exploratory case study in which they conducted, in pairs, a randomization test using TinkerPlots. The evaluation of this study is also explained in detail in the second part of this thesis. From this, factors for successful collaborative working processes are derived and the exact usage of the software TinkerPlots in these processes is examined. Based on findings about ideas and difficulties in carrying out randomization tests with TinkerPlots, hints for the further development of the learning trajectory are developed and general recommendations for dealing with this topic are generated.

Thus, the present work contributes to the state of mathematics education research, firstly through the analysis of the TinkerPlots software, and secondly by examining an introduction to the logic of inference with randomization tests using TinkerPlots.

Inhaltsverzeichnis

Danksagung ..... **III**

Zusammenfassung ..... **V**

Abstract ..... **VI**

Inhaltsverzeichnis ..... **VII**

1 Einleitung **1**

1.1 Forschungszusammenhang 1

1.2 Ziele der Arbeit 4

1.3 Aufbau der Arbeit 5

TEIL I: THEORETISCHE WERKZEUGANALYSE ZUR STOCHASTISCHEN SIMULATION MIT TINKERPLOTS **9**

2 Hintergründe der theoretischen Werkzeuganalyse **11**

2.1 Nutzen von Simulationen 11

2.2 Theorie der Mensch-Werkzeug Interaktion 14

2.2.1 Der Ansatz der instrumentellen Genese 15

2.2.1.1 Vom Artefakt zum Instrument 17

2.2.1.2 Instrumentation und Instrumentalization als konstituierende Prozesse 19

2.2.1.3 Zusammenfassung 22

2.2.2 Didaktik der instrumentellen Genese: Instrumentelle Orchestrierung 23

2.2.2.1 Konkrete Gestaltungsvorschläge für Unterricht (exploitation mode) 27

2.2.2.2 Die instrumentelle Genese des Lehrenden 30

2.3 Ein Modell des Wissenserwerbs: Cognitive Load Theory 32

2.4 Der Simulationsplan als didaktisches Konzept 33

3 Theoretische Werkzeuganalyse der Software TinkerPlots **36**

3.1 Ziele und Methoden der Werkzeuganalyse 36

3.2 Kurzüberblick über die Software TinkerPlots 37

3.2.1 Software-Ergonomie ... 37
3.2.2 Kurzüberblick zur Datenanalyse mit TinkerPlots ... 42
3.2.3 Kurzüberblick zur Simulation mit TinkerPlots ... 43
3.3 TinkerPlots als kognitives Werkzeug/expressives Medium ... 45
3.4 Strukturierung der Software: Bereiche zur Simulation ... 49
3.5 Interface Modellieren: (Pseudo-)Zufall erzeugen in TinkerPlots ... 52
3.5.1 Die Bauteile der Zufallsmaschine im Detail ... 53
3.5.2 Verschiedene Simulationsarten ... 85
3.5.2.1 Simultane Simulationsarten: Simultan-gebündelte und simultan-hintereinander-geschaltete Simulation ... 87
3.5.2.2 Sequentielle Simulation ... 93
3.5.2.3 Simulation als Stichprobenziehung ... 99
3.5.2.4 Simulation durch Randomisierung ... 102
3.6 Interface Ereignisse und Zufallsgrößen ... 105
3.7 Simulieren mit Abbruchbedingungen ... 122
3.8 Interface Auswerten ... 128
3.9 Weitere Tools ... 130
4 Detailanalyse des Simulationspotentials von TinkerPlots ... 132
4.1 Klassifizierung stochastischer Problemstellungen ... 133
4.1.1 Einstufige Zufallsexperimente ... 133
4.1.2 Mehrstufige Zufallsexperimente ... 134
4.1.3 Wartezeitprobleme ... 135
4.2 Simulation verschiedener Zufallsexperimente ... 136
4.2.1 Simulation einstufiger Zufallsexperimente ... 136
4.2.2 Simulation mehrstufiger, stochastisch unabhängiger Zufallsexperimente ... 139
4.2.3 Simulation mehrstufiger, stochastisch abhängiger Zufalls-experimente ... 152
4.2.4 Simulation von Wartezeitproblemen ... 165

4.3 Ergänzungen ... 171
4.3.1 Ergänzung: Gesetz der großen Zahlen ... 172
4.3.2 Ergänzung: Die versteckte Zufallsmaschine ... 175
4.4 Zusammenfassung ... 178
4.5 Fazit ... 181

TEIL II: STATISTISCHES SCHLIEßEN MIT RANDOMISIERUNGSTESTS: LERNEINHEIT UND EXPLORATIVE FALLSTUDIE ZUR DURCHFÜHRUNG VON RANDOMISIERUNGSTESTS MIT TINKERPLOTS ... 183

5 Grundlagen zu Teil II: Inferenzstatistik und Randomisierungstests ... 185
5.1 Statistical Literacy, Statistical Reasoning und Statistical Thinking ... 186
5.2 Informelle Inferenzstatistik ... 188
5.3 Resamplingverfahren als Einstieg in Inferenzstatistik ... 190
5.4 Der Randomisierungstest als ein Resamplingverfahren ... 191
5.4.1 Experimental Research ... 199
5.4.2 Mit welchen Stichprobengrößen kann im Randomisierungstest gearbeitet werden? ... 201
5.4.3 Was lässt sich aus den Ergebnissen eines Randomisierungstests schließen? ... 202
5.4.4 Kritik am Randomisieren ... 202

6 Randomisierungstests in der didaktischen Literatur ... 204
6.1 Lernerfolge, Schwierigkeiten und Unterrichtshinweise bei Randomisierungstests ... 205
6.2 Verschiedene Schemata zu Randomisierungstests ... 236
6.3 Die „Kernlogik“ des Randomisierungstests ... 249
6.4 Konzepte für die Durchführung eines Randomisierungstests ... 255
6.5 Verwendete Software bei Randomisierungstests ... 261
6.6 Zusammenfassung ... 264

7 Forschungsfragen und Methodik für Teil II ... 267
7.1 Forschungsfragen ... 267

7.2 Design Based Research 268
7.3 Qualitative Forschung 269
7.3.1 Fallstudie 271
7.3.2 Erhebungsverfahren: Das Leitfadeninterview 273
7.4 Mathematisch Argumentieren beim Randomisierungstest 274
7.5 Anlage der Studie 277
8 Die Lerneinheit „Inferenzstatistik mit Randomisierungstests“ 278
8.1 Rahmenbedingung: Die Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ im Sommersemester 2017 279
8.2 Der Baustein Inferenzstatistik 285
8.2.1 Ziele des Bausteins „Inferenzstatistik“ 285
8.2.2 Die Sitzungen im Detail 287
8.2.2.1 Vorlesung am 20.07.2017 287
8.2.2.2 Seminar am 25.07.2017 295
8.2.2.3 Vorlesung am 27.07.2017 304
8.3 Reflexion zum Baustein Inferenzstatistik 307
8.4 Fazit zur Lerneinheit „Inferenzstatistik mit Randomisierungstests“ 309
9 Design der explorativen Fallstudie „Randomisierungstests mit TinkerPlots“ 312
9.1 Forschungsfragen zur Studie 313
9.2 Die Interviewstudie 313
9.2.1 Die Aufgabe der Studie 314
9.2.2 A priori Analyse der Aufgabe 317
9.3 Geplanter Ablauf der Interviewstudie 333
9.4 Konkreter Ablauf der Interviewstudie 338
9.5 Die Teilnehmer 340
9.6 Datenerhebung 341
9.7 Transkription 341

10 Methodologische Grundlagen der Auswertungen 344

10.1 Aushandlung von Bedeutung als konstruktiver, interaktionistischer Prozess 345

10.2 Kommunikationsprozesse analysieren 345

10.2.1 Struktur von Gesprächen 346

10.2.2 Interaktionsanalyse nach Krummheuer und Naujok 346

10.2.3 Funktionale Argumentationsanalyse nach Toulmin 348

10.3 Generieren der Unterforschungsfragen 349

10.4 Auswertungsmethodologie zur Fallstudie – angepasstes Vorgehen 351

10.4.1 Gliederung in Interaktionseinheiten 351

10.4.2 Interpretative Rekonstruktion des Lösungsprozesses 353

10.4.3 Rekonstruktion einzelner Argumentationen 355

10.4.4 Analyse der TinkerPlots-Benutzung 356

10.5 Gütekriterien qualitativer Forschung 366

11 Auswertung der explorativen Fallstudien „Randomisierungstest mit TinkerPlots“ 370

11.1 Überblick über den zeitlichen Bearbeitungsprozess aller Paare 371

11.2 Fallstudie Rebecca und Selina 373

11.2.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Rebecca und Selina 373

11.2.2 TinkerPlots-Benutzung zur Simulation von Rebecca und Selina 425

11.2.3 Fallzusammenfassung Rebecca und Selina 427

11.3 Fallstudie Fabia und Laura 429

11.3.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Fabia und Laura 429

11.3.2 TinkerPlots-Benutzung zur Simulation von Fabia und Laura 476

11.3.3 Fallzusammenfassung Fabia und Laura 478

11.4 Fallstudie Mandy und Alisa 480

11.4.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Mandy und Alisa 480

11.4.2 TinkerPlots-Benutzung zur Simulation von Mandy und Alisa 527

11.4.3 Fallzusammenfassung Mandy und Alisa 529

11.5 Vergleichende Analyse in Bezug auf die Unterforschungsfragen 530

11.5.1 Welche Problemstellen treten auf? Welche davon sind unerwartet? Welche Gründe lassen sich identifizieren für die unerwarteten Problemstellen? 531

11.5.2 Wie stark sind die Phasen der Gesprächsführung ausgeprägt? Wie gestalten die Paare die Kommunikation untereinander? Was bedeutet das in Bezug auf die kollaborativen Bearbeitungsprozesse? 535

11.5.3 Arbeiten die Teilnehmer der Studie kollaborativ zusammen? An welchen Merkmalen lässt sich die Kollaboration erkennen? 538

11.5.4 Können die Teilnehmer die verschiedenen, in Kapitel 6.4 identifizierten Konzepte (einen Unterschied in den beobachteten Daten identifizieren, Nullhypothese aufstellen, etc.), die zur Argumentation eines Randomisierungstests gehören, verwenden und erläutern? Welche Schwierigkeiten treten dabei auf? 539

11.5.5 In welcher Weise gelingt es den Teilnehmern, die drei Ebenen Kontext, Statistik und Software aufeinander zu beziehen? 547

11.5.6 Können die Teilnehmer die Software TinkerPlots als Instrument einsetzen, um einen Randomisierungstest durchzuführen? An welchen Stellen bei der Benutzung der Software TinkerPlots benötigen die Studierenden Hilfe? 549

11.6 Einschränkung und Ausblick 553

12 Fazit 555

12.1 Ausgangslage 555

12.2 Zusammenfassung der Ergebnisse 557

12.2.1 Zusammenfassung der Ergebnisse der Werkzeuganalyse 557

12.2.2 Zusammenfassung der Ergebnisse des zweiten Teils 559

12.2.2.1 Gestaltung einer Lerneinheit „Inferenzstatistik mit Randomisierungstests“ 561
12.2.2.2 Problemstellen bei der selbständigen Durchführung eines Randomisierungs-tests 561
12.2.2.3 Faktoren für gelingende, kollaborative Kommunikationsprozesse 564
12.2.2.4 Auffälligkeiten hinsichtlich der verschiedenen Konzepte im Hinblick auf ein mögliches Redesign der Lerneinheit 566
12.2.2.5 TinkerPlots als Instrument beim Randomisierungstesten und auftretende Schwierigkeiten 568
12.2.2.6 Hinweise für ein Redesign der Lerneinheit 570
12.3 Bedeutung der vorgelegten Studie für die Stochastikdidaktik 571
12.3.1 Bedeutung der Werkzeuganalyse 571
12.3.2 Bedeutung des Forschungsüberblicks zu Randomisierungstests 571
12.3.3 Bedeutung der explorativen Fallstudie 572
12.3.4 Empfehlungen für die Gestaltung einer Lerneinheit zur Logik der Inferenzstatistik mit Randomisierungstests 573
12.4 Ausblick 574
Literaturverzeichnis 576

1 Einleitung

1.1 Forschungszusammenhang

Statistisches Schlussfolgern ist ein Eckpfeiler, auf dem die Praxis der Statistik beruht. In nahezu allen Bereichen des täglichen Lebens spielen Daten und somit auch aus Daten gezogene Schlussfolgerungen eine wichtige Rolle. Dabei sind computergestützte Auswertungen und Methoden nicht mehr wegzudenken aus der statistischen Praxis. In der Industrie, Medizin, Politik und in vielen anderen Feldern werden Entscheidungen mehr und mehr auf Grundlage von Daten geschlossen, „Big Data“ ist hier das Modewort. Beim Blick in die Zeitung, in einen Fernsehbeitrag oder einen Eintrag im World Wide Web begegnen dem interessierten Bürger immer häufiger die Schlagworte „Eine Studie hat gezeigt …“ oder „Der Effekt von X liegt an Y“. Dabei wird jedoch häufig unterschlagen, dass die so gelieferten Ergebnisse und Interpretationen keinesfalls sicher sind, wie es in den Medien gerne suggeriert wird. Bereits 2003 formuliert der Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik dazu eine Stellungnahme:

> „Immer mehr Entscheidungen und Vorhersagen beruhen auf der Analyse statistischer Daten, die Gefahr von Fehlinterpretationen und Missbrauch von Daten nimmt zu. Der Einsatz stochastischer Modelle zum Treffen von Entscheidungen in Situationen der Ungewissheit gewinnt an Bedeutung.“ (Arbeitskreis Stochastik 2003, 21)

Aus diesem Grund sollte „statistisches Denken“, wie beispielsweise auch im grundlegenden Artikel von Wild und Pfannkuch (1999) gefordert, ein integraler Bestandteil des Stochastikunterrichts sein.

Ein zentraler Bereich des „statistischen Denkens“ ist das „schlussfolgernde Denken“, bei dem Schlüsse über vorliegende Daten hinaus gezogen werden, und darauf Entscheidungen aufgebaut werden, so wie es in den Medien gängige Praxis ist. Diesen Prozess des Schlussfolgerns zu verstehen schätzen auch Garfield und Ben-Zvi (2008a, 262) als wichtige Fähigkeit eines jeden Erwachsenen ein. „Drawing inferences from data is part of everyday life and critically reviewing results of statistical inferences from research studies is an important capability for all adults.” Um nun Schlüsse aus Daten ziehen zu können oder postulierte Ergebnisse kritisch reflektieren zu können, ist es erforderlich, die Logik zu verstehen, auf der statistische Schlussfolgerungen beruhen. Die Logik der Inferenzstatistik zu verstehen ist jedoch ein kritischer Punkt, der häufig zu (Lern-) Schwierigkeiten führt (z. B. Engel 2014a, Garfield & Ben-Zvi 2008a, Haller & Krauss 2002, Pfannkuch, Wild & Regan 2013, Thompson, Liu & Saldanha 2012).

In der Inferenzstatistik gibt es zwei große Bereiche: Parameterschätzung und Hypothesentesten, und es gibt zwei Arten von Schlussfolgerungen, die gezogen werden können. Zum einen werden Verallgemeinerungen von einer Stichprobe auf die Population gesucht und zum anderen wird danach gefragt, ob Muster in vorliegenden Daten einer be-

stimmten Behandlung oder einem bestimmten Effekt zugeschrieben werden können. Die letztere ist eine Frage, die auch im Statistikunterricht thematisiert werden sollte:

> „Statistics education should really be telling students something every scientist knows, 'The quest for causes is the most important game in town.' It should be saying: 'Here is how statistics helps you in that quest'." (Wild & Pfannkuch 1999, 238)

Als eine statistische Methode, mit der solche kausalen Schlussfolgerungen gezogen werden können, wurde von G. Cobb (2007) und Rossman (2008) für die Einführung in die Logik des statistischen Schließens das Randomisierungstesten besonders hervorgehoben. Ein Randomisierungstest ist ein nichtparametrisches Verfahren, das durch die Leistungsfähigkeit moderner Computer über statistische Simulationen einen Zugang zur Inferenzstatistik bietet, der nahezu ohne Formeln und ohne formale Berechnungen auskommt und somit die Logik des Verfahrens stärker in den Vordergrund stellen kann. Nichtparametrische Verfahren sind flexibler als herkömmliche statistische Tests und können für Anfänger intuitiver zu verstehen sein (Pfannkuch & Budgett 2014, Tintle, Topliff, Vanderstoep, Holmes & Swanson 2012). Über Randomisierungstests lassen sich Schlüsse aus Daten ziehen, selbst wenn diese aus kleine Stichproben oder von nichtzufälligen Erfassungsmethoden stammen, so wie es in empirischer Forschung häufig der Fall ist (Edgington 1995, Ernst 2004, Zieffler, Harring & Long 2011).

Bei dem Verfahren des Randomisierungstests spielen stochastische Simulationen eine elementare Rolle, insofern ist die Wahl einer geeigneten Software entscheidend für den Lernprozess (Budgett, Pfannkuch, Regan & Wild 2012, Burrill & Biehler 2011, Holcomb, Chance, Rossman & Cobb 2010, Lock, Lock, Lock Morgan, Lock & Lock 2014, Rossman & Chance 2014). Um stochastische Simulationen für Randomisierungstests durchzuführen, muss eine entsprechende Werkzeugsoftware als Instrument beherrscht werden. Damit aus einem Werkzeug ein Instrument wird (Rabardel 2002), ist es nötig, dessen Möglichkeiten und Beschränkungen im Vorfeld genau zu analysieren. Durch den Einsatz von Simulationen werden jedoch nicht nur formale Berechnungen ersetzt, sondern gerade in Bezug auf Randomisierungstests haben sie einen entscheidenden Vorteil.

> „One advantage of this procedure [simulations] for introducing introductory students to the reasoning process of statistical inference is that it makes clear the connection between the random assignment in the design of the study and the inference procedure. It also helps to emphasize the interpretation of a p-value as the longterm proportion of times that a result at least as extreme as in the actual data would have occurred by chance alone under the null model." (Rossman 2008, 10)

Neben den von Rossman erwähnten inhaltlichen Vorteilen, die der Simulationsansatz für das Verstehen des Randomisierungstests bietet, können Randomisierungstests auch motivationale Vorteile bieten, und eine passende Simulation kann und sollte durch Hands-on Aktivitäten begleitet werden und so verschiedene Lerntypen bedienen (vgl. Rossman & Chance 1999).

> „The arguments for a randomization-based curriculum go beyond merely improving student's conceptual understanding. Anecdotal evidence is mounting that students may be more engaged in the course through its integral use of tactile simulations and other hands-on active learning strategies (e.g., a permutation test or simulation of a binomial distribution is easily adapted to an in-class, hands-on activity, while a t-test may be less so)." (Swanson, Vanderstoep & Tintle 2014, 2)

Darüber hinaus hat der Zugang über Randomisierungstests einen dritten Vorteil gegenüber anderen Methoden, denn er reduziert den cognitive load (Chandler & Sweller 1991) von Lernenden.

> „Norm-based and randomization methods have the same reasoning process but the norm-based methods rely on many invisible concepts behind mathematical manipulations. The randomization method will decrease cognitive load by reducing the number of concepts that need to be activated simultaneously. The method will enable tactile and visual experiences and allow students conceptual access to the logic of inference." (Pfannkuch et al. 2011, 911)

Die Software TinkerPlots[1], entwickelt in den USA von Cliff Konold und Craig Miller, adaptiert für Deutschland von Rolf Biehler, Daniel Frischemeier und der Autorin dieser Arbeit, ist eine Datenanalyse- und Simulationssoftware, die speziell für Lernende entwickelt wurde und international hoch angesehen ist. „The most recent and innovative of computer applications in the field of probability education, [is] TinkerPlots 2" schreiben zum Beispiel Pratt und Ainley (2013, 167). Die Software TinkerPlots erscheint für die Durchführung von Randomisierungstests sehr gut geeignet, weil sie es erlaubt, das Simulationsmodell selbst zu erstellen, wodurch eine tiefere Einsicht der benötigten Annahmen ermöglicht wird. Auch die Auswertung geschieht nicht „auf Knopfdurck", sondern die wesentlichen Schritte der Analyse müssen aktiv vollzogen werden (nähere Ausführungen siehe Kapitel 3), was ebenfalls zu einem inhaltlichen Verständnis beitragen kann. Um eine Software jedoch kompetent einsetzen zu können, müssen ihre Möglichkeiten und Beschränkungen in Bezug auf Simulationen genau analysiert werden, damit die Software als Instrument von Lernenden eingesetzt werden kann (Rabardel 2002, Trouche 2004). Eine solche Analyse der Software TinkerPlots für stochastische Simulationen soll im ersten Teil dieser Arbeit geschehen.

Zum Einsatz von Randomisierungstests als Einführung in Inferenzstatistik liegen nur wenige Erkenntnisse vor. Es finden sich einige wenige Studien, in denen einzelne Aspekte dazu untersucht wurden. Pfannkuch, Budgett und Arnold (2015) beispielsweise haben eine kurze Lerneinheit entwickelt, in der in Randomisierungstests mittels der von ihnen entwickelten VIT-Software eingeführt wird und mit einer Studie untersucht, wie Studierende damit kausale Schlussfolgerungen ziehen. Biehler, Frischemeier und Podworny (2015) und Frischemeier (2017) haben untersucht, welche Schritte eines Randomisierungstests Studierende nach einer kurzen Einführung ausführen können. Noll und Kirin (2017) haben untersucht, welche Modelle Studierende in TinkerPlots erstellen, um einen Randomisierungstest abzubilden und wie sie diese begründen. Wie

[1] http://www.tinkerplots.com (abgerufen am 20.05.2018)

sich der Einsatz von Randomisierungstests auf affektive Einstellungen von Studierenden auswirkt im Vergleich zu einem traditionellen Kurs, haben Swanson et al. (2014) untersucht. Die Ergebnisse der verschiedenen Untersuchungen zeigen positive Effekte bezüglich einzelner Aspekte und gehen häufig mit einer hohen Motivation der Lernenden einher. Eine Darstellung und Untersuchung des gesamten Argumentationsprozesses von Lernenden während der Durchführung eines Randomisierungstests ist bisher jedoch noch nicht geschehen und soll mit dieser Arbeit detailliert erfolgen.

Nach dem Vorbild der neuseeländischen Autoren um Maxine Pfannkuch und Stephanie Budgett wurde nach dem Design-Based Research Ansatz für die etablierte Lehrveranstaltung „Modellieren, Größen, Daten und Zufall I“ der Universität Paderborn eine Lerneinheit zur Inferenzstatistik entwickelt, in der über Randomisierungstests in die Logik des inferentiellen Schließens eingeführt wird. Daran anschließend wurde eine Studie durchgeführt, in der der Bearbeitungsprozess von Studierenden aufgenommen wurde, die zu zweit kollaborativ einen Randomisierungstest mit der Software TinkerPlots durchführen. Die neu designte Lerneinheit und ausführliche Einblicke in die Bearbeitungsprozesse der Studierenden, die einen Randomisierungstest mit TinkerPlots durchführen, werden im zweiten Teil dieser Arbeit vorgestellt. Dies kommt der Forderung nach Forschung in diesem Bereich nach, die explizit gefordert ist.

> „Areas in which research attention is needed include: (…) The critical evaluation of the use of alternative methods (e.g. randomisation tests) when first introducing statistical inference.” (Harradine, Batanero & Rossman 2011, 12)

1.2 Ziele der Arbeit

Mit der vorliegenden Arbeit werden drei Ziele verfolgt. Das erste Ziel liegt darin, das Unterstützungspotential der Software TinkerPlots für stochastische Simulationen zu analysieren. Zur Erreichung dieses Ziels wird eine theoretische Werkzeuganalyse durchgeführt, die an die Arbeit von Maxara (2009) anknüpft, die eine solche Werkzeuganalyse konzeptionalisiert und für die Software Fathom durchgeführt hat. Eine solche Werkzeuganalyse ist hilfreich, um im Prozess der instrumentellen Genese von Lernenden (Rabardel 2002) eine Software beherrschbar zu machen und die genauen Bedingungen im Hinblick auf einen besonderen Einsatz zu kennen.

Nachdem im ersten Teil dieser Arbeit das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation herausgearbeitet wurde, wird auf dieser Vorarbeit und auf eigenen Lehrerfahrungen aus dem Seminar „Angewandte Stochastik – Mit TinkerPlots komplexe Probleme simulieren und lösen“ (Podworny 2013, Podworny & Biehler 2014), das von der Autorin dieser Arbeit gemeinsam mit Rolf Biehler geplant und im Sommersemester 2012 und im Wintersemester 2012/2013 an der Universität Paderborn gehalten wurde, die Software TinkerPlots in der Vorlesung „Modellieren,

Größen, Daten und Zufall I“ für Studierende des Grundschullehramts seit 2013 intensiv für die Datenanalyse und für stochastische Simulationen eingesetzt.

Das zweite Ziel der Arbeit liegt darin, eine Lerneinheit zur Inferenzstatistik mit Randomisierungstests zu entwickeln und in die Veranstaltung „Modellieren, Größen, Daten und Zufall I“ zu implementieren, die Forderungen der aktuellen didaktischen Forschung nachkommt. Darin soll die Software TinkerPlots zur stochastischen Simulation genutzt werden, um Randomisierungstests durchzuführen. An diese Lerneinheit schließt eine explorative Studie an zur Analyse von Bearbeitungsprozessen von Lernenden, die eine Randomisierungstestaufgabe mit der Software TinkerPlots bearbeiten.

Das dritte Ziel dieser Arbeit ist es, diese Studie auszuwerten. Dazu werden Bearbeitungsprozesse von Studierenden, die einen Randomisierungstest durchführen, offengelegt und untersucht, wie sich die kollaborativen Bearbeitungsprozesse dabei gestalten. Für die Durchführung des Randomisierungstests wird die Software TinkerPlots zur Simulation genutzt, deren Verwendung im Bearbeitungsprozess ebenfalls untersucht werden soll.

1.3 Aufbau der Arbeit

Die Arbeit gliedert sich entsprechend der Ziele in zwei Teile. Im ersten Teil dieser Arbeit (Kapitel 2 bis 4) wird die Werkzeuganalyse zur allgemeinen stochastischen Simulation der Software TinkerPlots begründet und durchgeführt. Der zweite Teil dieser Arbeit (Kapitel 5 bis 11) widmet sich aufbauend auf den durchgeführten Analysen des ersten Teils dem Thema Inferenzstatistik. Hier wird die neu designte Lerneinheit vorgestellt und die dazu durchgeführte explorative Studie vorgestellt und ausgewertet.

Teil I: Theoretische Werkzeuganalyse der Software TinkerPlots zur stochastischen Simulation

Im zweiten Kapitel wird kurz der Nutzen von Simulationen als Lerngegenstand umrissen und ausführlich die Mensch-Werkzeug Interaktion durch die Theorie der instrumentellen Genese erläutert. Ein Modell des Wissenserwerbs in multimedialen Lernumgebungen, zu denen die Software TinkerPlots gezählt wird, wird vorgestellt und der Simulationsplan als didaktisches Hilfsmittel erläutert.

Die theoretische Werkzeuganalyse der Software TinkerPlots zur stochastischen Simulation wird im dritten Kapitel durchgeführt. Zu Beginn des Kapitels werden die Ziele und Methoden der Werkzeuganalyse vorgestellt und ein allgemeiner Überblick über die Software gegeben. Es folgt eine Einordnung der Software als kognitives Werkzeug. Daraufhin wird die Software für die stochastische Simulation strukturiert und in den nächsten Abschnitten die verschiedenen Interfaces vorgestellt. Dabei werden vier verschiedene Arten von Simulationsumsetzungen identifiziert und erklärt.

Im vierten Kapitel wird ein Perspektivwechsel vorgenommen von der Software auf die Stochastik. Im dritten Kapitel wurde ausgehend von der Software als Werkzeug erläutert, welches Unterstützungspotential zur stochastischen Simulation die Software bietet. In diesem vierten Kapitel wird nun von ausgewählten stochastischen Problemstellungen (aufbauend auf der Klassifikation von Maxara 2009) ausgegangen und die Simulationskapazität der Software TinkerPlots daraufhin detailliert untersucht. Dazu werden die verschiedenen, in Kapitel 3 identifizierten Simulationsarten miteinander verglichen.

Die Erkenntnisse aus diesem ersten Teil der Arbeit werden für die Entwicklung der Lerneinheit zur Inferenzstatistik und für das Design der Fallstudie im zweiten Teil genutzt.

Teil II: Statistisches Schließen mit Randomisierungstests: Lerneinheit und Explorative Fallstudie zur Durchführung von Randomisierungstests mit TinkerPlots

Zu Beginn des zweiten Teils werden im fünften Kapitel Grundlagen zur Inferenzstatistik dargelegt und der Nutzen von Randomisierungstests als Einführung in die Logik des inferentiellen Schließens aufgezeigt. Daran anschließend wird der fachliche Hintergrund von Randomisierungstests beleuchtet und dargelegt.

Didaktische Literatur zu Randomisierungstests wird im sechsten Kapitel aufgearbeitet. Neben Vorstellungen, die Lernende zu Randomisierungstests haben, sollen auch mögliche Fehlvorstellungen aufgedeckt werden. Verschiedene Schemata (im Sinne von Ablaufplänen) zu Randomisierungstests werden vorgestellt und verglichen. Aus den Darstellungen werden Konzepte synthetisiert, die bei der Durchführung eines Randomisierungstests adressiert werden müssen. Diese werden für die Auswertung der Fallstudie im neunten Kapitel verwendet.

Im siebten Kapitel werden Forschungsfragen für den zweiten Teil entwickelt und die methodischen Hintergründe für die Entwicklung der Lerneinheit und für die Fallstudie erläutert.

Die nach dem Design-Based Research Ansatz entwickelte Lerneinheit „Inferenzstatistik mit Randomisierungstests“ wird im achten Kapitel vorgestellt. Daran schließt sich die Implementierung der Lerneinheit im Sommersemester 2017 an der Universität Paderborn in den Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen für Modellieren, Größen, Daten und Zufall“ an.

Das Design der Fallstudie wird im neunten Kapitel vorgestellt. Die Aufgabe der Fallstudie wird dargelegt und ausführlich analysiert, auch im Hinblick auf mögliche

Schwierigkeiten der Teilnehmer[2]. Der geplante und der tatsächliche Ablauf der Studie werden beschrieben und die Teilnehmer und die erhobenen Daten werden vorgestellt.

Im zehnten Kapitel wird die Auswertungsmethodik zur Fallstudie beschrieben. Da sich die Auswertungen der Fallstudie in Kapitel elf in hohem Maße auf die verbalen Äußerungen der Teilnehmer stützen, wird zunächst erläutert, wie von Gesprächsteilnehmern Bedeutung in interaktionistischen Prozessen ausgehandelt wird. Darauf aufbauend wird dargelegt, mit welchen Methoden Kommunikationsprozesse und Aktionen mit der Software analysiert werden, die neben den sprachlichen Äußerungen im Mittelpunkt des Forschungsinteresses stehen. Am Ende des Kapitels wird das angepasste Vorgehen zur Auswertung der Fallstudie umfassend dargelegt.

Die Auswertungen der Fallstudie werden im elften Kapitel dargelegt. Zunächst werden die Einzelfallanalysen ausführlich vorgestellt und am Ende des Kapitels die Ergebnisse der Analysen im Hinblick auf die Unterforschungsfragen verglichen und zusammengefasst.

Die Arbeit schließt mit dem zwölften Kapitel, in welchem ein Fazit gezogen und ein Ausblick gegeben wird. Die wichtigsten Ergebnisse der Arbeit werden hier noch einmal zusammengefasst, die Forschungsfragen beantwortet und in Bezug auf die Bedeutung für die Stochastikdidaktik eingeordnet. Ein Ausblick auf mögliche weitere Forschung beendet die vorliegende Arbeit.

[2] In der vorliegenden Arbeit werden „geschlechtsneutrale“ Bezeichnungen gewählt und auf die Angabe der femininen Form ausschließlich zugunsten der Lesbarkeit verzichtet.

TEIL I: THEORETISCHE WERKZEUGANALYSE ZUR STOCHASTISCHEN SIMULATION MIT TINKERPLOTS

2 Hintergründe der theoretischen Werkzeuganalyse

In diesem Kapitel werden die theoretischen Hintergründe zum Einsatz von computergestützten Simulationen in Lernsituationen umrissen. Dabei liegt ein besonderer Schwerpunkt auf der instrumentellen Genese nach Rabardel (1995), die die beiden folgenden Kapitel rechtfertigt, in denen das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation analysiert (Kapitel 3) und an ausgewählten Beispielen didaktisch untersucht (Kapitel 4) wird.

Die Darstellung des Nutzens von Simulationen im Unterricht wird zu Beginn nur kurz umrissen, da diese als Methode für den Unterricht etabliert sind (z. B. Biehler & Engel 2015, Borovcnik 2014, Burrill & Biehler 2011, Chaput, Girard & Henry 2011; Garfield & Ben-Zvi 2008a, Lock, Lock, Lock Morgan, Lock & Lock 2013, Prömmel 2013, Zieffler & Catalysts for Change 2013). Es folgt die Darstellung der Mensch-Werkzeug Interaktion nach dem Ansatz der instrumentellen Genese. Anschließend wird kurz ein Modell für den Wissenserwerb in multimedialen Lernumgebungen vorgestellt und das Kapitel schließt mit dem Simulationsplan als didaktischem Konzept zur Unterstützung des Wissenserwerbs mittels stochastischer Simulationen.

2.1 Nutzen von Simulationen

Mit dem Aufkommen von leistungsstarken Computern in den letzten Jahren und daraus folgend der Möglichkeit, Zufallszahlen in großer Anzahl zu produzieren, werden Computersimulationen als Ergänzung oder sogar als Ersatz für statistische Berechnung mehr und mehr verwendet. Simulationen sind ein mächtiges Instrument, um den Zufall erfahrbar zu machen (Biehler & Engel 2015). Mit Simulationen können Probleme näherungsweise gelöst werden, für die eine theoretische Berechnung zu lange dauern würde oder die schlicht für den Benutzer (noch) nicht zugänglich sind. Außerdem können reale Probleme mithilfe von Modellen der zugrunde liegenden Situation untersucht werden, die weit über typische Standardbeispiele der (Schul-)Stochastik hinausgehen.

Für das Unterrichten von Stochastik wurde das Potential der computergestützten Simulationen schon früh gesehen (z. B. Biehler 1991, Gnanadesikan, Scheaffer & Swift 1987, Schupp, Berg, Dabrock & Pohlmann 1992). Mittlerweile gibt es viele Vorschläge, Simulationen in die Stochastikausbildung zu integrieren, sie haben unter anderem Aufnahme in die Bildungsstandards Mathematik gefunden (Kultusministerkonferenz 2012, 26). Auch der Arbeitskreis Stochastik empfiehlt den Einsatz von Simulationen im Unterricht:

> „Der Stochastikunterricht sollte ferner durch einen hohen Stellenwert experimenteller Arbeiten und durch selbständige Datenerhebungen charakterisiert sein. Dabei sind oft Computer zur Darstellung

> und Auswertung von Daten oder zur Simulation sinnvoll einsetzbar. Die Verfügbarkeit neuer Technologien (Computer, grafikfähiger TR, Internet) bedeutet weittragende Änderungen im Stochastikunterricht. Neben den auch aus anderen Gebieten der Schulmathematik bekannten Möglichkeiten (Erledigung aufwendiger Berechnungen, interaktives Erstellen von Grafiken und Schaubildern) sind hier vor allem die Datenbeschaffung durch das Internet sowie die flexible und leichte Durchführbarkeit von Demonstrationen und Simulationen zu nennen. Vom Zufallsgenerator erzeugte Daten können maßgeblich dazu beitragen, bei Schülern eine Intuition für zufallsbedingte Variabilität in empirischen Daten zu entwickeln.“ (Arbeitskreis Stochastik 2003, 2)

Biehler (1991) unterscheidet zwei Perspektiven, unter denen Simulationen in der Stochastikausbildung eingesetzt werden können. Zum einen lassen sich Simulationen als Werkzeug einsetzen, um stochastische Probleme zu lösen. Zum anderen können sie benutzt werden, um Erfahrungen über zufallsabhängige Vorgänge zu gewinnen, und dadurch die Begriffsbildung fördern (vgl. Biehler 1991, 128ff). Biehler und Maxara (2007) differenzieren diese Perspektiven weiter aus und unterscheiden für den Werkzeugeinsatz drei Perspektiven, unter denen ein Werkzeug für Simulationen in der Stochastikausbildung eingesetzt werden kann:

> „1. Simulation zur Repräsentation von Zufallsexperimenten
> 2. Simulation als Werkzeug im Wechselspiel mit analytischen (kombinatorischen) Methoden
> 3. Simulation als Werkzeug, als Methode sui generis“ (Biehler & Maxara 2007, 45)

Diese drei Perspektiven werden nun kurz zusammengefasst.

Simulation zur Repräsentation. Computersimulationen bieten die Möglichkeit, Zufallsexperimente in großer Anzahl durchzuführen, an den Stellen, an denen dies in Unterrichtssituationen aus Zeitmangel kaum möglich ist. Dadurch lassen sich Zufallsexperimente eigenständig durchführen, wodurch stochastische Konzepte erfahrbar gemacht werden können.

Simulation als Werkzeug im Wechselspiel. Simulationen können helfen, analytisch gewonnene Ergebnisse zu überprüfen und können anders herum einen Anhaltspunkt geben für analytische Ergebnisse. Über die Modellbildung kann die Simulation mit weniger Fachmathematik auskommen und dadurch einen einfacheren Zugang zu Problemen bieten, die später theoretisch aufgearbeitet werden (vgl. Eichler & Vogel 2013, 232).

Simulation als Werkzeug sui generis. Im Gegensatz zur vorherigen Perspektive können Simulationen auch eingesetzt werden bei Problemen, zu denen es (noch) keine analytische Lösung gibt. Die Simulationsmethode bietet hier den einzig möglichen Ansatz. Durch diese Perspektive als Problemlösemethode kann einem einseitigen Bild auf die Mathematik vorgebeugt werden.

Diese Perspektiven werden ähnlich auch von Kütting und Sauer (2011) formuliert und die letztere betont: „Das eigentliche Anwendungsgebiet der Simulationsverfahren ist freilich die Lösung stochastischer Probleme, deren Komplexität eine rechnerische Lö-

sung nicht zulässt.“ (Kütting & Sauer 2011, 226). Hierin sehen diese Autoren auch die größte Motivation, sich mit Simulationen in der Stochastikausbildung zu beschäftigen.

Tietze, Klika und Wolpers (2002) fordern, dass Simulationen ein fester Bestandteil des gesamten Stochastikunterrichts sein sollen und führen folgende Gründe dafür an:

- „Die Simulation ist ein wichtiges Verfahren zur Modellbildung in Theorie und Praxis.
- Die Modellkonstruktion durch Simulation vermittelt epistemologische Einsichten in die Rolle von Modellen bei der Mathematisierung von Ausschnitten der Realität, indem mit Hilfe von Simulationen Erfahrungen und Einsichten in den Zusammenhang von stochastischer Theorie und den empirischen Entsprechungen gewonnen werden können. Für die Aufhellung der Wechselbeziehungen zwischen Empirie und Theorie sind insbesondere solche Probleme geeignet, deren Lösung analytisch und empirisch-experimentell möglich ist.
- Simulationen fördern Fähigkeiten im Modellbilden.
- Simulationen sind wichtig für den Erwerb stochastischen Denkens: Dies gilt z. B. für den Erwerb und die Einschätzung zentraler probabilistischer Begriffe wie Zufall, Wahrscheinlichkeit, Erwartungswert, Signifikanzintervall usw.
- Durch Simulationen lassen sich auch dann Probleme lösen, deren vollständige analytische Lösung im Unterricht nicht möglich oder zu aufwendig wäre.
- Simulationen verlangen planerische, ausführende und beurteilende Tätigkeiten, also Projektarbeit. Eigentätigkeit hat positive Auswirkungen auf das Lernverhalten, weil die aktive Auseinandersetzung mit den Begriffen und Verfahren der Stochastik eine bessere Einbettung von deklarativem oder operativem Wissen in die kognitive Struktur ermöglichen. Insbesondere sind positive Auswirkungen auf die Veränderung falscher primärer Intuitionen und die Entwicklung angemessener sekundärer Intuitionen zu erwarten.
- Simulationen fördern die Motivation. Dies gilt besonders, wenn Probleme bearbeitet werden, deren Lösung ungewiss (z. B. Paradoxa der Stochastik) oder überraschend ist (z. B. Geburtstagsproblem).“ (Tietze et al. 2002, 130)

Auch diese Autoren sehen die Unterstützung durch Computereinsatz für Simulationen als äußerst sinnvoll an. Batanero und Borovcnik (2016) betonen, dass Simulationen helfen können, die Sicht auf die verschiedenen Naturen von subjektiver, klassischer und frequentistischer Sicht zu verbessern. Auch Engel & Grübel kommen zu ähnlichen Schlüssen: „Simulationen unterstützen valide Vorstellungen bezüglich Zufall und Wahrscheinlichkeit und konfrontieren Fehlvorstellungen. […] Sie [die Simulationen] erlauben aber einen entdeckenden Arbeitsstil, indem Lernende selbst aktiv Daten produzieren und analysieren und mit Zufallsstichproben einer Population experimentieren, deren Parameter bekannt sind.“ (Engel & Grübel 2008, 116).

Eichler und Vogel (2013, 173) kommen zu folgendem Schluss: „Simulationen können helfen, ein intuitives Verständnis von zufälligen Vorgängen auf- und auszubauen. […] Durch die externe Darstellung des Simulationsvorgangs kann so die mentale Modellbildung angeregt werden. […] Das wesentliche stochastikdidaktische Potenzial der Simulationen ist darin zu sehen, dass damit das Verständnis des Wahrscheinlichkeitsbegriffs vertieft und schließlich das Tor zur schließenden Statistik aufgestoßen werden kann.“ Sie sehen Simulationen als „didaktische Brücke von der beschreibenden Statistik in die Wahrscheinlichkeitsrechnung und weiter in die schließende Statistik hinein“ (ebd.

S. 323). In diesem Sinne wird in dieser Arbeit die Software TinkerPlots eingesetzt. Vor einem solchen Einsatz steht jedoch die Auseinandersetzung auf theoretischer Ebene mit der Software, was in Abschnitt 2.2 geschieht.

2.2 Theorie der Mensch-Werkzeug Interaktion

Eine Software, auch zur Simulation, für sich alleine ist nichts weiter als eben das. Sie wird manifestiert durch ein Datenpaket auf dem Computer, aber solange sie nicht benutzt wird, ist sie nichts weiter als eine Ansammlung von Einsen und Nullen. Erst durch Gebrauch wird eine Software zu einem Werkzeug, mit dem sich (sinnvolle) Tätigkeiten ausführen lassen. Wie aus einer Software ein Werkzeug, ja sogar ein Instrument für ein Subjekt wird, beschreibt der theoretische Ansatz der „Instrumentellen Genese“. Dieser Ansatz geht zurück auf Rabardel (1995), der ausführlich das Zusammenspiel zwischen einem Subjekt und einem Artefakt beschreibt und den Prozess, wie aus diesem Zusammenspiel ein Instrument wird, erläutert. Rabardels Zugang zu einem Werkzeug wird hier als Grundlage für das Verständnis der Mensch-Werkzeug Interaktion gesehen und mit weiterer Literatur und darauf aufbauenden Studien aufgearbeitet. In Abschnitt 2.2.2 wird aus didaktischer Sicht der Ansatz der „instrumentellen Orchestrierung“ als Anwendung in Unterrichtssituationen zur Förderung der instrumentellen Genese von Lernenden beschrieben. Die instrumentelle Orchestrierung geht zurück auf Guin & Trouche (2002) und Trouche (2004).

Der Ansatz der instrumentellen Genese ist in **Abbildung 1: Ansatz zur instrumentellen Genese**Abbildung 1 zu sehen.

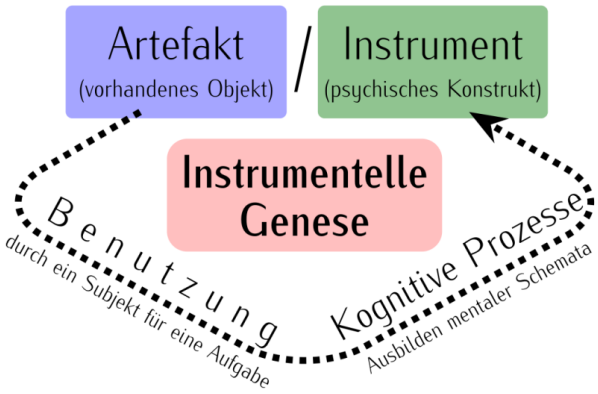


Abbildung 1: Ansatz zur instrumentellen Genese

Ausgegangen wird von einem Artefakt, im speziellen von Computersoftware. Durch die Benutzung durch ein Subjekt zur Aufgabenbearbeitung werden mentale Prozesse aktiviert und es geschieht eine kognitive Anpassung des Subjekts an das Artefakt. Diese kognitive Anpassung, die während der Benutzung geschieht, macht das Werkzeug zu

einem sinnvoll nutzbaren Instrument für das Subjekt. Erst wenn das Artefakt das Stadium des Instruments für ein Subjekt erreicht hat, lassen sich Aufgaben mit dessen Hilfe sinnvoll bearbeiten.

In Abschnitt 2.2.1 wird zunächst der instrumentelle Ansatz allgemein für die Benutzer-Werkzeug Beziehung beschrieben. Im anschließenden Abschnitt 2.2.2 wird als Ansatz der Didaktik die instrumentelle Orchestrierung besprochen, durch die die instrumentelle Genese in Lehrsituationen gelenkt wird.

2.2.1 Der Ansatz der instrumentellen Genese

Im Mittelpunkt der folgenden Betrachtung steht die Benutzer-Werkzeug Interaktion nach der Rabardelschen Theorie. 1995 hat Pierre Rabardel eine Schrift aus Sicht der kognitiven Ergonomie verfasst mit dem Titel „Les Hommes et les Technologies – une approche cognitive des instruments contemporains“ (Rabardel 1995/Rabardel 2002)[3]. Wie im Titel bereits anklingt, geht es darin um die Beziehung zwischen Mensch und „Technologie“, also zeitgenössischen Werkzeugen bzw. Instrumenten. Für die vorliegende Arbeit ist dies von besonderem Interesse, da beim Einsatz von Computersoftware die Entwicklung der Beziehung zwischen dem Benutzer und der Software als Instrument eine besondere Rolle spielt. Rabardel selbst nimmt alle möglichen Instrumente in seine theoretische Betrachtung mit auf, von materiellen Artefakten wie zum Beispiel einem Hammer über kulturelle Artefakte wie Sprache hin zu komplexen Werkzeugen, wie zum Beispiel eine Werkbank, die vorrangig von ihm betrachtet werden. Im weitesten Sinne wird von ihm nahezu alles, was benutzbar ist, als Instrument betrachtet. Rabardel baut seine Theorie unter anderem auf Grundlagen des russischen Psychologen Lew S. Wygotski (1896 – 1934) auf. Dieser gilt als Begründer der Tätigkeitstheorie in der Psychologie. Wygotski beschreibt in dem Werk „Die instrumentelle Methode in der Psychologie“ (Wygotski 1985) instrumentell vermittelte Handlungen als psychische Prozesse. Er bezieht „psychische Instrumente“ beziehungsweise „Werkzeuge“ in den Verhaltensprozess des Menschen ein, der dadurch beherrschbar wird. Die Werkzeuge selbst sind dabei Produkte kulturhistorischer Evolution. Wygotski beschreibt den „instrumentellen Akt“ wie folgt:

> „Die psychischen Prozesse insgesamt, die eine komplizierte strukturelle und funktionale Einheit darstellen, was das Gerichtetsein auf die Lösung der vom Objekt gestellten Aufgabe sowie die Stimmigkeit und die durch das Werkzeug vorgeschriebene Verlaufsweise anbelangt, bilden ein neues Ganzes, den instrumentellen Akt.“ (Wygotski 1985, 313)

Bei den psychischen Prozessen hat Wygotski die kulturhistorische Komponente im Blick: „Die zweiten [instrumentellen Formen des Verhaltens] stellen eine späte Errun-

[3] Nachfolgend zitiert nach der englischen Übersetzung von 2002 der französischen Schrift von Rabardel (1995).

genschaft der Menschheit dar, sie sind ein Produkt der historischen Entwicklung und eine spezifisch menschliche Verhaltensform.“ (Wygotski 1985, 314). Auch die Vermittlung des Werkzeuggebrauchs geschieht im kulturellen Umfeld. Durch „Erziehung“ (ebd.) kann direkter Einfluss auf den instrumentellen Akt genommen werden.

Dieser instrumentelle Akt ist die Grundlage dafür, eine Aufgabe zielgerichtet mit Hilfe eines Werkzeugs bearbeiten zu können. Die dabei ablaufenden kognitiven Prozesse konstituieren diesen Akt nach Wygotski. Die Werkzeuge werden dabei von Wygotski in psychische und technische unterschieden. Er selbst nennt als Beispiele für psychische Werkzeuge und daraus resultierende komplexe Systeme „die Sprache, verschiedene Formen der Nummerierung und des Zählens, mnemotechnische Mittel, die algebraischen Symbole, Kunstwerke, die Schrift, Schemata, Diagramme, Karten, Zeichnungen, alle möglichen Zeichen und ähnliches mehr“ (Wygotski 1985, 310). Von diesen „psychischen Werkzeugen“ grenzt er „technische Werkzeuge“ ab, die auf die materiell fassbare Umwelt des Menschen wirken. Im Sinne Rabardels können die technischen Werkzeuge als materielle Objekte betrachtet werden (z. B. ein Hammer oder eine Maschine). Das Werkzeug spielt beim Aufbau des instrumentellen Akts eine zentrale Rolle:

> „Das in den Verhaltensprozess eingeschlossene psychische Werkzeug bestimmt mit seinen Eigenschaften den Aufbau des instrumentellen Aktes und verändert den gesamten Verlauf sowie die gesamte Struktur der psychischen Funktionen in derselben Weise, wie technisches Werkzeug den Prozess der natürlichen Anpassung verändert, indem es die Form von Arbeitsoperationen bestimmt.“ (Wygotski 1985, 310)

Die Eigenschaften eines Werkzeugs, gleich ob psychisches oder technisches Werkzeug, bestimmen demnach die Nutzungsart desselben mit. Wygotski setzt dabei voraus, dass eine bestimmte Aufgabe mit Hilfe des instrumentellen Akts gelöst werden soll, bei der ein passendes Werkzeug eingesetzt wird. Das Werkzeug hat nach Wygotski vor allem eine kulturell vermittelte verbindende Funktion zwischen den ablaufenden psychischen Prozessen und dem Objekt.

An dieser Stelle baut Rabardel auf der Theorie des instrumentellen Akts von Wygotski auf und erweitert sie. Er unterscheidet bei einem Werkzeug zwischen Artefakt und Instrument, wobei der Begriff Artefakt das reine Objekt bezeichnet, also das, was Wygotski als „technisches Instrument“ beschreibt. Im Gegensatz dazu handelt es sich beim Instrument nach Rabardel um ein psychisches Konstrukt, das im konstruktivistischen Sinn vom Benutzer gebildet werden muss und somit an Wygotski anknüpft. Den Prozess, wie aus einem Artefakt ein Instrument wird, nennt Rabardel instrumentelle Genese und ordnet ihn wie folgt ein:

> „The instrumental approach is thus situated on a level of analysis of techniques as psychological constituents. It contributes to theoretical reflection and the empirical examination of relations between

> humans and human-centered technical systems from the perspective of operators engaged in real activities and actions situated in their work, training or daily life contexts." (Rabardel 2002, 14)

Aus verschiedenen Forschungsrichtungen entwickelt Rabardel seine Definition eines Instruments, das aus einem Artefakt gebildet wird: „We will define the instrument in the essence of its constituting relation: **the subject's use of the artifact as a means he/she associates with his/her action**" (Rabardel 2002, 18, Hervorhebung im Original). Der Bezug zu einer tatsächlichen Aufgabe oder Tätigkeit unter Einbezug des Instruments steht für Rabardel immer im Vordergrund. Auch beim Aufbau der psychischen Komponenten, die bei der Benutzung eines Artefakts eine Rolle spielen, steht der Bezug zu konkreten Aufgaben im Vordergrund.

Bis hierher wurden die Begriffe Werkzeug, Artefakt und Instrument benutzt. In dieser Arbeit geht es um die instrumentelle Benutzung von Software, eine genaue Einordnung der benötigten Begriffe ist notwendig. Wygotski benutzt die Bezeichnung Werkzeug, das er in psychisches und technisches unterscheidet. Allerdings liegt bei Wygotski der Fokus nicht auf dem Werkzeug, sondern auf den zugehörigen psychischen Prozessen, die den instrumentellen Akt konstituieren. Rabardel legt den Fokus auf das Werkzeug und entwickelt den Werkzeugbegriff weiter. Er unterscheidet ihn in Artefakt und Instrument. Unter Artefakt fasst er sowohl die technischen als auch die psychischen Werkzeuge nach Wygotski zusammen. Im Folgenden wird die Unterscheidung zwischen Artefakt und Instrument näher herausgearbeitet, da sie für die instrumentelle Genese von fundamentaler Bedeutung ist. Der weniger eindeutige Begriff des Werkzeugs wird vernachlässigt bzw. im Sinne eines Oberbegriffs sowohl für Artefakt als auch für Instrument verwendet.

2.2.1.1 Vom Artefakt zum Instrument

Was genau ist nun ein Artefakt und wie wird es zum Instrument? Der Begriff Artefakt soll neutral sein und keine besondere Beziehung zum Objekt spezifizieren. Der sprachlichen Ebene wird hier einige Bedeutung beim Aufbau der Theorie zugemessen („Choosing vocabulary is always a very important question." Trouche 2004, 1), deshalb ist die genaue Differenzierung zwischen Artefakt und Instrument wichtig. Erst wenn ein Artefakt in einer Situation Anwendung findet, wird es zum Instrument (Rabardel 2002, 39). Für ein Individuum hat ein Artefakt zunächst keinen instrumentellen Wert, sondern erst durch den Prozess der instrumentellen Genese wird es zum nutzbaren Instrument (vgl. Artigue 2002). Wie oben bereits erwähnt, kann ein Artefakt ein materielles Objekt wie ein Hammer sein, ein Klavier, eine Maschine, aber auch etwas metaphysisches wie Sprache oder ein Computerprogramm können als Artefakt angesehen werden. Bei einem Computerprogramm beispielsweise stellt sich die Frage, was als Artefakt betrachtet wird. Ist der Computer das Artefakt? Ist eine spezielle Software ein Artefakt? Oder ist eine bestimmte Funktion der Software das Artefakt und das Programm dazu eine

Sammlung von Artefakten? Je nach Sichtweise lassen sich verschiedene Ebenen eines Objekts als Artefakt beschreiben: „… it is a matter of granularity if one considers the [dynamic geometry] software as a whole as one single artifact, or if one sees it as a collection of artifacts, such as the construction artifact, the measurement artifact, the dragging artifact, and so on." (Drijvers, Kieran, et al. 2010, 108). Hierzu gibt es keine einheitlichen Vorgaben, verschiedene Sichtweisen sind durchaus gängig. Als kontrastierende Beispiele seien zum einen Ben-Zvi und Ben-Arush (2014) genannt, die eine einzelne Operation innerhalb des Programms TinkerPlots als Artefakt betrachten. Diese beiden Autoren sehen beispielsweise die Operation „stack" als einzelnes Artefakt an und untersuchen dies genauer. Zum anderen sei als weiteres Beispiel Kieran und Drijvers (2006) genannt, die den Taschenrechner TI-92 mit all seinen Funktionen insgesamt als Artefakt betrachten und analysieren. Da in den Folgekapiteln eine Werkzeuganalyse bezüglich der gesamten Simulationskapazität der Software TinkerPlots durchgeführt werden soll, wird hier in dieser Arbeit der Ansatz verfolgt, eine Software insgesamt als ein einziges Artefakt zu betrachten, mit einer Ansammlung von Funktionen und Interfaces.

Interessant ist nun der Übergang vom Artefakt zum Instrument. Damit ein Artefakt als Instrument angesehen werden kann, muss die Tätigkeit eines Benutzers mit diesem Artefakt vorliegen. Der Benutzung eines Artefakts als Werkzeug zum Bearbeiten einer bestimmten Aufgabe liegen mentale Schemata zugrunde, die Rabardel *utilization schemes*[4] nennt. Sie sind grundlegend für die Konstituierung eines Instruments, also für den Übergang vom bloßen Artefakt als Objekt zum kognitiven Instrument. Schemata sind in der Psychologie mentale Wissensstrukturen, die es einem Individuum erlauben, sich in einer Situation schnell zurechtzufinden und sich sinnvoll zu verhalten. Solche Schemata müssen für die Benutzung eines Artefakts vom Subjekt konstruiert und trainiert werden. Ein Schema hat drei Hauptfunktionen, erstens eine epistemische Funktion, die dem Individuum erlaubt zu verstehen, was es tut; zweitens eine pragmatische Funktion, die es einem Individuum erlaubt, etwas zu tun und drittens eine heuristische Funktion, die dem Individuum erlaubt, die eigenen Handlungen einzuordnen und zu kontrollieren (Rabardel 2002, 85).

Rabardel betont, dass *utilization schemes* immer eine soziale Komponente haben und im sozialen Zusammenhang erworben werden:

> „Schemes are the object of more or less formalized transmissions and transfers: information passed on from one user to another; training structured around complex technical systems; various types of

[4] Die in dieser Arbeit aufgegriffenen Theorien beinhalten teilweise Termini, für die es keine prägnante oder eindeutige Übersetzung ins Deutsche gibt. Um Verwechslungen oder weitschweifende Erläuterungen zu vermeiden, wird deshalb in der Regel der englische Terminus übernommen. Deutlich gemacht wird dies durch eine Kursivsetzung des jeweiligen Begriffs.

> users' support (instruction manuals, users' guides and various other supports introduced or not in the artifact itself). This is why we have called **u. sch. social utilization schemes (S. U. Sch.)**." (Rabardel 2002, 84, Hervorhebung im Original)

Bei diesen *utilization schemes* können zwei Arten von Schemata unterschieden werden. Sogenannte *usage schemes* orientieren sich in Richtung der Benutzung des Artefakts, dazu gehört das Anstellen des PCs, Aufrufen des Programms, Bedienen der Maus, etc. Weiterhin gibt es die *instrumented action schemes*. Diese beziehen sich auf das Lösen einer speziellen Aufgabe mit Hilfe des Instruments (z. B. den Grenzwert einer Funktion zu berechnen). In der nachfolgenden Werkzeuganalyse und beim Benutzen von Software spielen diese *instrumented action schemes* eine übergeordnete Rolle im Hinblick auf zu lösende Aufgaben und spiegeln das Wissen wider, dass ein Subjekt über ein spezielles Instrument hat. Wie zu sehen ist, wird durch diese Unterscheidung von Rabardel beispielsweise eine Computersoftware als ein Artefakt betrachtet, zu dem sogar die physischen Komponenten des Computers gezählt werden.

Entsprechend werden von Rabardel diese beiden Komponenten, Artefakt und Gebrauchsschema, in die Definition eines Instruments mit aufgenommen:

> „We propose defining the instrument as a mixed entity, born of both the subject and object (in the philosophical sense of the term): the instrument is a composite entity made up of an artifact component (an artifact, a fraction of an artifact or a set of artifacts) and a scheme component (one or more utilization schemes, often linked to more general action schemes). An instrument therefore consists of two types of entities:
>
> - a material or symbolic artifact produced by the subject or others;
> - one or more associated utilization schemes, resulting from an autonomous construction specific to the subject, or from an appropriation of S. U. Sch. [Social Utilization Scheme] already formed outside of him/her." (Rabardel 2002, 86)

Schemata können nicht im quasi luftleeren Raum gebildet werden, sondern sind vielmehr verknüpft mit speziellen Anforderungen einer bestimmten Situation bzw. Aufgabe. Umgesetzt werden sie in Form einer Prozedur entsprechend den Besonderheiten einer Situation (Rabardel 2002, 85). Hierdurch entsteht die psychische Komponente des Instruments. Drijvers, Kieran, et al. (2010) betonen dies am Beispiel eines Stifts, der, solange das Individuum keine Kenntnis von Buchstaben hat, zum reinen Malen verwendet wird. Sobald jedoch ein Wissenszuwachs im Sinne von Schreiben-können stattgefunden hat, bekommt der Stift als Instrument eine weitere Bedeutung. „Following Rabardel, we speak of an instrument if a meaningful relationship exists between the artifact and the user for a specific type of task." (Drijvers, Kieran, et al. 2010, 108).

2.2.1.2 Instrumentation und Instrumentalization als konstituierende Prozesse

Die Entwicklung zum Instrument beinhaltet nun die materielle oder symbolische Komponente des Artefakts und die Komponente des kognitiven Schemas. Wie werden diese

Komponenten konstituiert? Rabardel prägt hierfür die beiden Begriffe *instrumentation* und *instrumentalization*:

> „Whether it concerns the scheme or the artifact, this construction generally does not occur in a vacuum. The artifacts are most often pre-existing but are nonetheless instrumentalized by the subject. Schemes most often spring from the subject's repertory and are extended or adapted to the new artifact. Sometimes, entirely new schemes must be constructed: these processes as a whole can be described as instrumentation and instrumentalization processes." (Rabardel 2002, 88f).

Diese beiden Prozesse sind grundlegend für die Beziehung zwischen dem Subjekt und dem Artefakt und führen erst dazu, dass ein Instrument entsteht:

> „This instrumental genesis process is carried by the subject. Because it concerns the two poles of the instrumental entity – the artifact and the utilization schemes, it also has two dimensions and two orientations that are both distinguishable and often related: instrumentalization directed toward the artifact and instrumentation relative to the subject him/herself." (Rabardel 2002, 101).

Kurz gesagt ist der Prozess der *instrumentalization* auf das Artefakt gerichtet und der Prozess der i*nstrumentation* auf das Subjekt. Beide Prozesse gehen vom Subjekt aus, da dieses die Prozesse als mentale Schemata adaptiert.

> „Instrumentalization processes concern the emergence and evolution of artifact components of the instrument: selection, regrouping, production and institution of functions, deviations and catachreses, attribution of properties, transformation of the artifact (structure, functioning etc.) that prolong creations and realizations of artifacts whose limits are thus difficult to determine" (Rabardel 2002, 104).

Der *instrumentalization process* ist der Prozess, durch den das Subjekt bestimmte Komponenten des Artefakts verwenden kann. Dabei werden verschiedene Herangehensweisen genutzt, die die Eigenschaften des Artefakts anreichern oder verändern. Auch innovative Benutzungsweisen des Artefakts, die nicht offensichtlich sind, können auftreten, dies beschreibt der Begriff „catachreses" in der obigen Definition. Eine solche Katachrese beschreibt das Beispiel eines Arbeiters, der mit Hilfe einer speziellen Werkbank Autotüren repariert, das Sich-zu-eigen-Machen seiner Werkbank, in diesem Fall ein physikalisch höchst individuell angepasstes Instrument (Rabardel 2002, 105). Eine genaue Kenntnis und ggf. Individualisierung des Artefakts erleichtert das Arbeiten damit und führt zu einem verbesserten und ggf. physisch veränderten Instrument.[5]

Es gibt zwei Ebenen, auf denen der *instrumentalization process* operieren kann, zum einen kann der Prozess lokal sein und sich auf spezielle Handlungen mit dem Artefakt und spezielle Umstände, unter denen es verwendet wird, beziehen. Auf dieser lokalen Ebene wird das Artefakt vorübergehend für eine spezielle Aufgabe instrumentalisiert. Auf der zweiten Ebene ist die durch den aufgabenbezogenen Prozess angereicherte

[5] Beim Arbeiten mit dem Computer kann dieses individuelle Anpassen beispielsweise das Verändern der Maus sein, die durch das Anbringen zusätzlicher Gewichte den eigenen Bedürfnissen angepasst wird.

Funktion des Artefakts dauerhaft und wird für das Subjekt zu einer Eigenschaft des Artefakts mit einer Klasse von Handlungen, Objekten, Aktivitäten und Situationen. In beiden Fällen findet keine physikalische Transformation des Artefakts statt, es wird lediglich mit neuen extrinsischen Funktionen angereichert, vorübergehend oder dauerhaft, allerdings bezieht sich Rabardel auch auf tatsächliche physische Objekte, wie beispielsweise einen Bleistift und nicht auf Software, die ja auch den eigenen Bedürfnissen angepasst werden kann. Je nach Anforderung kann allerdings auch eine physische Transformation des Artefakts wie oben beschrieben stattfinden.

> „Instrumentation processes are relative to the emergence and evolution of utilization schemes and instrument-mediated action: their constitution, their functioning, their evolution by adaptation, combination coordination, inclusion and reciprocal assimilation, the assimilation of new artifacts to already constituted schemes, etc." (Rabardel 2002, 104)

Der *instrumentation process* hingegen ist vom Artefakt auf das Subjekt gerichtet. Hierbei werden die benötigten mentalen Schemata gebildet, angepasst oder kombiniert.

> „The genesis of these schemes, the assimilation of new artifacts to schemes (thus giving new signification to artifacts), the adaptation of schemes (contributing to their changes in signification), make up this second dimension of instrumental genesis: instrumentation processes." (Rabardel 2002, 109)

Trouche präzisiert diesen Prozess wie folgt, wobei er den Begriff des Schemas im Sinne Rabardels nicht explizit nennt, jedoch umschreibt:

> „[…] instrumentation is precisely this process by which the artifact prints its mark on the subject, i.e., allows him/her to develop an activity within some boundaries (the constraints of the artifact). One might say, for example, that the scalpel instruments a surgeon. This is obviously the case in every CLE (=Computer Learning Environment). To understand this process we first need to study the constraints and "enablements" of an artifact." (Trouche 2004, 290)

Bei diesem Prozess steht die Ausbildung von Schemata im Fokus und sowohl Rabardel als auch Trouche fordern, um diesen Prozess besser verstehen zu können, dass die Rahmenbedingungen und Möglichkeiten eines vorliegenden Artefakts genau untersucht werden. Guin und Trouche (2002) differenzieren bereits für Software drei Arten von Bedingungen eines Artefakts:

- *Internal constraints*, diese stehen in unmittelbarer Beziehung zur Hardware,
- *Command constraints*, diese beziehen sich auf die Existenz und die Form (z. B. Syntax) der verschiedenen Befehle,
- *Organization constraints*, diese beziehen sich auf das Interface zwischen Artefakt und Benutzer. (Guin & Trouche 2002, 10).

Für eine bestimmte Anwendung (Grenzwertberechnung) des graphischen Taschenrechners werden diese *constraints* exemplarisch von Guin und Trouche (2002) untersucht.

Für die Simulationsmöglichkeiten von TinkerPlots geschieht dies detailliert in den Kapiteln 3 und 4.

Direkt beobachtbar ist der *instrumentation process* nicht, denn er drückt sich in der Ausbildung kognitiver Schemata aus, die das Subjekt beim Benutzen des Artefakts entwickelt oder verändert. Die Schemata führen zur Herausbildung von Strategien, die zu effektiven Antworten auf gegebene Aufgaben führen (Artigue 2002). Auch von dieser Autorin wird betont, dass im Prozess der instrumentellen Genese das Werkzeug hinsichtlich der Möglichkeiten und Bedingungen genau analysiert werden muss. Der Unterschied zwischen einem Experten und einem Anfänger beim Gebrauch eines bestimmten Artefakts zeigt sich unter anderem in den herausgebildeten *utilization schemes*. Guin und Trouche (2002) verweisen in diesem Zusammenhang auf das Ausbilden von Strategien, bei denen die mentalen Komponenten der Schemata unberücksichtigt bleiben, die dafür jedoch als Aktivitäten beobachtbar sind. Trouche (2004) unterscheidet zwischen „gesture“ und „scheme“, wobei für ihn die Geste ein beobachtbarer Teil des zugrunde liegenden Schemas ist. Beispielsweise können die Eingaben beim Taschenrechner als Abfolge von Gesten beobachtet werden, die einem Löseschema bestimmter mathematischer Aufgaben zugrunde liegen. „This scheme allows the students to solve a system of two equations with two unknowns. It can appear as a sequence of gestures on the keypad of the calculator, but it requires considerable knowledge, although not necessarily the same knowledge for each student.” (Trouche 2004, 6).

Die beiden Prozesse *instrumentalization* und *instrumentation* sind nicht immer klar voneinander abzugrenzen und treten nicht unabhängig voneinander auf. Vor allem der *instrumentation process* ist kaum direkt beobachtbar, da er im Ausbilden mentaler Schemata mündet, die nur mittelbar durch bestimmte Aktivitäten beobachtbar sind.

2.2.1.3 Zusammenfassung

Die instrumentelle Genese ist hier in Abbildung 2 dargestellt, die die Zusammenhänge und Richtungen deutlich macht.

Dabei ist das Instrument als ein psychologisches Konstrukt zu verstehen, das durch die Beziehung zwischen dem Artefakt und dem Subjekt gebildet wird. Von zentraler Bedeutung sind die beiden Prozesse i*nstrumentalization* und i*nstrumentation*, die einmal auf das Werkzeug und einmal auf das Subjekt gerichtet sind. Durch den *instrumentation process* werden beim Subjekt mentale Schemata aufgebaut, die eine wiederholbare sinnvolle Beschäftigung mit dem Instrument ermöglichen. Im *instrumentalization process* lernt das Subjekt das Werkzeug mit seinen Eigenheiten und Bedingungen kennen. Der gesamte Prozess, der dazu führt, dass ein Werkzeug zu einem Instrument wird, mit

dem eine zuvor gestellte Aufgabe bearbeitet werden kann, wird unter „instrumenteller Genese“ zusammengefasst.

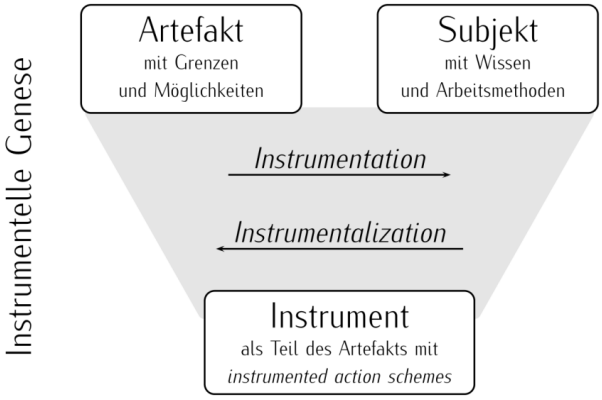


Abbildung 2: Instrumentelle Genese in Anlehnung an Trouche (2004, 289)

2.2.2 Didaktik der instrumentellen Genese: Instrumentelle Orchestrierung

> „Research suggests a close relationship between tool use, cognitive development and social practice. Identification of the nature of these relationships and exploiting them for educational goals, however, is a non-trivial issue. The central problem with which education is confronted, therefore, is to identify the relation between the use of technological tools and learning, and to use these insights as guidelines for technology-rich teaching practices. How can the use of technological tools be embedded in innovative learning/teaching arrangements so that it improves learning?” (Drijvers, Doorman, Boon, van Gisbergen & Gravemeijer 2007, 1)

Drijvers et al. (2007) werfen die zentrale Frage auf, wie technologische Werkzeuge so in Lernsequenzen integriert werden können, dass sie das Lernen verbessern oder bestimmte Aufgaben Lernenden überhaupt erst zugänglich machen. Dieser zentralen Frage wird - aufbauend auf der instrumentellen Genese - als Ansatz der Didaktik mit Hilfe der instrumentellen Orchestrierung nachgegangen. Der Begriff der instrumentellen Orchestrierung geht zurück auf Guin und Trouche (2002) und Trouche (2004). Die Metapher des musikalischen Orchesters hat dabei einen hohen Bezugswert. Während Rabardel generell alle Arten von Objekten in seine Theorie aufnimmt, bezieht sich Trouche (2004) auf die Mensch-Werkzeug Beziehung in sogenannten *computerized learning environments* (CLE). Dabei steht der Gebrauch der Software durch die Lernenden im Vordergrund. Da später in dieser Arbeit auf den Einsatz von Software tiefer eingegangen wird, wurde dieser Ansatz gewählt. Aufbauend auf der instrumentellen Genese steht diese als Prozess von Lernenden in einer Unterrichtssituation im Mittelpunkt der instrumentellen Orchestrierung.

> „We will call instrumental orchestration a plan of action, partaking in a didactic exploitation system which an institution (the school institution, in this case) organizes with the view of guiding students‘ instrumented action. Instrumented orchestration is defined by four components:
>
> - a set of individuals;

- a set of objectives (related to the achievement of a type of task or the arrangement of a work-environment);
- a didactic configuration (that is to say a general structure of the plan of action);
- a set of exploitation of this configuration.“ (Guin & Trouche 2002, 208)

Zu der instrumentellen Orchestrierung gehören demnach als erstes die Lernenden als Individuen mit jeweils eigenen Wegen der instrumentellen Genese. Weiterhin gehören dazu Ziele, die in Verbindung mit gestellten Aufgaben oder Arbeitsumgebungen stehen und eine allgemeine didaktische Konfiguration, die vom Lehrenden vorgegeben wird und den Zugang und Gebrauch eines Werkzeugs strukturiert. Ebenso werden die gesamte Anordnung und das übergreifende Zusammenspiel von didaktischer Hardware, wie den Artefakten, Anleitungen dazu, technischen Dokumenten und didaktischer Software als Lernsequenz, als definierende Komponente der instrumentellen Orchestrierung miteinbezogen.

Von Trouche wird immer wieder betont, dass die instrumentelle Genese Lernender gelenkt werden sollte. Die instrumentelle Orchestrierung ist dafür der theoretische Rahmen.

> „I introduce the term instrumental orchestration to point out the necessity (for a given institution – a teacher in her/his class, for example) of external steering of students' instrumental genesis.” (Trouche 2004, 296)

Trouche zieht in die instrumentelle Orchestrierung die gesamte „*environmental organization*” (ebd.) mit ein. Herausgestellt wird hierbei, dass mit Hilfe dieses Ansatzes die gesamte Lernumgebung und -situation bei der Erstellung von Aufgaben einbezogen werden muss.

Wie bereits oben erwähnt, ist die Ausbildung von Schemata zentral bei der instrumentellen Genese von Lehrenden. Dabei handelt es sich sowohl um einen individuellen, als auch um einen sozialen Prozess: „schemes are elaborated and shared in communities of practice and may give rise to an appropriation by subjects, or even result from explicit training processes” (Rabardel & Samurcay 2001, 20). Diese expliziten Trainingsprozesse werden durch die instrumentelle Orchestrierung des Lehrenden für eine Gruppe von Lernenden gelenkt.

Die instrumentelle Orchestrierung beinhaltet nach Trouche (2004) zwei wesentliche Elemente: die didaktische Konfiguration und deren Nutzungsweise im Unterricht. Diese beiden Punkte werden von Drijvers, Doorman, Boon und van Gisbergen (2010) um den weiteren Punkt der konkreten didaktischen Umsetzung, ergänzt.

1. Die didaktische Konfiguration („*didactical configuration*“) ist die Zusammenstellung der Artefakte in der Umgebung bzw. die Konfiguration

der Lehrsituation mit den involvierten Artefakten. Sowohl technologische Werkzeuge als auch Aufgaben können als Artefakte angesehen werden.
In der Welt der Musik kann die didaktische Konfiguration mit der Auswahl und Positionierung im Raum der einzelnen Instrumente eines Orchesters verglichen werden. (Trouche 2004, 296)

2. Eine Nutzungsweise („*exploitation mode*“) der didaktischen Konfiguration ist die Art und Weise, auf die der Lehrer die didaktische Konfiguration umsetzt. Hierin sind Entscheidungen über die Einführung einer Aufgabe, über das Arbeiten an einer Aufgabe, über die Rolle der Artefakte und über die auszubildenden Schemata und Abläufen eingeschlossen.
In der Welt der Musik kann dies mit dem Festlegen der Partitur für jedes Instrument verglichen werden im Hinblick auf die später entstehenden Harmonien. (Trouche 2004, 296)
3. Die didaktische Umsetzung („*didactical performance*“) beinhaltet alle ad hoc Entscheidungen, die während des Unterrichtens getroffen werden müssen. Welche Frage muss als nächstes gestellt werden, wie kann auf einen Schülerbeitrag eingegangen werden, wie wird mit einem unerwarteten technologischen oder mathematischen Aspekt umgegangen?
In der Welt der Musik kann die didaktische Umsetzung mit dem musikalischen Auftritt verglichen werden. (Drijvers, Doorman, Boon & van Gisbergen 2010, 1350f)

Drijvers, Doorman, Boon und van Gisbergen (2010) ergänzen den Ansatz von Trouche um den weiteren Punkt *didactical performance*. Dies ist aus der konkreten Anwendung der instrumentellen Orchestrierung nach Trouche entstanden, als es darum ging, Lehrer beim Unterrichten zu analysieren. Von diesen Autoren stammt die folgende genauere Charakterisierung:

> „An instrumental orchestration is defined as the intentional and systematic organisation and use of the various artefacts available in an – in our case computerised – learning environment by the teacher in a given mathematical task situation, in order to guide students‘ instrumental genesis.” (Drijvers, Doorman, Boon & van Gisbergen 2010, 1350f)

Die instrumentelle Orchestrierung wird, wie jeder andere Unterricht auch, teilweise vorbereitet, teilweise jedoch spontan durchgeführt in der jeweiligen Situation, weshalb der letzte Punkt von den Autoren ergänzt wurde. Das Ziel der instrumentellen Orchestrierung ist es, die instrumentelle Genese von Lernenden zu verbessern und zu lenken (Drijvers, Doorman, Boon & van Gisbergen 2010, 1351).

Einen allgemeinen Überblick über verschiedene Arten der Orchestrierung geben Drijvers, et al. (2010c). Dabei erfolgt die Einordnung bezüglich der Punkte *Didaktische Konfiguration* und *Nutzungsweise*, betrifft also die Planung des Unterrichts. Insgesamt

identifizieren die Autoren sechs Arten der Orchestrierung und bezeichnen diese wie folgt: „*Technical-demo*, *Explain-the-screen*, *Link-screen-board*, *Discuss-the-screen*, *Spot-and-show*, and *Sherpa-at-work*“ (Drijvers, Doorman, Boon, Reed & Gravemeijer 2010, 219). Dabei unterscheiden sich diese Orchestrierungsarten in Bezug auf ihren Fokus entweder auf den Lehrenden oder auf den Lernenden und den Grad an zusätzlichem Technologieeinsatz.

Die *Technical-demo* Orchestrierung ist im Wesentlichen die Demonstration einzelner Werkzeugtechniken durch den Lehrenden. Die didaktische Konfiguration hierzu beinhaltet Zugriff auf die Werkzeugsoftware, auf einen Beamer und eine Lernsituation, die es den Lernenden erlaubt, der Demonstration zu folgen. Als Nutzungsweise bietet sich eine neue Situation oder Aufgabe an, oder es wird die Arbeit eines Lernenden zu Demonstrationszwecken vorgestellt. Dies ist eine Orchestrierung, die in der Analyse der Autoren häufig zu Beginn von Unterrichtssequenzen gefunden wurde, da hierin über den Lehrenden Einführungen und Erklärungen zur Software stattfinden.

Die folgenden beiden Orchestrierungen *Explain-the-screen* und *Discuss-the-screen* identifizieren die Autoren als technologische Varianten herkömmlichen Unterrichts. Befragte Lehrende geben dazu an, dass sie mit diesen Formen der Orchestrierung aus ihrer Lehrerfahrung vertraut sind.

Die *Explain-the-screen* Orchestrierung ist ebenfalls eine lehrerzentrierte Demonstration des Computerbildschirms. Diese Erklärung geht jedoch über technische Aspekte hinaus und beinhaltet mathematischen Kontext. Die didaktische Konfiguration ist ähnlich zu *technical-demo*. Als Ausgangspunkt für die Nutzungsweise kann die Arbeit eines Lernenden verwendet werden oder eine komplett neue Lösungsstrategie einer Aufgabe.

Die *Discuss-the-screen* Orchestrierung umfasst eine Diskussion darüber, was auf dem Computerbildschirm passiert. Das Ziel der Diskussion ist es, die kollektive instrumentelle Genese zu verbessern. Die didaktische Konfiguration ist wie oben. Für die Nutzungsweise bietet sich wahlweise die Arbeit eines Lernenden oder eine Aufgabe, ein Problem oder ein Zugang durch den Lehrenden an.

Sowohl *Technical-demo* als auch die folgende *Link-screen-board* Orchestrierung sind technologiespezifisch und dadurch charakteristisch für Unterricht mit einem hohen Grad an Technologieeinsatz.

In der *Link-screen-board* Orchestrierung betont der Lehrende die Beziehung zwischen dem, was in der Technologieumgebung passiert mit dem, was durch konventionelle Artefakte wie Papier, Buch oder Tafel repräsentiert ist. Zusätzlich zu Software und Beamer muss für die didaktische Konfiguration beispielsweise die Tafel sichtbar sein. Die Nutzungsweise kann aussehen wie bei *explain-the-screen*.

Im Wesentlichen lernendenzentriert und ebenfalls spezifisch für Unterricht mit Technologieeinsatz sind die Orchestrierungen *Spot-and-show* und *Sherpa-at-work* (s. u.).

In einer weiteren Studie wird diese Liste um eine zusätzliche Art der Orchestrierung ergänzt und gleichzeitig eingeräumt, dass die Liste nicht vollständig sei:

> „As a new orchestration type, the Work-and-walk-by orchestration was identified. We expect the list of possible orchestrations to be extended in future, not as to strive for a complete list, but as to provide teachers with a diverse repertoire of possible orchestrations as source of inspiration to their professional activity." (Drijvers 2012, 278)

Die *Work-and-walk-by* Orchestrierung ist auf den einzelnen Lernenden zentriert und tritt häufig in Arbeitsphasen auf. Der Lehrende beobachtet die einzelnen Lernenden und hilft oder erklärt individuell. Als didaktische Konfiguration können hier auch die Tafel oder weitere Hilfsmittel eingesetzt werden, der Fokus bleibt jedoch auf dem einzelnen Lernenden und richtet sich nicht an die gesamte Gruppe.

2.2.2.1 Konkrete Gestaltungsvorschläge für Unterricht (exploitation mode)

Im Zusammenhang mit dem Einsatz von graphikfähigen Taschenrechnern beschäftigen sich Guin & Trouche bereits 1999 mit der instrumentellen Genese bei Lernenden aus didaktischer Sicht. Am Ende ihres Artikels kommen sie zu Vorschlägen, wie Unterricht, gerichtet auf instrumentelle Genese, aufgebaut werden sollte. Dabei kommt der Rolle des Lehrenden und von ihm ausgehend der Gestaltung der einzelnen Unterrichtsphasen eine fundamentale Rolle zu. Sie betonen, dass sowohl auf mathematischer, als auch auf technologischer Ebene der Lehrende einen guten Einblick in den Inhalt und die Werkzeuge haben muss, um bei Lernenden eine zielgerichtete instrumentelle Genese anzustoßen.

> „Therefore, we argue for strong teacher involvement in the instrumentation process and full recognition of the constraints and potential of the artifact as well as various profiles of student behaviour so as to design and implement appropriate mathematical activities." (Guin & Trouche 1999, 224)

Sie schlagen vor, die Unterrichtssituation so zu organisieren, dass den Schülern viel Zeit für eigene, praktische Untersuchungen mit Hilfe des Instruments bleibt.

Folgende allgemeine Vorschläge für die Unterrichtsgestaltung haben sie:

- In jeder Aktivität sollte nur eine begrenzte Anzahl neuer Befehle/Funktionen eingeführt werden, damit die mathematische Arbeit nicht verdeckt wird und Vermeidungsstrategien seitens der Lernenden möglichst limitiert werden.
- Es sollte genug Zeit zur Verfügung stehen, damit verschiedene Repräsentationen und Strategien ausprobiert werden können, und um eine passende Sprache aufbauen zu können.

- Zwischenfragen bezüglich mathematischer Inhalte sollten ermöglicht und begünstigt werden, um einer zu starken Abhängigkeit vom Instrument zugunsten des mathematischen Verstehens vorzubeugen.
- Effiziente Strategien und Techniken sollten mittels Projektor gezeigt und besprochen werden.
- Größere Zusammenhänge zum Curriculum sollten aufgezeigt werden. (Guin & Trouche 1999, 224).

Die instrumentelle Genese läuft bei jedem Subjekt individuell verschieden ab, allerdings besitzt sie auch eine soziale Komponente. Das Arbeiten mit einem Werkzeug findet generell häufig in einem sozialen Umfeld (z. B. Schule) statt. Gerade dem Lehrer kommt dabei eine zentrale Stellung zu (s. o. und Guin & Trouche 2002, 208). Der Ansatzpunkt der instrumentellen Orchestrierung ist in der Tat die Sozialisierung der instrumentellen Genese der Lernenden mit dem Ziel, möglichst ähnliche Schemata zum Bearbeiten einer Aufgabe zu erzeugen. Da sich die einzelnen Bearbeitungen dennoch unterscheiden durch den individuellen Prozess der *instrumentalization* ist eine solche Normierung des Wissens durch den Lehrenden nötig.

Guin und Trouche (2002) nennen drei Phasen, in denen der *instrumentalization process* ablaufen kann: „a stage of discovery and selection of the relevant keys, a stage of personalization (one fits the tool to one's hand) and a stage of transformation of the artifact, sometimes in directions unplanned by the designer (…). The instrumentalization is also a process of differentiation of the artifacts themselves." (Guin & Trouche 2002, 205).

Die instrumentelle Genese hat dadurch individuelle und soziale Komponenten. Der *instrumentalization process* befördert die individuellen Komponenten. Zusammengeführt werden kann diese Individualität durch den *instrumentation process*. Guin und Trouche (2002) umschreiben diesen Prozess als Sammlung von Gesten als sichtbare Komponenten der herausgebildeten Schemata, die ein Subjekt zum Erreichen einer bestimmten Aufgabe herausbildet. Mehrere solche Gesten, die zu einem bestimmten Schema gehören, fassen sie zu sogenannten Techniken zusammen. Im Gegensatz zu einem Schema kann eine solche Technik verbal beschrieben und damit auch als „*instrumented technique*" (Guin & Trouche, 2002, S. 206) zentral durch den Lehrenden gelehrt werden. Um den sozialen Teil der instrumentellen Genese der Lernenden zu verstärken, sollten neue Bearbeitungsweisen von Aufgaben deshalb zentral durch den Lehrer gezeigt werden.

Eine Quelle der Sozialisierung der instrumentellen Genese sehen die Autoren jedoch in einer besonderen didaktischen Situation: Der Orchestrierung der Lernsituation mit einem „Sherpa-Schüler", dessen Arbeit ein Vorbild und per Beamer für alle sichtbar ist.

> „[The sherpa-student] will thus be used, for both class and teacher, as reference, a guide, an auxiliary and a mediator. This orchestration favors collective management of a part of the instrumentation and instrumentalization processes: what a student does with her/his calculator, - traces of her/his activity – is seen by all, this allows to compare different instrumented techniques and give information to the teacher on the instrumented actions schemes being built by the sherpa-student." (Guin & Trouche 2002, 209).

Dieser Sherpa-Schüler hat eine zentrale Bedeutung in der Arbeit von Guin und Trouche. Er ist Teil der Gruppe der Lernenden und kann mehrere Funktionen innehaben, sowohl für den Lehrer als auch für die restlichen Lernenden. In verschiedenen Arrangements ist er Referenz, Vorbild, Hilfslehrer oder Mediator. Die jeweilige Orchestrierung der Unterrichtssituation gibt für alle Lernenden Einblicke in den beim Sherpa-Schüler ablaufenden *instrumentalization process* und den *instrumentation process* und soll so diese Prozesse bei allen anderen wirkungsvoll unterstützen oder voranbringen.

Drijvers, Doorman, Boon, Reed, et al. (2010) identifizieren als zwei weitere Orchestrierungsarten (vgl. S. 214) die *Sherpa-at-work* und *Spot-and-show* Orchestrierung, durch die die instrumentelle Genese der Lernenden besonders gelenkt wird. Bei der *Sherpa-at-work* Orchestrierung präsentiert der so genannte Sherpa-Schüler seine eigene Arbeit oder führt verschiedene, durch den Lehrenden genannte, Aktionen aus. In der Nutzungsweise kann der Lehrende durch Fragen oder Anweisungen in die Präsentation des Sherpa-Schülers eingreifen. In der *Spot-and-show* Orchestrierung steht die interessante Aufgabenbearbeitung eines Lernenden im Vordergrund der Demonstration und Diskussion (vgl. Drijvers, Doorman, Boon, Reed, et al. 2010, 219).

Die *exploitation mode* bezogen auf die Integration des Sherpa-Schülers in einer Lehrsituation (in einer Arbeitsphase) kann durch den Lehrer auf unterschiedliche Arten organisiert werden.

- Stark gelenkt: Wenn der Sherpa-Schüler etwas zeigt (z. B. über einen Projektor), das alle nacharbeiten sollen, dann hat das Auswirkungen auf den *instrumentalization* und den *instrumentation process* aller Lernenden. Beide Prozesse sind dann starkem Zwang unterworfen und laufen bei allen Anwesenden möglichst ähnlich ab.
- Relativ gelenkt: Der Sherpa-Schüler dient als Beispiel, seine Arbeit ist durch einen Projektor für alle sichtbar, aber nicht verbindlich zu kopieren. Die beiden Prozesse laufen dann unter weniger starken Bedingungen ab.
- Kaum gelenkt: Die Arbeit des Sherpa-Schülers ist für alle anderen nicht sichtbar, die Prozesse laufen unter kaum kontrollierten Bedingungen ab und sind wenig vergleichbar. (Trouche 2004, 299)

Weitere didaktische Entscheidungen müssen vom Lehrenden getroffen werden. Zum Beispiel ob es nur einen Sherpa-Schüler gibt, oder ob diese Rolle wechselt, ob der

Sherpa-Schüler nach vorne kommt oder an seinem Platz bleibt, etc. Durch diese spezielle Orchestrierung ergeben sich einige Vorteile, unter anderem können lernschwächere Schüler stärker gefördert werden, weil sie sich eher mit einem Mitschüler als mit dem Lehrer identifizieren. Der Sherpa-Schüler kann auch ein Mitschüler in einer Lerngruppe sein, dann ist die Funktion wie im letzten Spiegelpunkt angedeutet.

Kurz zusammengefasst beschreibt die instrumentelle Orchestrierung, wie ein Lehrender die instrumentelle Genese von Lernenden managt. Wie sieht es nun aber auf der Seite des Lehrenden aus?

2.2.2.2 Die instrumentelle Genese des Lehrenden

Der sinnvolle Einsatz eines Instruments in einer didaktischen Situation hängt von verschiedenen Bedingungen ab. Um als Lehrender ein Instrument didaktisch nutzen zu können, muss es zunächst beherrscht werden, in dem Sinne, dass durch die instrumentelle Genese der Lehrende sich das Artefakt zu eigen macht, sprich als Instrument nutzen kann. Das Wissen über Möglichkeiten und Grenzen des zugrunde liegenden Artefakts ermöglicht erst den sinnvollen Einsatz des Instruments in Lehrsituationen.

Ein Instrument muss auf zwei Ebenen klar differenziert werden:

- Die Ebene eines pädagogischen Instruments, bei der das Subjekt der Lehrer ist und das Objekt das spezielle durch den Lehrer definierte zu erreichende Lernziel.
- Die Ebene des praktischen Instruments, bei der das Subjekt ein beliebiger Benutzer ist, besonders ein Schüler. (Artigue et al. 2006, 15)

Im Fall der ersten Ebene gibt es zwei Arten der instrumentellen Genese beim Lehrenden: Zum einen muss sich der Lehrende das Artefakt zu eigen machen um es wie ein Lernender als Instrument nutzen zu können. Zum anderen muss sich der Lehrende das Artefakt als Lehrinstrument aneignen, mit dem er bestimmte Lernziele erreichen möchte.

Nach einer Studie von Bueno-Ravel und Gueudet (2007) machen sich Lehrende als Lernende als erstes vertraut mit einem Artefakt und untersuchen dessen Möglichkeiten. Im nächsten Schritt arbeiten sie sorgfältig ihre eigenen Möglichkeiten und Nutzungsweisen aus, die durchaus von den intendierten Ideen des Artefakt-Entwicklers abweichen. So machen sich die fünf Lehrenden der Studie das Artefakt EEB (e-exercises bases, dieses besteht aus verschiedenen Übungen, die nach dem mathematischen Inhalt und/oder dem Schwierigkeitsgrad klassifiziert werden und bei denen verschiedene mathematische Werkzeuge zum Einsatz kommen) zu eigen als Lehrinstrument (Bueno-Ravel & Gueudet 2007).

Drijvers und Trouche (2008) arbeiten heraus, dass bei der Aneignung eines Artefakts als Instrument der Lehrende verschiedene Ressourcen nutzen, beziehungsweise kreieren muss. Eine solche Ressource kann beispielsweise das Internet sein, eine Ressource ist das Artefakt als Instrument um einen bestimmten Aufgabentyp zu bearbeiten, eine weitere Ressource ist das gleiche Artefakt als Instrument um einen anderen Aufgabentyp zu bearbeiten. Zusätzlich sollte eine solche Ressource benutzt oder erarbeitet werden, die in Form von Schaubildern, Diagrammen o. ä. die Lösungsstrategien und Arbeitsmethoden für bestimmte Aufgaben dokumentiert (siehe auch Guin, Ruthven & Trouche 2005). Diese Sammlung verschiedener Ressourcen bleibt dabei offen für Neuzugänge und sollte im Rahmen der instrumentellen Genese des Lehrenden laufend ergänzt werden.

Im zweiten Teil dieser Arbeit wird eine Lehrveranstaltung vorgestellt (Kapitel 8.1), in der die Software TinkerPlots im Sinne der instrumentellen Genese eingesetzt wird. Diese Lehrveranstaltung wurde für Lehramtsstudierende entwickelt. In diesem Zusammenhang werden die Studierenden jedoch als Lernende betrachtet, der Aspekt der speziellen „doppelten“ instrumentellen Genese bei Lehrenden wird deshalb hier nicht weiter ausgearbeitet. Weiterführende Literatur dazu findet sich beispielsweise in Drijvers, Kieran, et al. (2010), bei Gueudet, Pepin und Trouche (2012) oder bei Gueudet und Trouche (2009).

Die Notwendigkeit, Lernsequenzen zu designen im Hinblick auf eine zielgerichtete soziale instrumentelle Genese der Lernenden, beschreiben diverse Autoren.

> „Instrumental orchestrations, integral parts of a didactic exploitation system are plans of actions, allowing to guide students' instrumented action. These obviously require some didactic engineering – arranged sequences of carefully thought-out situations aiming at reaching mathematics-learning objectives (…).” (Guin & Trouche 2002, 210)

> „In the design of the learning sequence, the issue of orchestration is dealt with by developing a teacher guide, in which didactical scenarios are described that suggest working arrangements as well as possible ways to use them” (Drijvers et al. 2007, 2)

An diesen zwei Beispielen zeigt sich, dass ein Durchplanen und eine genaue Beschreibung der Lernsequenzen im Sinne der instrumentellen Orchestrierung als Startpunkt für die instrumentelle Genese von Lernenden steht. Aus Lehrerperspektive geht es auch darum, „die Handhabung, Einflüsse auf kognitive Entwicklungen, Möglichkeiten der Anwender und zukünftige Softwarekonzeptionen besser beurteilen zu können. Mithilfe dieser Erkenntnisse lassen sich im Idealfall verbesserte Lernumgebungen generieren“ (Hattermann 2011, 1). Auch aus stochastisch-didaktischer Sicht ist ein genaues Analysieren eines Simulations-Werkzeugs für den Unterricht sinnvoll: „But the status of a computer simulation in class must be analysed, (…) and its didactical pertinence must be discussed“ (Chaput et al. 2011, 92). Und in Bezug auf die Software TinkerPlots findet sich: „Because of the diversity of the constraints in TinkerPlots it is important for

teachers to be aware of their potential either to support or hinder students' developing understanding. Again the pedagogical content knowledge of the teacher is built on knowing well both the affordances and constraints of the software." (Watson & Fitzallen 2016, 589). Die hier ausgeführte instrumentelle Orchestrierung wird deshalb mit der genauen Analyse der Software (Kapitel 3 und 4) als Grundlage für die Gestaltung der Lernsequenzen genommen, die in der Lehrveranstaltung eingesetzt werden (Kapitel 8).

2.3 Ein Modell des Wissenserwerbs: Cognitive Load Theory

Wenn Wissen erworben wird, zum Beispiel beim Umgang mit Software, so wird das Arbeitsgedächtnis belastet. Einen Erklärungsansatz hierzu bietet die *Cognitive-Load-Theory* (Chandler und Sweller 1991) Die Cognitive-Load-Theory (kurz CLT) geht von zwei Grundannahmen aus.

Die erste Grundannahme lautet, dass die Kapazität des Arbeitsgedächtnisses beschränkt ist (nach Baddeley 1992). Wenn neue Informationen aufgenommen werden, so werden nur bestimmte Aspekte davon gleichzeitig im Arbeitsgedächtnis bearbeitet. Die zweite Grundannahme lautet, dass Wissen in Schemata abgespeichert wird, so wie dies auch im vorherigen Abschnitt bereits ausgeführt wurde. Beim Lernen werden sowohl neue Schemata konstruiert als auch neues Wissen mit bereits vorhandenen Schemata verknüpft. Das Arbeitsgedächtnis muss genügend Kapazität haben und darf kognitiv nicht überlastet werden, damit effektives Lernen möglich ist.

Durch die CLT werden drei unterschiedliche Arten von Belastung definiert (nach Sweller, van Merrienboer & Paas 1998):

- Intrinsische kognitive Belastung: Diese wird durch die Schwierigkeit und Komplexität des Lernmaterials und das Vorwissen des Lerners bestimmt;
- Extrinsische kognitive Belastung: Diese wird durch die Gestaltung des Lernmaterials bestimmt;
- Lernbezogene Belastung: Diese wird durch den Aufwand des Lernenden, das Lernmaterial zu verstehen, bestimmt.

Beim Lernen sind die intrinsische und die extrinsische kognitive Belastung möglichst gering zu halten, damit für die lernbezogene Belastung genügend Kapazität bleibt. Für die Gestaltung von Lernumgebungen ist demnach darauf zu achten, dass die extrinsische Belastung möglichst gering gehalten wird, da auf diese Belastung unmittelbar Einfluss genommen werden kann.

Beim Lernen mit Software geschieht sowohl Lernen von stochastischen Inhalten, als auch prozessbezogenes Lernen mit der Software, dies erhöht die intrinsische kognitive

Belastung. Eine Möglichkeit, die extrinsische kognitive Belastung zu verringern, besteht darin, das prozessbezogene Lernen zu unterstützen, indem hierzu eine Struktur, beispielsweise in Form eines Plans (siehe nächster Abschnitt) vorgegeben wird. Eine andere Möglichkeit ist durch die Bereitstellung von ausgearbeiteten Beispielen gegeben (Renkl 2002), die ebenfalls die extrinsische kognitive Belastung verringern können.

2.4 Der Simulationsplan als didaktisches Konzept

Es hat sich gezeigt, dass ein Simulationsplan Lernende nachhaltig beim Simulationsprozess unterstützen kann (Hofmann 2012, Maxara & Biehler 2006, Prömmel 2013) und eine Möglichkeit darstellt, den *cognitive load* zu verringern. Ein allgemeiner Simulationsplan für stochastische Simulationen wurde von Maxara (2006) entwickelt. Einen speziellen 6-Schritte-Plan zur Simulation für die Verbindung der stochastischen Ebene und der „Simulationsebene“ wurde im weiteren ebenfalls von Maxara erstellt (2009), der später von Hofmann (2012) zu einem graphischen Simulationsplan weiterentwickelt wurde.

Maxara (2009) hat vier Schnittstellen zwischen einem Zufallsexperiment und einer Simulation identifiziert, die sie als wesentlich für das erfolgreiche Durchführen einer Computersimulation charakterisiert. Diese Schnittstellen werden von ihr „Interface“ genannt und sind

„1. Modellierungsinterface
2. Ereignisse/Zufallsgrößen-Interface
3. Wiederholungs+Abbruchbedingungs-Interface
4. Auswertungsinterface“ (Maxara 2009, 39)

Bezüglich dieser Interfaces soll eine Software Unterstützungspotential bieten und an diesen Stellen soll ein Simulationsplan ansetzen (ebd.).

Es lassen sich drei Einsatzmöglichkeiten eines Simulationsplans unterscheiden:

Der Simulationsplan zur Planung einer Simulation

Soll der Simulationsplan zum Planen einer Simulation eingesetzt werden, so ist er vor dem Beginn des Simulierens auszufüllen. Bei diesem Einsatz liegt das Ziel darin, dass im Vorfeld, d. h. vor der eigentlichen Arbeit am PC, die Simulation geplant werden muss. Dies erfordert eine gute Kenntnis der Funktionen der Software.

Der Simulationsplan zur Prozessunterstützung bei einer Simulation

Der Einsatz des Simulationsplans zur Prozessunterstützung bei einer Simulation geschieht während der Simulation. Parallel zur Simulation kann der Simulationsplan als Orientierungshilfe genutzt werden, welcher Schritt als nächstes durchzuführen ist und

dieser Schritt kann parallel im Simulationsplan notiert werden. Das Ziel hierbei ist die Vorgabe eines roten Fadens durch den Plan für die Simulation.

Der Simulationsplan zur Dokumentation einer Simulation

Die dritte Möglichkeit ist es, den Simulationsplan zur Dokumentation einzusetzen. Dies geschieht nach einer durchgeführten Simulation. Die einzelnen Schritte und Ergebnisse werden im Simulationsplan notiert. Das Ziel liegt hierbei zum einen auf einer ordentlichen Offline-Dokumentation der Simulation und zum anderen auf der Förderung der verbalen Beschreibungskompetenz einer Simulation. Zusätzlich regt die schriftliche Dokumentation einer Simulation erneut zur Auseinandersetzung mit den Inhalten an und verhilft somit dazu, die einzelnen Schritte genauer zu reflektieren. Weiterhin kann ein ausgefüllter Simulationsplan später als „worked example“ (Renkl 2002) eingesetzt werden, um eine ähnliche Simulation durchzuführen.

Darauf aufbauend wurde für die Simulation mit der Software TinkerPlots (deutsche Version) von der Autorin dieser Arbeit ein eigener Simulationsplan entwickelt, der speziell auf diese Software zugeschnitten ist. Dieser graphische Simulationsplan (siehe Abbildung 3) ist in den vergangenen Jahren im Rahmen verschiedener universitärer Lehrveranstaltungen weiterentwickelt und erprobt worden (Podworny 2013, Podworny & Biehler 2014).

Der Simulationsplan basiert auf der graphischen Verknüpfung der TinkerPlots Elemente mit auszufüllenden Textfeldern. Jedes TinkerPlots-Element, das zur Simulation benötigt wird, ist graphisch im Simulationsplan repräsentiert. So kann direkt eingezeichnet werden, wie die Simulation in TinkerPlots durchgeführt wird. Zusätzliche Erläuterungen bzw. Verbalisierungen des Simulationsprozesses sollen in Textfeldern geschehen. Eine Nummerierung bestimmter Funktionen hilft als Orientierung und Verweismöglichkeit.

Der in Abbildung 3 gezeigte Simulationsplan für TinkerPlots unterstützt die von Maxara (2009) identifizierten Interfaces für eine Simulation mit TinkerPlots und stellt somit eine wichtige Ressource im Sinne von (Guin et al. 2005) für den Prozess der instrumentellen Genese eines Lernenden dar.

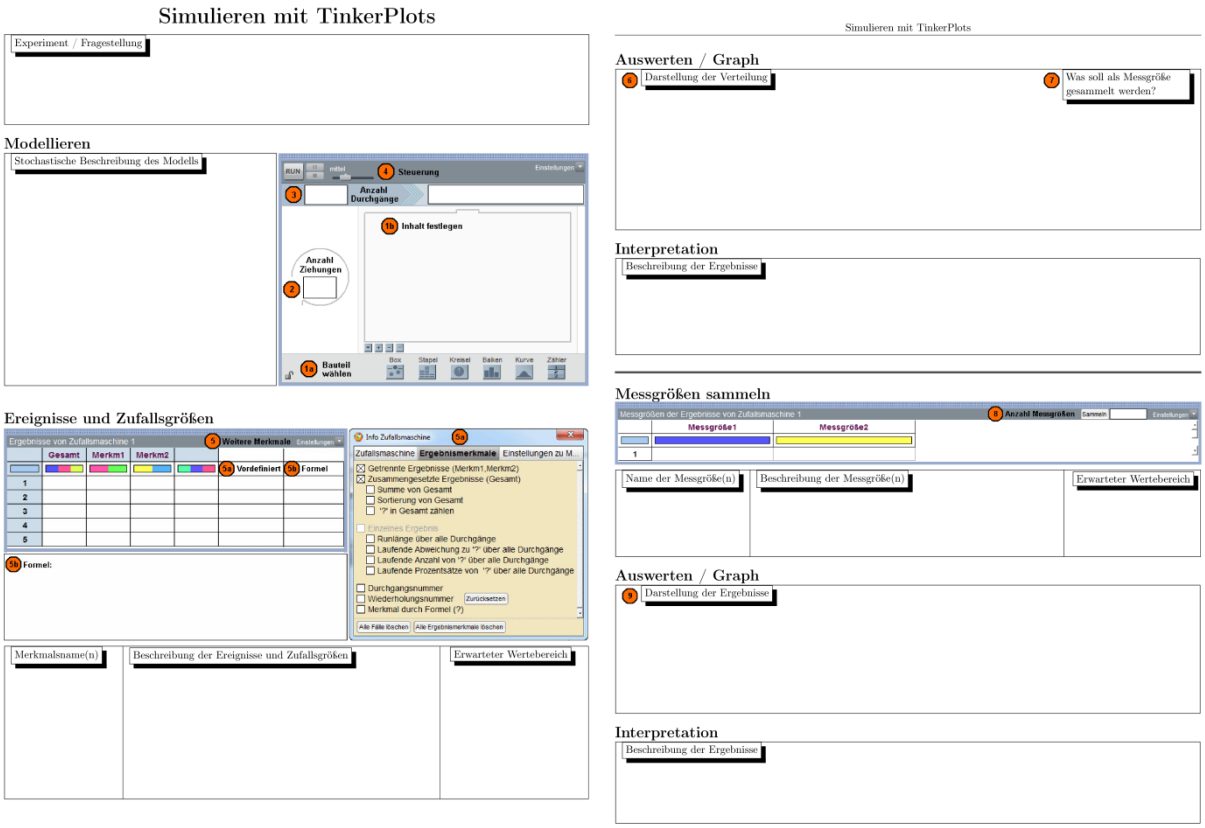


Abbildung 3: Simulationsplan für TinkerPlots

3 Theoretische Werkzeuganalyse der Software TinkerPlots

In diesem Kapitel wird das folgende Ziel verfolgt:

Z1: Analyse der Software TinkerPlots bezüglich des Unterstützungspotentials zur Konstruktion stochastischer Simulationen.

Dabei soll untersucht werden, inwieweit TinkerPlots die Umsetzung der Schritte *Modellierung, Ereignisse und Zufallsgrößen, Wiederholung* und *Auswertung* in intuitiver Weise unterstützt, welche Möglichkeiten der Interfacegestaltung bereitgestellt werden und welche Aspekte nicht oder nur in geringem Maß unterstütz werden.

Am Ende jeden Abschnitts findet sich eine Zusammenfassung und Bewertung hinsichtlich des jeweiligen Interfaces.

3.1 Ziele und Methoden der Werkzeuganalyse

Die sinnvolle Nutzung eines Werkzeugs wird gefördert durch eine gute Kenntnis der Eigenschaften und Grenzen desselben. Damit eine zielgerichtete instrumentelle Genese (siehe Kapitel 2.2) bei Lernenden und Lehrenden vonstatten gehen kann, ist das genaue Herausarbeiten und Wissen dieser Bedingungen des Instruments vonnöten (z. B. Artigue 2002, Drijvers, Doorman, Boon, Reed, et al. 2010, Guin & Trouche 2002, Rabardel 2002, Trouche 2004). Aus der Kenntnis der Möglichkeiten und Grenzen eines Instruments erwachsen unterschiedliche Möglichkeiten an Problemstellungen heranzugehen. Es können Aufgaben des unterschiedlichsten Typs mit Hilfe eines Instruments sinnvoll bearbeitet werden (siehe dazu auch Kapitel 4). Die Eigenschaften und Grenzen eines Instruments herauszuarbeiten, bedeutet bei Software eine genaue Werkzeuganalyse durchzuführen. Eine solche Analyse ist sinnvoll, um Software als mathematisches Instrument nutzbar zu machen und bei Lernenden die stochastischen Inhalte mit den technischen zu verknüpfen. Lehrende können aufbauend auf diesem Wissen Aufgaben und Lernumgebungen erstellen oder nutzen, die diese Verknüpfung sinnvoll unterstützen und die technische Inhalte nicht über stochastische stellen.

Im vorangegangenen Kapitel wurde der Simulationsplan als didaktisches Konzept vorgestellt (vgl. Abschnitt 2.4). Die dort definierten Interfaces *Modellierung, Ereignisse und Zufallsgrößen, Wiederholung (und Abbruchbedingung)* und *Auswertung* bilden die Struktur für die Werkzeuganalyse. Wie bereits Maxara (2009, 55) bemerkt, kann der universelle Simulationsplan als Anwendungsschema für die Bewertung der einzelnen Interfaces genutzt werden, auch wenn er nicht für alle Simulationen von Zufallsexperimenten in dieser Reihenfolge abgearbeitet werden kann. Durch die Besonderheiten des Programms TinkerPlots wird bei der Umsetzung von Zufallsexperimenten in der Software direkt bei der Modellierung das Interface Wiederholen benutzt. Aus diesem Grund

wird das Interface Wiederholen in das Interface Modellierung integriert. Es wird analysiert, inwieweit TinkerPlots bei der Realisierung der einzelnen Interfaces unterstützt und welche Möglichkeiten und Grenzen in der Benutzung der Software liegen. Weiterhin wird darauf geachtet, welche Aspekte mit TinkerPlots kaum oder gar nicht realisiert werden können. Aus der Analyse der Möglichkeiten, die TinkerPlots bezüglich der einzelnen Interfaces bereitstellt, ergeben sich verschiedene Arten der Simulation. Diese werden in Kapitel 3.5 detailliert dargestellt und an Beispielen erläutert.

3.2 Kurzüberblick über die Software TinkerPlots

TinkerPlots ist eine Software, die zwei große Bereiche der Stochastik abdeckt: erstens die dynamische Datenanalyse und zweitens die Simulation von Wahrscheinlichkeitsprozessen. Es handelt sich um eine Software, die speziell zum Lernen dieser beiden Bereiche entwickelt wurde.[6] Sie wurde in den USA nach dem „bottom-up“-Prinzip entwickelt, „which takes into account not only where we want students to end up, but also where they are coming from” (Konold 2007, 7). Es handelt sich um eine Software, die laut den Entwicklern Cliff Konold und Craig Miller ab der Schulklasse drei eingesetzt werden kann und nicht um die heruntergebrochene („top-down“) Version einer professionellen Software (ebd.). Das heißt, der Gebrauch der Software knüpft an wenig Vorwissen an und erfordert unter anderem keine Programmierkenntnisse. Ein Vorteil von TinkerPlots ist die kurze Einarbeitungszeit in die wesentlichen Objekte. Im Sinne von Bakker handelt es sich um einen „landscape-type of software“ (Bakker 2002), in der sich Anwender in einer Lernlandschaft weitgehend frei bewegen können. Die Idee ist, dass Lernende spielerisch erste Erfahrungen mit Datenanalyse und Simulationen machen können und in ihren Lernprozessen unterstützt werden, um später gezielt stochastische Aufgaben bearbeiten zu können. Die Software gibt es seit 2011 als Version 2.0 mit dem Simulationsmodul. Eine deutsche Version liegt seit 2016 vor und kann unter www.tinkerplots.com/get (abgerufen am 27.06.2018) heruntergeladen werden.

3.2.1 Software-Ergonomie

Die Software TinkerPlots bietet darüber eine intuitive Bedienung bei einer nutzerfreundlichen Oberfläche. Dazu werden nun einige Grundlagen der Software-Ergonomie betrachtet. In der internationalen Norm DIN EN ISO 9241[7] sind Standards der Mensch-Computer-Interaktion festgelegt. Besonders ISO 9241 Teil 11 befasst sich mit Anforderungen an Gebrauchstauglichkeit von Software, welche insbesondere von der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin ausgearbeitet ist (Bauer & Röckmann

---

[6] TinkerPlots ist wesentlich auf Grundlagen der Software Fathom aufgebaut und für jüngere Schüler in der Bedienung und Oberfläche weiterentwickelt.

[7] DIN EN ISO bedeutet, dass es eine deutsche, europäische und international gültige Norm darstellt, im Folgenden wird die Kurzbezeichnung ISO verwendet.

1999). Darin festgelegt sind drei Leitkriterien: *Effektivität*, *Effizienz* und *Zufriedenheit* des Anwenders. Das bedeutet, der Benutzer kann mit der Software seine Aufgaben erledigen (Effektivität), kann dies mit minimalem oder angemessenem Aufwand tun (Effizienz) und ärgert sich nicht über die Software, sondern hat bestenfalls Freude an ihrer Benutzung (Zufriedenheit) (Dahm 2006). Aufbauend auf der Definition der Gebrauchstauglichkeit nach ISO 9241 Teil 11, einschlägiger Literatur zur Software-Ergonomie und zum Mediendesign, untersucht Hartwig (2007) allgemein die Grundlagen von Software-Ergonomie und gibt umfassend Hinweise, die für die Gebrauchstauglichkeit von Software nützlich sind. Notwendige Voraussetzungen für den Erfolg einer jeglichen Software sind nach Hartwig eine einfache, verständliche Bedienung, Programmästhetik und die technische Qualität der Software, wozu unter anderem Zuverlässigkeit und Leistungsfähigkeit gezählt werden (Hartwig 2007, 2f). Bei TinkerPlots zeigt sich eine einfache Bedienung unter anderem daran, dass die meisten Aktionen über graphisch repräsentierte Objekte ausgeführt werden können und somit eine intuitive Bedienung ermöglicht wird. Eine weitere Besonderheit in der Bedienung von TinkerPlots ist, dass an keiner Stelle Variablennamen vergeben werden müssen, überall, wo Variablen eine Rolle spielen, erfolgt zunächst eine automatische Namensvergabe durch das Programm. Ein solch automatisch vergebener Name kann angepasst werden, notwendig ist dies im Gegensatz zu den meisten anderen Programmen jedoch nicht, somit wird eine fehlerhafte Eingabe vermieden. Darüber hinaus arbeitet Hartwig folgende Punkte aus, die technische Qualitätsmerkmale einer Software darstellen: *Korrektheit*, *Effizienz*, *Robustheit*, *Verfügbarkeit*, *Zuverlässigkeit*, *Datensicherheit*, *Verständlichkeit*, *Wartbarkeit* und *Wiederverwendbarkeit* (Hartwig 2007, 27; siehe auch Bauer & Röckmann 1999, Dahm 2006, Herczeg 2009). Hartwig räumt ein, dass eine Überprüfung dieser Merkmale an einer konkreten Software teilweise schwer nachzuweisen ist (Hartwig 2007, 28), soweit eine jahrelange Benutzung einer Software jedoch Rückschlüsse auf deren Qualität zulässt, werden diese Eigenschaften für TinkerPlots nun knapp überprüft.

Korrektheit ist eine wesentliche und notwendige Eigenschaft von Software, die die Übereinstimmung von computergenerierten Ergebnissen mit Ergebnissen im Kontext sicherstellt. Nur mit korrekt arbeitenden Systemen können Aufgaben gelöst werden. Es wird davon ausgegangen, dass etwa zehn Prozent Fehler im System verbleiben. Bezogen auf die Simulation mit TinkerPlots kann diese Eigenschaft als erfüllt angesehen werden. Sicherlich arbeitet TinkerPlots nicht fehlerfrei, grobe Fehler sind mir jedoch nicht bekannt.[8] *Effizienz* meint den Aufwand bei der Nutzung des Systems. Zu den verwendeten Algorithmen kann ich mangels Kenntnis keine Aussage treffen. Die Auslastung eines Rechners bei komplexen Vorgängen (damit sind rechentechnisch komplexe

[8] Durch jahrelange Benutzung der Software und eine intensive Überprüfung nahezu aller Funktionen während meiner Tätigkeit beim Lokalisieren des Programms für Deutschland, nehme ich diesbezüglich eine gewisse Kompetenz für mich in Anspruch.

Simulationen gemeint, die die Verwendung mehrere Bauteile in der Zufallsmaschine und zusätzlich das Sammeln von Messgrößen beinhalten) kann je nach Leistung des Computers für einige Sekunden bis Minuten sehr hoch sein. Vor allem die *Robustheit*, das ist der Umgang der Software mit fehlerhaften Eingaben, erscheint bei TinkerPlots gut umgesetzt, denn gerade bei der Benutzung von Software durch unerfahrene Anwender kann dies leicht geschehen. Fehlerhafte Eingaben führen bei TinkerPlots in vielen Fällen zu automatisch umgewandelten, sinnvollen Eingaben und einem entsprechenden Hinweis (vgl. Kapitel 3.5.1). *Verfügbarkeit* ist die Wahrscheinlichkeit, dass eine Software zu einem beliebigen Zeitpunkt des Einsatzes tatsächlich funktionsfähig ist, im Idealfall beträgt diese Wahrscheinlichkeit 1. Bei TinkerPlots ist die Verfügbarkeit der Software dann eingeschränkt, wenn Simulationen mit sehr großen Wiederholungszahlen durchgeführt werden, da während des Simulationsvorgangs die Software nicht anderweitig benutzt werden kann. Abhängig von der Rechenleistung des zugrunde liegenden Computers und der Komplexität der Simulation kann ein Simulationsvorgang durchaus mehrere Minuten in Anspruch nehmen, in denen die Software „lahmgelegt“ ist. Unter dem Punkt *Zuverlässigkeit* wird das Produkt aus Korrektheit, Robustheit und Verfügbarkeit gebildet, das zum Ausdruck bringt, mit welcher Wahrscheinlichkeit die Software zu einem bestimmten Zeitpunkt tatsächlich korrekt arbeitet. Bei der *Datensicherheit* geht es um das Ausgespähtwerden von Außen. Da TinkerPlots weder webbasiert ist noch eine aktive Internetleitung benötigt, sind die mit Hilfe der Software genutzten Daten letztlich genauso sicher und geschützt wie der zugrunde liegende Computer. Die letzten drei Punkte, *Verständlichkeit*, *Wartbarkeit* und *Wiederverwendbarkeit* beziehen sich vor allem auf die Entwicklung von Software, die *Austauschbarkeit* und *Dokumentation* unter verschiedenen Entwicklern und spielen in Bezug auf die Werkzeuganalyse kaum eine Rolle.

Weiterhin nennt Hartwig fünf didaktische Prinzipien als Anforderungen an die Ergonomie einer Software, die der Strukturierung und Gestaltung der Inhalte und deren Implementierung als Lernmedium dienen: *Selbstständigkeit*, *Adaptierbarkeit*, *Adaptivität*, *Motivation* und *Interaktivität* (Hartwig 2007, 74f, vgl. auch Dahm 2006). Die *Selbstständigkeit* des Benutzers als didaktisches Prinzip zielt auf die Kontrolle über den Nutzungsverlauf der Software. Der Anwender kann selbst entscheiden, in welcher Verwendung und Reihenfolge er die Software nutzt. Dies kann jedoch bei unerfahrenen Benutzern zu Überforderung führen, sofern die fachlichen und technischen Inhalte nicht ausreichend bekannt sind. Abhilfe kann hier das Prinzip der Führung bzw. Leitung schaffen, das den Nutzen der Software steuert, bis der Lernende mit steigender Vertrautheit mit den Inhalten eigene Wege gehen kann. Eine Möglichkeit hierzu ist die instrumentelle Orchestrierung, die in Kapitel 2.2.2 vorgestellt wurde. Prinzipiell kann das Arbeiten mit TinkerPlots in hohem Maße selbständig erfolgen, damit ist gemeint, dass von der Softwareseite keine feste Bearbeitungsreihenfolge vorgegeben ist. Das Prinzip der

Adaptierbarkeit beinhaltet die Möglichkeit, die Software nach eigenen Wünschen anzupassen. In TinkerPlots beschränkt sich ein Nutzer in der Arbeitsfläche auf die Objekte, die er tatsächlich benötigt. Beispielsweise wird zu einer Simulation kein Datenkartenstapel benötigt. Somit muss dieser auch nicht als Objekt verwendet werden und die Arbeitsfläche beinhaltet nur die zum Simulieren notwendigen Objekte. Zusätzlich können diese Objekte minimiert werden, so dass eine bessere Übersichtlichkeit durch den Benutzer erzeugt werden kann. Darüber hinaus kann in TinkerPlots die Bedienoberfläche insoweit angepasst werden, dass einige wenige Menüleisten ein- oder ausgeblendet werden können, wie zum Beispiel das Menü zur Textverarbeitung. Außerdem können für den Bereich der Datenanalyse in Graphen viele Funktionen komplett ausgeblendet werden. Dies bedeutet, dass TinkerPlots durch einen Lehrenden für einen unerfahrenen Lernenden soweit angepasst werden kann, dass bestimmte Inhalte, die zum einen mathematisch (noch) nicht bekannt sind und die zum anderen aus diesem Grund zum nicht zielgerichteten Spielen mit der Software einladen würden, unsichtbar gemacht werden können. Für den Bereich der Simulation ist eine solche Anpassung jedoch nicht möglich. Mit dem Prinzip der *Adaptivität* wird die automatische Anpassung der Software an den Lerner beschrieben. Hierzu ist ein gewisses Maß an künstlicher Intelligenz auf Seiten des Systems nötig, um Nutzerhandlungen zu erkennen und sinnvoll für den Nutzer auszuwerten. Eine solche Funktionalität ist bei TinkerPlots nicht vorgesehen, geschieht allenfalls in einem Lehr-Lernarrangement durch die Hilfestellung eines erfahrenen Benutzers. Unter dem Prinzip der *Motivation* wird der Anreiz verstanden, sich mit der Software zu beschäftigen. Dies wird vor allem durch die Gestaltung der Software beeinflusst. Im Hinblick auf die Zielgruppe von TinkerPlots, junge Lernende zwischen der dritten und zehnten Klasse, ist die Software ansprechend farbig und dennoch schlicht gehalten. Die Animationen des Objekts Zufallsmaschine für die Simulation tragen außerdem zur Motivation und kognitiven Orientierung eines Nutzers bei, da in gewisser Weise der Zufallsprozess, u. a. in Form von Urnenmodellen, sichtbar gemacht wird. Den Nutzen eines Computerprogramms, das mit visuellen Urnenmodellen ausgestattet ist, hat Sedlmeier für ein von ihm entwickeltes Training untersucht und kommt zu dem Schluss „More important than the immediate improvement in performance is the lasting effect of the urn model training. In terms of long-term effectiveness, training with the urn model was remarkably successful“ (Sedlmeier 1999, 139). Da TinkerPlots jedoch nicht mit einem systemimmanenten Tutor ausgestattet ist und keine interaktiven Elemente zwischen Benutzer und Software bietet, wird von Seiten des Programms keine weitere Motivation geboten. Das letzte Prinzip, die *Interaktivität*, ist die Klassifizierung von Interaktionsniveaus und bezieht sich vor allem auf eLearning Varianten. Dies ist als Prinzip für die Software TinkerPlots wenig passend, da die Software beispielsweise weder als Nachschlagmedium, noch als Multiple-Choice-Test noch mit einem Chatmodul konzipiert ist.

Mit diesen Ausführungen zu Software-Ergonomie und deren didaktischen Prinzipien als Grundlage wurde die Software als Ganzes eingeordnet. Eine inhaltliche Detailanalyse bezüglich der Möglichkeiten und Grenzen der Software TinkerPlots zum Simulieren stochastischer Zufallsexperimente folgt. Teilweise wird dabei zurückgegriffen auf Aspekte der Werkzeuganalyse der Software Fathom, die von Maxara (2009) im Rahmen ihrer Dissertation durchgeführt wurde. Dabei wird jedoch für die vorliegende Arbeit eine eigene Struktur entwickelt, da sich die beiden Programme zwar grundlegend ähneln, in der Benutzung und in der visuellen Repräsentation der einzelnen Objekte jedoch stark unterscheiden.

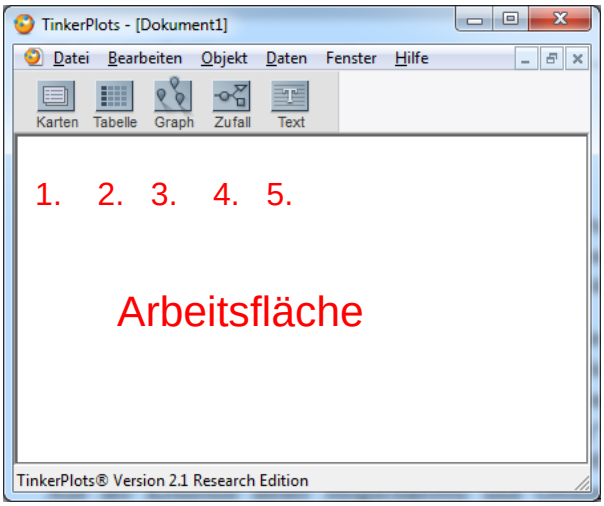


Abbildung 4: TinkerPlots Oberfläche

Es folgt ein kurzer Einblick in die Arbeitsweise mit TinkerPlots, um die anschließende detaillierte Werkzeuganalyse besser einordnen zu können.

Es gibt fünf grundlegende Objekte in TinkerPlots (vgl. Abbildung 4):

1. Datenkarten,
2. Tabellen,
3. Graphen,
4. Zufallsmaschinen und
5. Textfelder.

Diese Objekte werden in die Arbeitsfläche gezogen und können dort beliebig positioniert und benutzt werden. Bei Benutzung in der Arbeitsfläche sind die ersten vier Elemente miteinander vernetzt, das bedeutet, eine Änderung der Daten in einem Objekt führt zur entsprechenden Änderung in den vernetzten Objekten. Somit ist das dynamische Arbeiten gesichert. Darüber hinaus gibt es noch die Objekte Bild, Regler und Flashobjekt, die nur über das Menü *Objekt* zugänglich sind.

3.2.2 Kurzüberblick zur Datenanalyse mit TinkerPlots

Die Datenanalyse wird hier nur kurz dargestellt, eine ausführliche Beschreibung findet sich in Frischemeier (2017).

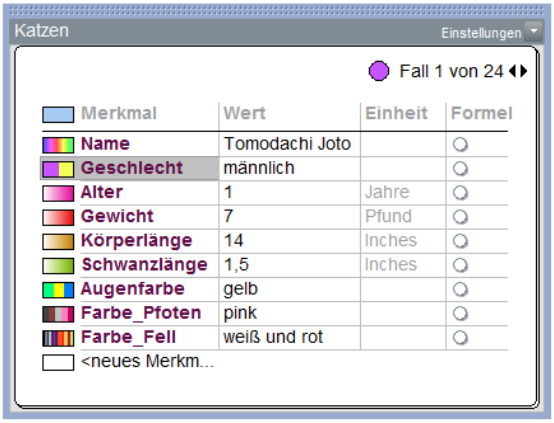


Abbildung 5: Datenkartenstapel

Daten sind in TinkerPlots in Datenkarten (Abbildung 5) organisiert. Der Datenkartenstapel beherbergt alle Daten. Jeder Fall bekommt eine eigene Karte, auf der beliebige Merkmale eingetragen werden können, welche sich auf den folgenden Karten wiederfinden. Die Karten werden zu einem Datenkartenstapel zusammengefasst und bilden so die Datengrundlage (Abbildung 5). Der Stapel kann durchgeblättert werden und jeder Fall mit seinen Merkmalswerten kann einzeln angesehen werden. Die Fälle können zusätzlich in einer Tabelle repräsentiert werden, wobei jeder Fall einer Zeile und jedes Merkmal einer Spalte entspricht (Abbildung 6).

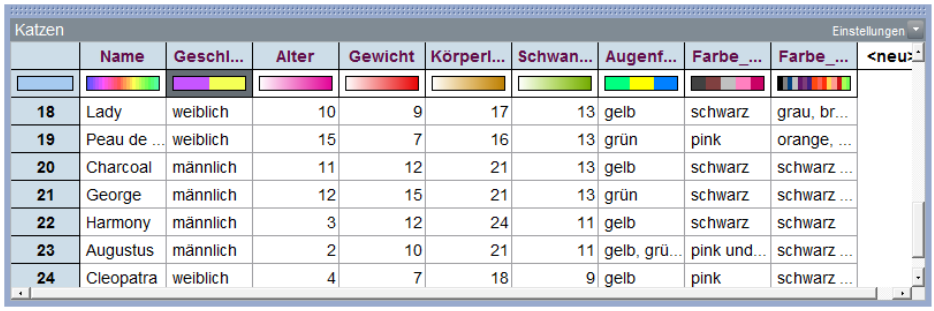


Abbildung 6: Repräsentation von Daten in der Tabelle

Dieses Gerüst ist fix, das heißt Zeilen und Spalten können nicht ausgetauscht werden. Wohl aber kann die Reihenfolge der Merkmale variiert werden. Änderungen in der Tabelle bewirken entsprechende Änderungen in den Datenkarten und umgekehrt. Ob die Daten in einer Tabelle, in Datenkarten oder in beidem repräsentiert werden, bleibt dem Anwender überlassen. Beide Objekte sind für das Programm gleichrangig, Änderungen werden jedoch in allen zusammenhängenden Objekten dynamisch aktualisiert. Liegen die Daten bereits in einem von beidem vor, so kann das zweite Objekt per Drag & Drop

in die Arbeitsfläche gezogen werden und ist bereits mit den Daten bestückt. Ein Ausfüllen von Hand des zweiten Objekts mit denselben Daten ist nicht nötig.

Abbildung 7: Repräsentation von Daten im Graph

Visualisiert, analysiert und ausgewertet werden die Daten in so genannten „Graph-Objekten“ (Abbildung 7). Dabei liegen zunächst alle Fälle als Punkte im Graph-Objekt ungeordnet vor und können dann mit verschiedenen Aktionen nach den gewünschten Gesichtspunkten geordnet werden. Es gibt in TinkerPlots keine vorgefertigten Graphen, sondern jeder Graph muss von Hand durch verschiedene Aktionen, z. B. Verschieben der Datenpunkte, Ordnen, Stapeln, Trennen, Einblenden von Hilfslinien, Kennwerten, etc., erstellt werden. Die Objekte Datenkarten, Tabelle und Graph sind dynamisch miteinander vernetzt, das bedeutet, Änderungen wirken sich auf alle beteiligten Objekte aus.

3.2.3 Kurzüberblick zur Simulation mit TinkerPlots

Bevor die Software hinsichtlich ihrer Möglichkeiten und Grenzen analysiert wird, wird hier ein kurzer Einblick in das Simulieren mit TinkerPlots gegeben. Zunächst muss jedoch eine wichtige Unterscheidung hinsichtlich des Begriffs Zufallsmaschine getroffen werden. In TinkerPlots bezeichnet der Begriff Zufallsmaschine das gesamte Objekt, das für die Visualisierung des zufallserzeugenden Prozesses genutzt wird (Abbildung 8). Dieser Begriff wurde bereits in der Software Fathom verwendet, allerdings unterscheidet er sich hier im Gebrauch von dem in Fathom. In Fathom bezeichnet Zufallsmaschine ein „Grundkommando“ (Maxara 2009, 57), das aus einer Funktion besteht, mit der Zufallszahlen erzeugt werden. Dies sind zum Beispiel die Funktionen *ZufallsWahl()*, *ganzeZufallszahl()*, etc. Mit Hilfe dieser Zufallsmaschinen werden in Fathom verschiedene Zufallsmodelle repräsentiert, jedoch immer in Form eines Programmierbefehls. In TinkerPlots wird der Begriff Zufallsmaschine auf das gesamte in Abbildung 5 zu sehende Objekt mit all seinen Funktionen angewendet. Somit ist die Verwendung des Begriffs Zufallsmaschine in TinkerPlots weitreichender.

Dieses vierte Objekt, die Zufallsmaschine, ist zunächst unabhängig von den anderen drei Objekten, da mit ihr Daten erzeugt werden können. Verschiedenste Zufallsexperi-

mente und Wahrscheinlichkeitsprozesse können mit Hilfe der Zufallsmaschine modelliert und erfassbar gemacht werden. Die erzeugten Daten werden in eine automatisch erstellte Tabelle geschrieben und können bei Bedarf zusätzlich in einem Datenkartenstapel repräsentiert werden. Diese zusätzliche Repräsentation muss durch den Benutzer ausgeführt werden, sie geschieht nicht automatisch.

Abbildung 8: Das Objekt Zufallsmaschine

Das Herzstück der Zufallsmaschine ist ihre mittlere Fläche. Für die Modellierung stehen sechs verschiedene Bauteile zur Verfügung, die eine intuitive Umsetzung (vgl. Kapitel 3.5) von Ereignissen und zugehörigen Wahrscheinlichkeiten erlauben: Es gibt die Bauteile Box (zu sehen in Abbildung 8), Stapel, Kreisel, Verteilungsbalken, Verteilungskurve und Stückzähler. Die Bauteile können durch Ziehen in die mittlere Fläche miteinander ausgetauscht werden. Außerdem können die Bauteile hintereinander gesetzt werden und mit Abzweigungen versehen werden. So können durch mehrfache Kombination von Bauteilen mehrstufige abhängige oder unabhängige Zufallsexperimente erstellt werden. Hiermit und durch die Variation der Anzahl von Ziehungen und Durchgängen können Zufallsexperimente auf vielfältige Weise umgesetzt werden.

Die durch Simulation erzeugten Daten werden automatisch in eine Tabelle geschrieben (analog zu der in Abbildung 6). Dabei werden die einzelnen Merkmale jeweils in eine eigene Spalte geschrieben und mit den simulierten Werten bestückt. Zusätzlich werden in einer Spalte alle Merkmalswerte eines Durchgangs als zusammengesetztes Ergebnis zusammengefasst.[9] Weiterhin können in der Tabelle zusätzliche Spalten zur Auswertung benutzt werden.

Besonders hilfreich für die Umsetzung von Zufallsgrößen ist ein Menü, aus dem voreingestellte Merkmale durch einfaches Anklicken erzeugt werden können (zum Beispiel die Summe mehrerer Merkmale). Ebenso können mithilfe von Formeln, die einer ge-

[9] Dies ist eine wesentliche Neuerung gegenüber der Software Fathom. Durch dieses zusammengesetzte Ergebnis lassen sich Auswertungen im Vergleich zu Fathom wesentlich vereinfachen oder werden überhaupt erst möglich.

wissen Syntax unterliegen, Auswertungsmerkmale erstellt werden. Hervorzuheben sind hier die „Listen“-Befehle, die viele Auswertungen abdecken. Dies ist nicht mehr ganz so intuitiv wie die restliche Bedienung des Programms, erlaubt jedoch eine große Fülle an Realisierungen. Hierzu steht ein Formeleditor zur Verfügung, der zu jeder Formel einen kurzen Hilfetext in Form von Syntax, Beschreibung und Beispiel anbietet.[10]

Die Untersuchung der simulierten Daten findet wie bei der Datenanalyse auch in Graphen statt.

Ein weiteres mächtiges Objekt der Simulation mit TinkerPlots ist das, was als das Sammeln von Messgrößen bezeichnet werden soll. Der Begriff Messgröße ist der Software Fathom entlehnt, da hier eine ähnliche Funktion in TinkerPlots hervorgehoben werden soll.[11] Eine Messgröße in TinkerPlots ist eine definierbare Größe, die sich auf einzelne Merkmale bezieht oder auf die Zufallsmaschine als Ganzes. Die Definition von Messgrößen geschieht entweder direkt in der Zufallsmaschine oder durch Auswahl bestimmter Werte von in Graphen repräsentierten Kennzahlen oder Häufigkeiten. Mit Hilfe dieses Konzepts werden zusätzliche Simulationsmöglichkeiten begründet, da auf diese Weise Ereignisse und Zufallsgrößen festgelegt werden können. Beim Sammeln von Messgrößen geschieht die Simulation und Wiederholung des zugehörigen Zufallsexperiments automatisch und die gesammelten Werte werden in einer eigenen (Messgrößen-)Tabelle abgelegt. Ereignisse und Zufallsgrößen werden in TinkerPlots entweder in Merkmalen oder in Messgrößen definiert, so dass diese beiden abstrakten Begriffe anwendungsbezogen in der Software umgesetzt werden und hierdurch gleichzeitig eine Unterscheidung verschiedener Simulationsarten herausgearbeitet werden kann (vgl. Kapitel 3.5.2).

3.3 TinkerPlots als kognitives Werkzeug/expressives Medium

TinkerPlots setzt auf Visualisierungen bei der Simulation. Das Herzstück der Simulation ist die sogenannte Zufallsmaschine, die als Werkstatt für die Konstruktion von Zufallsexperimenten betrachtet werden kann. Durch Benutzung dieses Instruments (vgl. Kapitel 2.2) können in dieser Zufallswerkstatt über verschiedene Bauteile verschiedene „Zufallsmaschinen“ realisiert werden. Dabei arbeitet der Benutzer bei der Realisierung

---

[10] Der Formeleditor ist komplett von Fathom adaptiert und um die erwähnten „Listen“-Befehle erweitert worden.

[11] Der Begriff „Messgrößen“ wurde für TinkerPlots aus der deutschen Fathomsoftware entlehnt, denn er hat eine ähnliche Funktionalität wie dort. In der englischen TinkerPlots-Version heißt diese Funktion „history“, was gerade nicht aus der englischen Version von Fathom stammt, dort heißt die Funktion „measure“. Ein persönliches Gespräch mit dem TinkerPlots-Entwickler Cliff Konold hat gezeigt, dass über den Begriff „history“ für TinkerPlots lange diskutiert wurde, da der Begriff „measure“ aufgrund der neuen Funktionalität des Abstandsmessens mit einer Linealfunktion in TinkerPlots bereits belegt war. Für die deutsche Version ist die Entscheidung wiederum zugunsten des bereits aus Fathom bekannten Begriffs „Messgröße“ gefallen, da dieser Begriff die zugehörige Funktionalität am bestens beschreibt und eine Dopplung wie in der englischen Version nicht zutrifft.

von zufallsgenerierten Daten mit intuitiv bedienbaren Bauteilen. Ein stochastisches Modell, das mental mit Boxen (Urnen), Glücksrädern etc. formuliert wurde, kann quasi direkt in TinkerPlots übertragen werden, ohne dass eine weitere Übertragung in spezifische Programmiersprachen/-befehle notwendig ist. Dabei ist das Erstellen von einstufigen, mehrstufigen, abhängigen oder unabhängigen Modellen in TinkerPlots gleichermaßen möglich und intuitiv per Drag & Drop (siehe Abschnitt 3.6 und Kapitel 4). Die Erstellung der Zufallsmaschine erfolgt weitgehend auf einer visuellen Ebene, zum Beispiel ist die Bestückung einer Urne mit nahezu beliebig beschrifteten Kugeln so möglich, wie es den Vorstellungen des Benutzers entspricht. Das mentale Modell entsteht dabei in Verbindung mit dem Realmodell im Modellierungskreislauf nach Blum (2006), für das eine gegebene Situation vereinfacht, idealisiert und strukturiert wird. Dadurch ist der Schritt der Mathematisierung (Abbildung 9, Schritt 3) im Modellbildungskreislauf wesentlich vereinfacht, da die Mittel der Software eine visuelle Repräsentation des mentalen Modells erlauben, ohne dass das zugrunde liegende mathematische Modell explizit definiert werden muss.

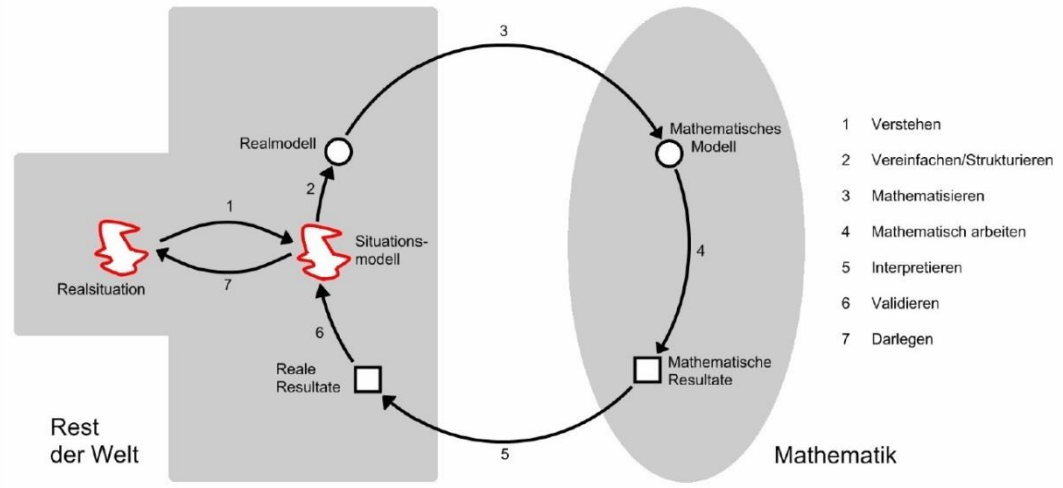


Abbildung 9: Modellbildungskreislauf nach Blum (2006)

Mit der Zufallsmaschine bietet TinkerPlots dem Benutzer die Möglichkeit eines konkreten visuellen Modells am Bildschirm. Anstelle des mathematischen Modells in Abbildung 9 tritt die visuelle (und maschinelle) Repräsentation durch die Zufallsmaschine. Schritt 3, das Mathematisieren, wird somit zum Programmieren, was jedoch in TinkerPlots benutzerfreundlich und einfach möglich ist. Der zufallserzeugende Prozess, also beispielsweise das Durchmischen der Kugeln in einer Box oder das Drehen des Kreisels, ist in TinkerPlots für den Nutzer sichtbar und somit nachvollziehbar, was eine der größten Stärken der Software ist. Durch die leichte Manipulierbarkeit kann das visuelle Modell jederzeit angepasst werden, das Modell ist also transparent für den Benutzer und

stellt keine Blackbox[12] dar, was ansonsten bei Zufallsbefehlen oder -funktionen in anderen Programmen (z. B. Fathom oder Excel) der Fall ist. Betrachtet man das Repräsentationsvermögen der Zufallsmaschine für das mathematische Modellieren von Zufallsexperimenten, so kann dies als hoch eingestuft werden durch die Art, in der dynamische, manipulierbare und interaktive Repräsentationen das mathematische Denken und Ausdrücken fördern. Genau diese Repräsentationsfähigkeit ist nach Specker (1997) in seiner Dissertation über Software Engineering die Definition eines kognitiven Werkzeugs:

> „Was zeichnet nun aber ein kognitives Werkzeug näher aus? Es ist die Möglichkeit, Wissen ausserhalb des menschlichen Geistes ‚künstlich' zu repräsentieren und somit die interne Repräsentation zu unterstützen." (Specker 1997, 17)

Gerade das äußere Abbilden einer internen Repräsentation eines Zufallsmodells unterstützt TinkerPlots. Dadurch sinkt zusätzlich die Einarbeitungszeit in das Programm, denn es müssen zur Zufallsgenerierung keine spezifischen Befehle gelernt werden und die Funktionsweise der einzelnen Bauteile für die Erstellung der Zufallsmaschine erschließt sich gut nach einer kurzen Einarbeitungszeit. Eine direkte Manipulation der Bauteile, z. B. das Hinzufügen weiterer Kugeln in einer Box, die Größenänderung eines Segments des Kreisels oder das Umbenennen von Elementen, verhilft dem Benutzer die innewohnende Wahrscheinlichkeitsverteilung leicht zu verändern. Ebenso führt die einfache Anpassung der Durchführungsanzahl eines Zufallsexperiments zu mehr Transparenz in der Auswirkung auf die Genauigkeit einer Simulation. Ob die Zufallsmaschine zehn, hundert oder zehntausend Durchläufe haben soll, ist nur abhängig von der entsprechenden Zahleneingabe und kann jederzeit durch den Benutzer verändert werden.

Ein kognitives Werkzeug soll dem Anwender helfen, sein Wissen zu organisieren und zu strukturieren, in einer entsprechenden Lernumgebung kann das Wissen geeignet repräsentiert werden. Die verschiedenen Bauteile der Zufallsmaschine erlauben eine gewisse Individualisierung der Repräsentation, wodurch jeder Benutzer die Lernumgebung in bestimmten Grenzen selbst gestalten kann. Nach Sedlmeier (1999) ist das Urnenmodell ein zentrales mentales Modell in der Stochastik, das universal eingesetzt werden kann (ebd., 121) und somit ein wichtiges mentales Hilfsmittel darstellt. Dieses mentale Modell kann direkt in TinkerPlots mit dem Bauteil der Box visualisiert werden, somit kann in die Zufallsmaschine dieses kognitive Modell direkt übertragen und benutzt werden. TinkerPlots ist demnach auch ein kognitives technisches Werkzeug, in das direkt ein mentales Hilfsmittel übertragen werden kann. So kommen nach der Definition eines „cognitive tool" von Grune (2000, 60) sogar zwei Aspekte zusammen: „Als cognitive tools können somit mentale oder technische Hilfsmittel bezeichnet werden,

[12] In der Tat stellt auch die Zufallsmaschine in TinkerPlots eine Blackbox dar, da im Hintergrund wie bei anderen Programmen auch über Programmierbefehle Zufallszahlen erzeugt werden. Diese werden jedoch durch die Zufallsmaschine metaphorisch deutlich gemacht, so dass der Eindruck entsteht, dass der Benutzer die Erzeugung der Zufallswerte nachvollziehen kann.

die den Denkprozess ihrer Nutzer unterstützen, anleiten oder erweitern.“ Der kognitive Lernprozess, der hier angesprochen wird, wird durch die Arbeitsweise unterstützt, die TinkerPlots vorgibt, da das mentale Hilfsmittel quasi direkt im technischen abgebildet werden kann und die Funktionsweise der Zufallsmaschine sichtbar ist. Vor allem die Bauteile Box und Kreisel ähneln ihren realen Gegenstücken und können leicht der Vorstellung angepasst werden.

Für den Bereich der Simulation kann TinkerPlots als Programmiersoftware betrachtet werden. Im Hinblick auf die Zielgruppe der Lernenden der Klassen drei bis zehn ist das Programmieren wie oben beschrieben mit Hilfe des kognitiven Werkzeugs repräsentiert durch das Objekt Zufallsmaschine weitgehend visuell und somit intuitiv. Folgt man dem Gedanken von „programming is represented as a tool for expression and articulation” von Noss und Hoyles (1996, 57), dann hat die Software als “microworld” (Noss & Hoyles 1996, 63) eine wichtige expressive Komponente. Durch den Blick auf TinkerPlots als *microworld*, die ein Werkzeug für den Benutzer ist, kann das Sinnstiften beim Benutzer unterstützt werden. Gefördert wird dadurch auch das Arbeiten an offenen Fragen, die eine gewisse Exploration erfordern, so wie die Autoren es von einem Computerwerkzeug fordern:

> „The idea of microworlds involves an intention to develop an open and investigative stance to mathematical enquiry.” (Noss & Hoyles 1996, 65)

In einem weiteren Artikel unterscheiden Hoyles & Noss bei Software zwischen *programmable tools* und *expressive tools*, wobei der Begriff *microworld* dem ersteren zuzuordnen ist:

> „On the one hand, programming or building programmable tools, presents novel ways of modelling and representing mathematics, while, on the other hand, what we shall term expressive tools aim to provide ready access to the results of procedures and algorithms without the necessity for learners to attend to their production, to open up the tools or to evaluate alternative representations. The outcome of using the tools rather than the tool structures is the focus of the users’ thinking – to obtain an answer or some information, to calculate a result, to construct a graph.” (Hoyles & Noss 2003, 237)

In TinkerPlots sind beide Arbeitsweisen möglich, es kann ein Modell von Grund auf selbst erstellt bzw. programmiert werden oder es können in einer vorbereiteten Lernumgebung bestimmte Prozesse untersucht werden. Dadurch vereint die Software als *microworld* Aspekte von *programmable tools* und von *expressive tools* in sich. Gerade aber der Aspekt des Modellierens und des visuellen Repräsentierens von stochastischen Prozessen hat einen hohen Stellenwert beim Arbeiten mit TinkerPlots. Mit Hilfe von TinkerPlots als *expressive tool* können Zufallsmaschinen konstruiert und benutzt werden, die als Modelle untersucht werden können und zu bestimmten Problemen simulative Lösungen anbieten. Durch das visuelle Arbeiten werden die Gedanken des Benutzers gleichzeitig externalisiert und durch die Benutzung des Werkzeugs präzisiert. Dieser Gedanke wird auch von Doerr & Pratt aufgegriffen:

> „Building a model (or expressive modeling) provides learners with the opportunity to express their own concepts and to learn through the iterative process of representing their ideas, selecting objects, defining relationships among objects, operating on those relationships, and interpreting and validating outcomes." (Doerr & Pratt 2008, 265)

Betrachtet man TinkerPlots als *programmable microworld* oder als Software, mit der *expressive modeling* möglich ist, dann allerdings als solche, in der vorwiegend visuell programmiert wird. TinkerPlots basiert nicht auf einer textuellen Programmiersprache, allerdings gibt es zur Auswertung durchaus die Möglichkeit, Formeln einzugeben. Die visuelle Programmierung einer Zufallsmaschine in TinkerPlots hat den Vorteil, dass die zugrunde liegenden computerspezifischen Algorithmen nicht umfassend verstanden sein müssen. Schiffer (1998, 325) nennt weitere Vorteile des visuellen Programmierens:

> „Eine Reihe weiterer Vorteile, wie leichte Erlernbarkeit, intuitive Handhabung, Reduktion von Fehlerquellen und Anpaßbarkeit an persönliche Bedürfnisse, begründen die überragende Stellung grafischer Benutzungsschnittstellen als Mechanismus für die Interaktion zwischen Anwender und Applikation."

Dabei ist mit „grafischer Benutzungsschnittstelle" nicht eine ausschließliche Verwendung von graphischen Repräsentationen zur Programmierung gemeint. Es wird von Schiffer eingeräumt, dass an einigen Stellen immer Beschreibungen, Formeln oder Beschriftungen in Textform benötigt werden. Es geht darum, dass die überwiegende Programmierung mit visuell repräsentierten Objekten stattfindet, so wie es bei der Zufallsmaschine von TinkerPlots der Fall ist. Vor allem der Punkt „Reduktion von Fehlerquellen" ist von zentraler Bedeutung. Es ist in TinkerPlots fast unmöglich, eine Zufallsmaschine zu konstruieren, die nicht richtig funktioniert, also im Sinne einer Programmierung syntaktisch falsch ist. Bauteile können nur dort eingefügt werden, wo dies syntaktisch sinnvoll ist. Ebenso verhält es sich mit Eingaben zum Beispiel bei der Wiederholungsanzahl oder der Ziehungsanzahl, die nur so gewählt werden können, dass sie durch die Software weiterverarbeitet werden können. Fehlerhafte Eingaben werden erkannt und direkt durch eine sinnvolle Eingabe (in beiden genannten Fällen durch eine „1") ersetzt. Somit ist der Frustrierungsgrad recht gering, da ausschließlich Zufallsmaschinen erstellt werden können, die lauffähig sind. Natürlich heißt das noch lange nicht, dass im Hinblick auf eine Aufgabe nur sinnvolle, d. h. passende Zufallsmaschinen gebaut werden. Hier obliegt es dem Benutzer, aus den vorhandenen Mitteln die richtigen auszuwählen um einen gewünschten Prozess stochastisch korrekt abzubilden.

3.4 Strukturierung der Software: Bereiche zur Simulation

Angelehnt an die vier Interfaces zwischen Stochastik und Software (Kapitel 2.4) ergeben sich bestimmte Bereiche, die funktionale Einheiten für die Simulation definieren. Für diese funktionalen Einheiten müssen nacheinander gewisse Entscheidungen bzw. Modellierungen getroffen werden, um eine Simulation durchzuführen. Für die Simulationsumsetzung mit TinkerPlots werden neun Bereiche definiert, die in gewisser Weise

eigene Objekte sind, gleichzeitig jedoch voneinander abhängen, um ein bestimmtes Zufallsexperiment umzusetzen. Mit den so definierten Bereichen ergibt sich eine Analysestruktur, die für die später definierten Simulationsarten verwendet wird (siehe Abschnitt 3.5.2) und gleichzeitig eine gewisse Strukturierung anbietet. Diese Strukturierung wird auch im neu entwickelten Simulationsplan abgebildet (siehe Kapitel 2.4). Eine Zuordnung der Bereiche zu den Interfaces ist abhängig von der Simulationsart und geschieht im jeweiligen Abschnitt zum entsprechenden Interface. Tabelle 1 zeigt die neun definierten Bereiche und gibt durch die Nummerierung eine mögliche Bedienungsreihenfolge wieder, die für die Analyse einer stochastischen Problemstellung genutzt werden kann.

Tabelle 1: Bereiche der funktionalen Einheiten zur Simulation

Bereich	Funktionale Einheit
Bereich 1a und 1b	Bauteile
Bereich 2	Ziehungen (Merkmale)
Bereich 3	Durchgänge (Wiederholung)
Bereich 4	Steuerungselemente
Bereich 5	Ergebnistabelle
Bereich 5a und 5b	Weitere Merkmale
Bereich 6	Graph (Auswerten)
Bereich 7	Messgrößensammeln im Graph und Messgrößentabelle
Bereich 8	Anzahl Messgrößen
Bereich 9	Messgrößen auswerten (Graph)

Durch das Design der Software, speziell der Zufallsmaschine (als Objekt der Software), wird das Interface Wiederholung durch den geübten Anwender bereits bei der Modellierung der Zufallsmaschine bedient und steht deshalb recht früh in der Reihenfolge. Allerdings kann die Wiederholungsanzahl auch an späterer Stelle beliebig variiert werden. Die Defaulteinstellung sieht fünf Wiederholungen vor, muss also zunächst durch den Anwender nicht explizit angegeben werden, um eine Simulation durchführen zu können.

Die Bereiche sieben bis neun werden nur benötigt, sofern eine Simulation mit Messgrößen erstellt wird.

An einem Beispiel wird dies nun verdeutlicht. Die folgende Abbildung 10 zeigt farblich abgesetzt die verschiedenen Bereiche. Simuliert wurde hier zur historischen Fragestellung von Chevalier de Méré (1607-1684): *Tritt beim dreifachen Würfelwurf die Augensumme 11 oder die Augensumme 12 häufiger auf?* Die hier graphisch voneinander abgesetzten Bereiche werden im Folgenden in Bezug zum jeweiligen Interface beschrieben und mit ihren Möglichkeiten erläutert.

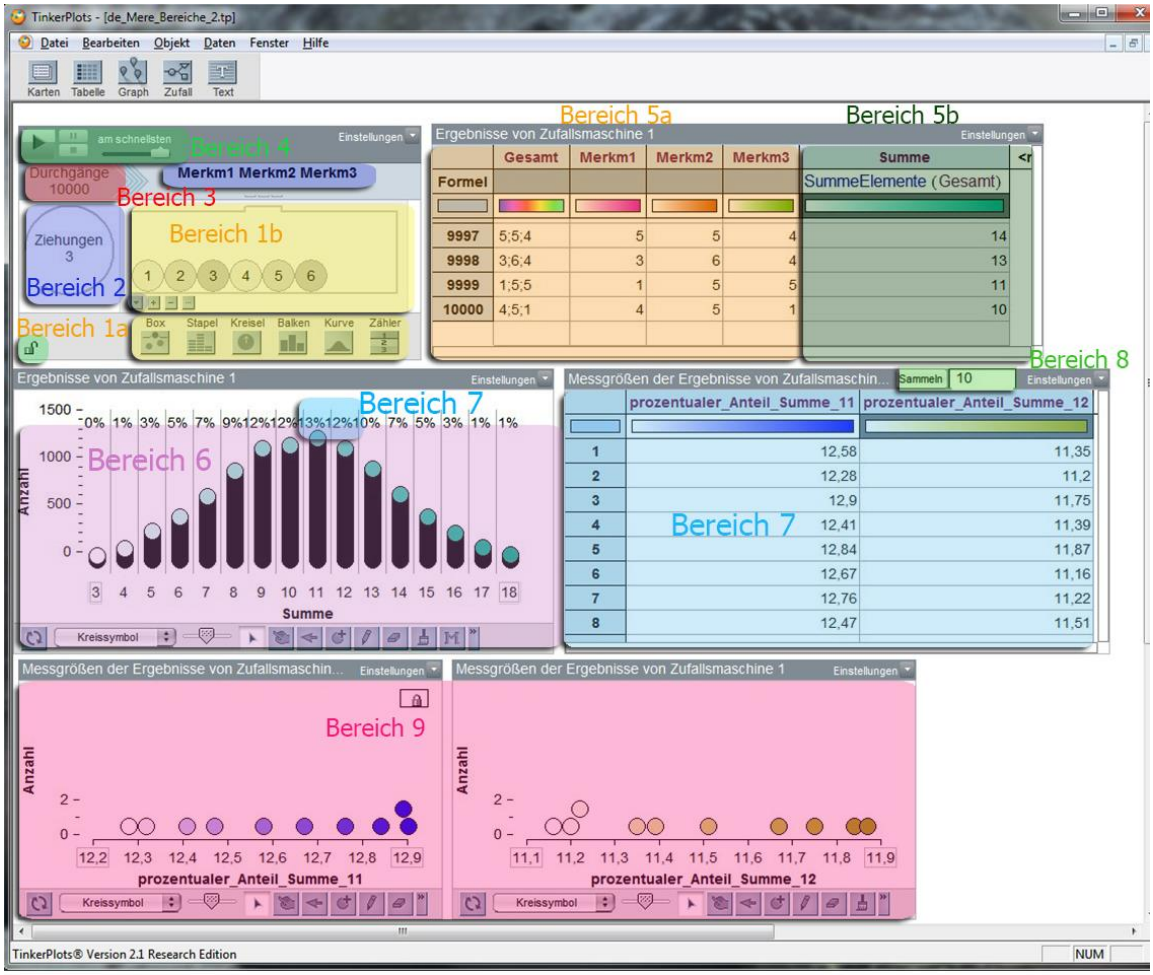


Abbildung 10: Übersicht zu den für eine Simulation notwendigen Bereichen

Der dreifache Würfelwurf wird durch den Wahrscheinlichkeitsraum $(\Omega, \mathcal{P})$ mit

$$\Omega = \{(x, y, z) \mid x \in \{1; 2; 3; 4; 5; 6\}, y \in \{1; 2; 3; 4; 5; 6\}, z \in \{1; 2; 3; 4; 5; 6\}\}$$

$$\text{und } \mathcal{P}(\omega) = \frac{1}{216} \; \text{für alle } \omega \in \Omega$$

beschrieben. Die Definition des Ergebnisraums Ω ist durch die Zufallsmaschine gegeben (Bereich 1-4 in Abbildung 10). Dazu wird für die Modellierung eines Würfels in Bereich 1a das Bauteil Box ausgewählt und in Bereich 1b mit sechs Kugeln versehen, die von 1 bis 6 beschriftet sind. Um den dreifachen Würfelwurf zu modellieren, werden in Bereich 2 drei Ziehungen mit Zurücklegen benötigt. Die Nutzervorstellung dazu kann sein, dass ein in Bereich 1b über das Bauteil Box modellierter Würfel dreimal geworfen wird, aus dem Bauteil also dreimal mit Zurücklegen gezogen wird. Jeder einzelne Würfelwurf bekommt ein eigenes Merkmal, diese sind per Standardeinstellung mit den Namen *Merkm1*, *Merkm2* und *Merkm3* versehen. Um per Simulation eine Genauigkeit von

± 1 Prozentpunkten zu erreichen (vgl. $1/\sqrt{n}$ – Gesetz, z. B. Biehler & Prömmel 2013), wird das Zufallsexperiment des dreifachen Würfelwurfs zehntausend Mal wiederholt, welches durch die Anzahl der Durchgänge in Bereich 3 abgebildet ist. Die Geschwindigkeit des Simulationsvorgangs wird in Bereich 4 gesteuert und ist hier auf „am schnellsten“ eingestellt. Die 3-Tupel des Ergebnisraums werden im Merkmal „Gesamt“ in Bereich 5a dargestellt, ebenso wie die einzelnen Ergebnisse der Würfelwürfe in den Merkmalen Merkm1, Merkm2 und Merkm3 repräsentiert sind. Für die Beantwortung der Frage wird die

Zufallsgröße X: „Augensumme beim dreifachen Würfelwurf“

benötigt. Diese ist in Bereich 5b als Merkmal „Summe“ in der Tabelle realisiert. In diesem Merkmal wird die Summe aller Einträge des Merkmals „Gesamt“ gebildet. Die Verteilung der Zufallsgröße mit dem jeweiligen prozentualen Anteil ist im Graph in Bereich 6 dargestellt (Abbildung 10, Mitte links). Es ist zu sehen, dass die Augensumme 11 in 13% der Fälle aufgetreten ist und die Augensumme 12 in 12% der Fälle. Somit lässt sich schätzen, dass die Augensumme 11 wahrscheinlicher ist, als die Augensumme 12.

Dies deckt sich mit theoretischen Überlegungen, die hierfür angestellt werden können. Für die Augensumme 11 gibt es insgesamt 27 Kombinationsmöglichkeiten beim dreifachen Würfelwurf, für die Augensumme 12 gibt es nur 25 Kombinationsmöglichkeiten. Somit ist theoretisch die Wahrscheinlichkeit P(„Augensumme=11“) = $\frac{27}{216}$ = 0,125 und P(Augensumme=12“) = $\frac{25}{216} \approx 0{,}116$.

Als Ergänzung kann zusätzlich die Schwankung der Simulation beobachtet werden. Dazu wird der Anteil der Augensummen 11 und 12 als Messgrößen gesammelt (Bereich 7). Dies wird in einer eigenen Tabelle protokolliert (Abbildung 10, Mitte rechts). Es werden zehn weitere Messgrößen gesammelt (Bereich 8), um einen ersten Eindruck der Schwankungen zu erhalten. Deren Auswertung erfolgt erneut in einem Graph (Bereich 9).

Im nächsten Abschnitt wird nun genau analysiert, wie die Zufallsmaschine mit ihren Bauteilen zur Erzeugung von Zufall benutzt werden kann.

3.5 Interface Modellieren: (Pseudo-)Zufall erzeugen in TinkerPlots

Das Herzstück der Simulation in TinkerPlots ist das Objekt Zufallsmaschine (vgl. Abbildung 8). Hiermit werden Zufallsgeräte definiert, mit denen Zufallsexperimente in TinkerPlots modelliert werden, und somit (Pseudo-)Zufallszahlen erzeugt. Da Computer grundsätzlich deterministisch arbeiten, werden durch die Computersimulation keine echten Zufallszahlen, wie beispielsweise beim tatsächlichen händischen Würfelwurf,

sondern sogenannte Pseudozufallszahlen erzeugt. Die durch den Computer erzeugten Zufallszahlen wiederholen sich nach einer gewissen Periode, die jedoch hinreichend groß ist. Der bei TinkerPlots eingesetzte Algorithmus zur Erzeugung von Pseudozufallszahlen ist derselbe wie bei der Software Fathom, welcher in Marsaglia und Zaman (1991) beschrieben ist.

Durch die verschiedenen Einstellmöglichkeiten der Zufallsmaschine lassen sich sehr viele stochastische Zufallsexperimente modellieren. Die Zufallsmaschine kann aus sechs verschiedenen Bauteilen gebaut werden. Je nach Fragestellung ist die Modellierung nach der Wahl des entsprechenden Bauteils recht intuitiv. Häufig gibt es mehr als eine günstige Umsetzung, so dass unterschiedliche Vorstellungen des Benutzers abgebildet werden können. Zum Interface Modellieren werden die Bereiche 1a, 1b und 2 zugeordnet, da in diesen ein Zufallsexperiment in TinkerPlots umgesetzt wird. Hiermit werden Ergebnismengen realisiert und Wahrscheinlichkeitsverteilungen erzeugt.

Aus didaktischer Sicht kann die Zufallsmaschine von TinkerPlots im Sinne von Konold, Harradine und Kazak (2007) als „Datenfabrik“ aufgefasst werden. Damit wird der Gedanke verfolgt, dass die Zufallsmaschine eine Maschine in einer Fabrik darstellt, die Objekte oder auch Personen mit bestimmten Eigenschaften „herstellt“.

3.5.1 Die Bauteile der Zufallsmaschine im Detail

Bereich 1: „Bauteile“

Der erste Bereich beinhaltet die Umsetzung verschiedener Zufallsgeräte der Zufallsmaschine: die Bauteile. Er teilt sich auf in Bereich 1a, der die sechs verschiedenen Bauteile ikonisch abbildet und Bereich 1b als das innere Feld, in dem die Bauteile benutzt bzw. bestückt werden inklusive dem Pfeilmenü und den Buttons +, -, … . Im Bereich 1a sind alle zur Verfügung stehenden Bauteile aufgelistet und als Icon repräsentiert: Bauteil Box, Bauteil Stapel, Bauteil Kreisel, Bauteil (Verteilungs-)Balken, Bauteil (Verteilungs-)Kurve und Bauteil Zähler (kein Zufallszahlengenerator). Durch Hineinschieben in Bereich 1b können die Bauteile dort genutzt werden. Dazu gehört das Auswählen und Benennen der Elemente, die für eine Realisierung eines Zufallsexperiments benötigt werden. Das Pfeilmenü (bis auf den Eintrag *Zurücklegen*, der zum nächsten Bereich gezählt wird) und die Buttons +, -, …, mit denen Einstellungen zu den Bauteilen vorgenommen werden können, gehören ebenfalls zu Bereich 1b. Die Bauteile können einzeln verwendet werden oder hintereinander geschaltet und mit Abzweigungen versehen werden. So können komplexe Zufallsmaschinen mit bis zu 100 Bauteilspalten und beliebigen Abzweigungen erstellt werden. Beim Erstellen von mehrstufigen Zufallsexperimenten entspricht jede Bauteilspalte einem Merkmal, die Anzahl der Merkmale (Bereich 2) kann also nicht mehr frei gewählt werden (vgl. Kapitel 4).

Im Weiteren werden die einzelnen Bauteile im Detail beschrieben. Im Anschluss findet ein Vergleich zwischen einzelnen Bauteilen statt, inwiefern sie bestimmte Modellvorstellungen unterstützen und welche Vor- und Nachteile bei der Realisierung typischer Zufallsgeräte (Würfel, Münze, Kreisel) vorliegen. Unter einem Zufallsexperiment wird hier folgendes verstanden:

Ein Zufallsexperiment (oder auch Zufallsversuch) ist ein Vorgang, der sich, zumindest theoretisch, unter im Wesentlichen unveränderten Bedingungen beliebig oft wiederholen lässt. Der Ausgang des Zufallsexperiments ist nicht vorhersagbar, allerdings sind die möglichen Ergebnisse des Zufallsexperiments vorab bekannt. Es wird davon ausgegangen, dass den möglichen Ergebnissen ω_i aus Ω Wahrscheinlichkeiten $p_i \in [0;1]$ mit $\sum p_i = 1$ zugeordnet werden können. Der Raum aller möglichen Ergebnisse eines Zufallsexperiments wird als Ergebnisraum Ω bezeichnet. Ein mehrstufiges Zufallsexperiment besteht aus mehreren Teilexperimenten von Zufallsversuchen, die nacheinander (oder gleichzeitig) ausgeführt werden. Hat ein Zufallsversuch n Teilexperimente, so wird dies hier als n-stufiger Zufallsversuch bezeichnet und durch die Produktmenge $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ beschrieben. Ein Ergebnis eines n-stufigen Zufallsexperiments ist dabei ein n-Tupel $(\omega_1, \omega_2, \ldots, \omega_n)$ mit ω_i aus Ω_i.

Erläuterung der einzelnen Bauteile:

Den Erläuterungen der einzelnen Bauteile wird zunächst jeweils ein Beispiel vorangestellt, um die Funktionsweise des Bauteils zu verdeutlichen. Daraufhin wird die Modellierung im Bauteil beschrieben und anschließend werden die allgemeinen Möglichkeiten und Grenzen des Bauteils aufgezeigt. Eine besondere Rolle spielt dabei jeweils die Beschriftung der Elemente des Bauteils, da durch diese der Ergebnisraum Ω definiert wird. Es folgen kognitive Überlegungen zur Benutzung des Bauteils aus Nutzersicht.

Bauteil Box:

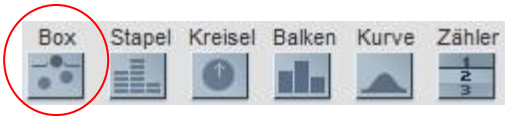


Abbildung 11: Bauteil Box

Ein Standardbeispiel der Stochastik ist das Ziehen von Kugeln aus einer Urne. Nehmen wir an, in einer Urne liegen vier rote und fünf blaue Kugeln wie in Abbildung 12 und es soll einmal daraus gezogen werden.

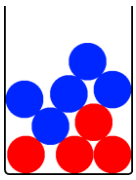


Abbildung 12: Eine Urne mit 4 roten und 5 blauen Kugeln

Ein zugehöriger Wahrscheinlichkeitsraum $(\Omega, \mathcal{P})$ kann hier gegeben sein durch neun gleichwahrscheinliche Kugeln, von denen vier rot und fünf blau gekennzeichnet sind, also den Ergebnisraum Ω = {rot, rot, rot, rot, blau, blau, blau, blau, blau} mit der Wahrscheinlichkeitsverteilung P(rot) = 4/9 und P(blau) = 5/9.

In der Zufallsmaschine lässt sich dies wie folgt abbilden:

Abbildung 13: Repräsentation einer Urne durch das Bauteil Box

Für das Beispiel liegen neun passend beschriftete Kugeln, vier rote und fünf blaue, in der Box in Abbildung 13. Die Kugeln als Elemente des Bauteils Box können beliebig beschriftet werden, allerdings nicht farblich geändert werden. Das Ziehen aus der Urne wird repräsentiert durch die Anzahl Ziehungen. Über die Anzahl der Durchgänge wird nun modelliert, wie häufig das Zufallsexperiment durchgeführt werden soll. Das Ziehen aus der Urne wird per Standardeinstellung fünfmal durchgeführt, wobei jeder Durchgang bei der Ausgangssituation wie in Abbildung 13 startet. Die Visualisierung in der Zufallsmaschine zu dieser Situation entspricht recht genau dem mentalen Modell, das ein Benutzer mental bilden könnte.

Das Bauteil Box entspricht einer stochastischen Urne, die mit maximal 10.000 Kugeln gefüllt werden kann. Mit dem + und – Button werden Kugeln hinzugefügt bzw. gelöscht. Die Kugeln können beliebig beschriftet werden. Die Beschriftung der Kugeln in der Zufallsmaschine definiert den Ergebnisraum Ω. Diskrete Verteilungen können hier einfach abgebildet werden, gleichverteilte oder nicht gleichverteilte Prozesse lassen sich gleichermaßen modellieren. Durch die maximale Befüllung von 10.000 beliebig beschrifteten Kugeln lassen sich Zufallsexperimente mit entsprechend vielen Ausgängen realisieren. Zahl- oder Buchstabenbereiche lassen sich komfortabel über das Eingabefenster … eingeben. Weiterhin kann beim mehrfachen Ziehen aus der Box eingestellt werden, ob mit oder ohne Zurücklegen gezogen werden soll. Beim Starten des Simulationsvorgangs (Klick auf Play) wird zufällig eine Kugel ausgewählt, ihr Wert entspre-

chend der Ziehung unter den Merkmalsnamen geschrieben und sie je nach Einstellung zurückgelegt oder herausgenommen.

Die Beschriftung der Kugeln erfolgt durch einzelnes Anklicken der Kugeln. Benutzerfreundlich ist die Funktion, einen Wertebereich über den Button … eingeben zu können. Soll eine Zahlen- oder Buchstabenfolge eingegeben werden, so kann dies direkt geschehen. Soll beispielsweise eine „Lotto“-Urne mit den Zahlen von 1 bis 49 modelliert werden, so kann dies nutzerfreundlich über die Eingabe „1-49“ geschehen. Alle zuvor in der Box liegenden Kugeln werden durch entsprechend beschriftete 49 Kugeln ersetzt. Durch diese Funktion wird die technische Eingabe von umfangreichen Ergebnisräumen für den Nutzer vereinfacht und eine fehleranfällige Einzeleingabe wird vermieden.

Die Beschriftung weiterer, manuell über das Plus-Zeichen hinzugefügter Kugeln orientiert sich an der letzten Kugel (d. h. in einer zeilenweise von links unten nach rechts oben orientierten Betrachtung). Ist diese letzte Kugel eine beliebige Zahl, so wird eine neue Kugel automatisch mit der nächstgrößeren natürlichen Zahl beschriftet. Bei allen anderen Zeichen auf der letzten Kugel wird eine neue Kugel mit einem Kleinbuchstaben, beginnend bei a, beschriftet. Diese intelligente Halbautomatik bei der Beschriftung der Kugeln folgt der Philosophie von TinkerPlots, bei der zunächst alle Variablennamen automatisch vergeben werden. Für viele Modellierungen ist dies eine hilfreiche Funktion.

Bei einer großen Anzahl an Kugeln (oder Zeichen pro Kugel) ist das Beschriften jedoch recht mühsam, sofern nicht durch die Bereichsfunktion eine Folge eingegeben werden kann. Liegen die Elemente in einer Liste (z. B. in Excel) oder als Textdatei in einzelnen Zeilen vor, so lassen sie sich einfach durch Kopieren und Einfügen als Kugeln in die Box legen. Dies ermöglicht mit dem Bauteil Box das Ziehen einer Stichprobe aus einem vorliegenden Datensatz (vgl. dazu Abschnitt 3.5.2.3 zur Simulation als Stichprobenziehung).

Mit Hilfe des Bauteils Box wird die Modellvorstellung des Nutzers einer stochastischen Urne unterstützt. Wie durch Sedlmeier (1999) beschrieben, ist die Grundvorstellung einer Urne sehr universell und mit seinen Möglichkeiten ist das Bauteil Box für den Nutzer sehr mächtig einsetzbar. Es können hierdurch Zufallsexperimente mit endlichen Ergebnismengen modelliert werden, die sich auf ein Urnenmodell zurückführen lassen.

Bauteil Stapel:

Abbildung 14: Bauteil Stapel

Das Bauteil Stapel (Abbildung 14) ist eine andere Darstellungsart des Bauteils Box und bedient somit ebenfalls die Modellvorstellung einer Urne in einem anderen Repräsentationsformat. Als Anschauungsbeispiel dient wieder das Ziehen von farbigen Kugeln aus einer Urne, diesmal allerdings mit einer größeren Anzahl an Kugeln zur Verdeutlichung der Vorteile dieses Bauteils gegenüber dem Bauteil Box. In einer Urne sollen nun 10 rote Kugeln, 10 blaue Kugeln und 20 schwarze Kugeln liegen. Ein zugehöriger Wahrscheinlichkeitsraum $(\Omega, \mathcal{P})$ kann durch 40 gleichwahrscheinliche Elemente gegeben sein, die entsprechend der Farbe markiert sind, also

$\Omega =$ {rot, rot, rot, rot, rot, rot, rot, rot, rot, rot, blau, blau, blau, blau, blau, blau, blau, blau, blau, blau, schwarz}

und der Wahrscheinlichkeitsverteilung P(rot) = P(blau) = $\frac{10}{40}$ und P(schwarz) = $\frac{20}{40}$. Diese Situation ließe sich analog zum obigen Beispiel mit dem zuvor beschriebenen Bauteil Box in der Zufallsmaschine abbilden. Dazu müsste für 40 Kugeln einzeln die Beschriftung erfolgen, entweder direkt, d. h. jede Kugel einzeln anklicken und beschriften, oder über den Umweg mit einer Liste. Das ist möglich mit dem Bauteil Box, ist aber zeitaufwändig und lässt sich mit dem Bauteil Stapel leichter realisieren (Abbildung 15).

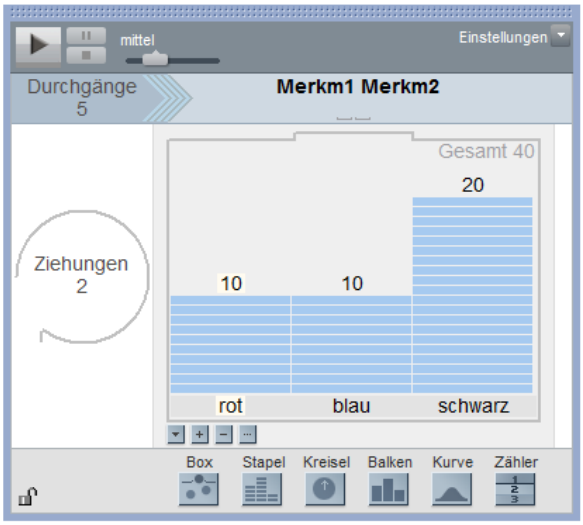


Abbildung 15: Eine Urne mit vielen gleichartigen Elementen repräsentiert durch das Bauteil Stapel

Für das Beispiel sind in der Zufallsmaschine in Abbildung 15 drei Stapel gebildet und mit der entsprechenden Farbe beschriftet. Der „rote“ Stapel hat 10 Elemente, der „blaue“ ebenfalls und der „schwarze“ Stapel hat 20 Elemente. Somit bildet die Beschriftung der Stapel den Ergebnisraum Ω ab.

Die Anzahl der Elemente in einem Stapel (z. B. 10 über dem mit rot beschrifteten Stapel in Abbildung 15) kann über das Pfeilmenü eingeblendet werden, ansonsten kann durch Klicken die Anzahl der Elemente in einem Stapel erhöht oder verringert werden. Ist eine genaue Anzahl erwünscht, so ist dies mit der Einblendung der Anzahlen leichter. Zusätzlich bekommt man auch noch die Information, wie viele Elemente insgesamt vorhanden sind (rechts oben in Bereich 1b der Zufallsmaschine in Abbildung 15). Entsprechend der Stapelmetapher werden hier also Stapel gebildet, in denen jedes Element die gleiche Beschriftung trägt.

Im Bauteil Stapel können maximal 10.000 Stapel gebildet werden. Dabei kann die Elementanzahl über jedem einzelnen Stapel beliebig variiert werden, die jeweilige absolute Anzahl kann eingeblendet werden. Prozente oder Anteile können für einzelne Stapel nicht angezeigt werden. Mit der Beschriftung verhält es sich wie bei der Box. Beim Starten wird zufällig ein Element eines Stapels ausgewählt und dessen Wert unter das entsprechende Merkmal geschrieben. Es kann mit oder ohne Zurücklegen gezogen werden. Benannte Stapel dürfen auch 0 Elemente enthalten und verschiedene Stapel dürfen gleich benannt werden, ohne dass sie automatisch zu einem Stapel zusammengefasst werden.

Gehören zu einem Zufallsexperiment viele gleichartige Ergebnisse, wie oben die gleichfarbigen Kugeln, so bietet das Bauteil Stapel einen Vorteil für den Nutzer bei der Visualisierung gegenüber dem Bauteil Box. Das Bauteil Stapel unterstützt die folgende Vorstellung der Situation der Urne mit 40 Kugeln. Stellt man sich dies als Urne vor, in der die Kugeln nach Farben geordnet sind, so kann man sich 10 rote Kugeln auf einem Stapel, 10 blaue Kugeln auf einem Stapel und 20 schwarze Kugeln auf einem Stapel vorstellen. Repräsentiert in der Zufallsmaschine sieht dies aus wie in Abbildung 15, dem Bild des Stapels folgend sind allerdings Blöcke und keine Kugeln aufgestapelt. Der Aufwand für den Nutzer, eine Urne zu erstellen, in der 400 oder sogar 4000 Kugeln in drei Farben liegen, ist mit dem Bauteil Stapel kaum höher als für 40 Kugeln, wodurch das Bauteil bei der Benutzung vorteilhaft ist.

Mit Hilfe des Bauteils Stapel wird ebenfalls die Modellvorstellung einer Urne unterstützt, ähnlich wie beim Bauteil Box. Allerdings wird hier das Modell unterstützt, in welchem die Elemente der Urne geordnet bwz. „gestapelt“ sind. Zufallsexperimente mit endlichen Ergebnismengen können hiermit realisiert werden. Es lassen sich diskrete Verteilungen realisieren, allerdings lassen sich hier einfacher, d. h. in der Bedienung der Software schneller, als bei der Box Zufallsexperimente mit nicht gleichverteilten Ergebnismengen modellieren. Durch die abgebildeten Stapel repräsentiert dieses Bauteil das Modell einer Urne, die in Stapeln gleichartiger Elemente vorsortiert ist. Auf die Ziehung hat diese Ordnung keinen Einfluss, es wird genauso zufällig ein Element aus-

gewählt wie bei der Box. Sowohl das Bauteil Stapel als auch das Bauteil Box beinhalten eine feste Anzahl diskreter Elemente.

Bauteil Kreisel

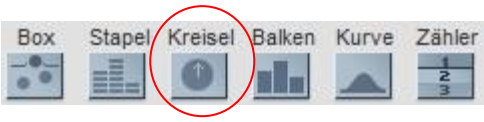


Abbildung 16: Bauteil Kreisel

Ein Glücksrad wie das in Abbildung 17 ist ein weiteres typisches Zufallsgerät. Das Drehen des Glücksrads aus Abbildung 17 soll nun mit der Zufallsmaschine simuliert werden.

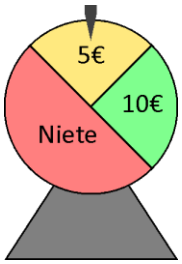


Abbildung 17: Ein Beispielglücksrad

Der Aufbau des Glücksrad ist wie folgt: In einem Viertel der Fälle bekommt man beim Drehen des Glücksrads 5€, in einem weiteren Viertel der Fälle erhält man 10€ und in der Hälfte der Fälle erhält man eine Niete.

Mithilfe des Bauteils Kreisel (Abbildung 16) kann dieses Glücksrad direkt in TinkerPlots abgebildet werden (Abbildung 18).

Abbildung 18: Ein Glücksrad in der Zufallsmaschine

In Abbildung 18 findet sich eine 1:1 Umsetzung des Beispielglücksrads in die Zufallsmaschine mithilfe des Bauteils Kreisel. Wie oben beschrieben ist zunächst keine weitere Abstraktion nötig, um das gegebene Glücksrad durch das Bauteil Kreisel abzubilden.

Beim Benutzen der Zufallsmaschine aus Abbildung 18 erscheint ein Zeiger, der zufällig in einem Segment stehen bleibt. Die Wahrscheinlichkeit, dass ein bestimmtes Segment ausgewählt wird, ist proportional zu dessen Anteil an der Gesamtfläche. Durch die Beschriftung der Segmente des Kreisels wird der Ergebnisraum definiert, in diesem Beispiel ist dieser:

$$\Omega = \{5\ €, 10€, \text{Niete}\}.$$

Betrachtet man zu diesem Beispiel die Zufallsgröße X: „Gewinn nach einmaligem Drehen des Glücksrads“, so liegt die folgende Wahrscheinlichkeitsverteilung zugrunde:

Tabelle 2: Wahrscheinlichkeitsverteilung der Zufallsgröße „Gewinn“

k (Gewinn in €)	0	5	10
P(X=k)	0,5	0,25	0,25

Dies ist die Wahrscheinlichkeitsverteilung, unter der simuliert wird. Durch den Anteil der Segmente in Abbildung 18 werden die möglichen Ausgänge des Zufallsexperiments mit den zugehörigen Wahrscheinlichkeiten dargestellt. Die Wahrscheinlichkeitsverteilung in Tabelle 2 stellt in der Notation eine Verkürzung des Beispielglücksrads aus Abbildung 17 dar, indem die „Niete“ durch „0“ dargestellt wird und das Eurosymbol ausgelassen wird. Auch dies lässt sich in TinkerPlots abbilden (Abbildung 19). Zusätzlich sind die Anteile für die einzelnen Segmente eingeblendet, um die Wahrscheinlichkeitsverteilung transparent darzustellen.

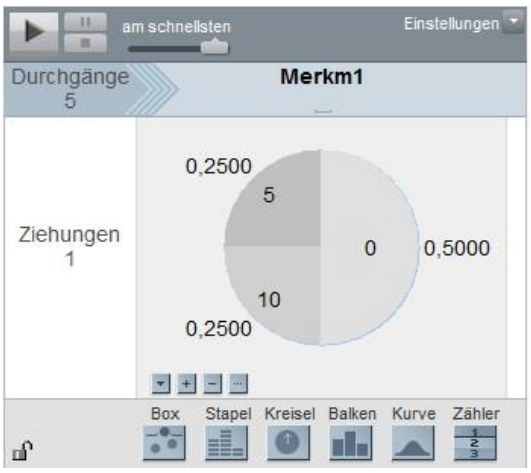


Abbildung 19: Verkürzte Darstellung des Glücksrads durch die Zufallsmaschine

Mit der Darstellung aus Abbildung 19 lässt sich beim Auswerten der simulierten Daten leichter weiterarbeiten als mit der beispielnahen Darstellung aus Abbildung 18 (siehe dazu Abschnitt 3.6).

Das Bauteil Kreisel entspricht einem idealen Kreisel, für den maximal 10.000 Segmente festgelegt werden können. Dabei ist die Größe jedes einzelnen Segments als Anteil oder

Prozentangabe individuell einstellbar (in der Summe muss sich 1 bzw. 100 Prozent ergeben). Die entsprechende Prozentzahl oder der entsprechende Anteil eines Kreiselsegments kann eingeblendet werden, dabei kann der Anteil auf vier Nachkommastellen genau angegeben werden. Allgemein werden zu einem endlichen Ergebnisraum mit

$$\Omega = \{x_1, x_2, \ldots, x_n\} \text{ und } P(x_1) = p_1, P(x_2) = p_2, \ldots, P(x_n) = p_n, \text{ mit } p_1 + p_2 + \cdots + p_n = 1$$

die Ergebnisse x_i (i = 1,…,n) eines Zufallsexperiments durch die Segmente des Kreisels repräsentiert. Dabei ist die Wahrscheinlichkeit p_i (i = 1,…,n) für ein bestimmtes Ergebnis proportional zu seinem Segmentanteil am gesamten Kreisel.

Die Beschriftung der Segmente unterliegt den gleichen Möglichkeiten, wie die der beiden vorherigen Bauteile und definiert ebenso wie dort den Ergebnisraum Ω. Die Einstellmöglichkeit „ohne Zurücklegen“ ist bei diesem Bauteil nicht vorhanden. Damit folgt das Bauteil Kreisel seinem realen Gegenstück, bei dem ein „Herausschneiden“ einzelner Segmente auch nicht möglich ist, nachdem sie einmal erdreht wurden.

Mit Hilfe des Bauteils Kreisel wird für den Nutzer die Modellvorstellung eines idealen Glücksrads unterstützt. Allerdings ist hiermit keine überabzählbare Ergebnismenge produzierbar, wie die Repräsentation es nahe legt. Wie auch mit den beiden vorherigen Bauteilen lassen sich hiermit nur endliche Ergebnismengen erzeugen. Für den Nutzer sind Segmente vorgegeben, von denen maximal 10.000 erstellt werden können, wodurch keine überabzählbare Ergebnismenge erzeugt werden kann. Diskrete Wahrscheinlichkeitsverteilungen lassen sich hiermit realisieren, durch die leichte Veränderbarkeit der Segmentgrößen ist dieses Bauteil gut für gleichverteilte und nicht gleichverteilte Verteilungen geeignet. Alle Situationen, die durch das mentale Modell eines Glücksrads als weitere Grundvorstellung des Nutzers zu einem typischen Zufallsgerät beschrieben werden, können mit dem Bauteil Kreisel direkt in TinkerPlots übertragen werden. Allerdings kann sich das Einstellen der Segmentgrößen und somit der Wahrscheinlichkeit für ein bestimmtes Segment als schwierig herausstellen. Solange die Segmentanteile durch Ziehen der Grenzen „verschoben“ werden, ändern sich nur die zwei beteiligten Segmentgrößen wie gewünscht. Wird allerdings eine genaue Angabe des Segmentanteils beispielsweise in Form von Prozenten gemacht, so ändern sich oft beide benachbarten Segmente in ihrem Anteil mit. Diese Änderung ist manchmal für den Nutzer nicht nachvollziehbar. Strategien, die häufig zur gewünschten Verteilung führen sind entweder die Segmente entgegen dem Uhrzeigersinn sukzessive in der Größe anzupassen oder die Segmente geordnet nach ihrer Größe in absteigender Reihenfolge anzupassen. Durch diese undurchschaubare Größenänderung gerade bei Kreiseln mit vielen Segmenten ist die Handhabung dieses Bauteils für den Nutzer jedoch manchmal schwierig.

Bauteil Balken

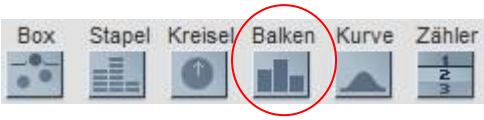


Abbildung 20: Bauteil Balken

Hat man eine Situation mit vielen verschiedenen Ausgängen, die zusätzlich unterschiedliche Wahrscheinlichkeiten haben, so bietet das Bauteil Balken (Abbildung 20) neben dem Bauteil Kreisel eine gute Unterstützungsmöglichkeit für die Modellierung. Es wird das Beispiel des „Schweinchen-Spiels“ betrachtet, bei dem ein schweinchenförmiger Würfel geworfen wird. Dabei ist die Landeposition des Würfels entscheidend. Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{P})$ ist durch $\Omega = \{\text{Sau, Suhle, Haxe, Schnauze, Backe}\}$ und der in Abbildung 21 angegebenen Wahrscheinlichkeitsverteilung gegeben.

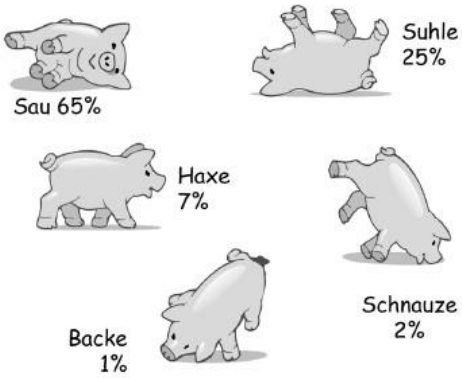


Abbildung 21: Verschiedene Landepositionen mit dem „Schweinchenwürfel" (entnommen aus: http://btmdx1.mat.uni-bayreuth.de/smart/buch/smart_sinus_j05_angabe.pdf, abgerufen am 29.06.2018)

Tabelle 3: Tabellarische Darstellung der Landepositionen mit zugehöriger Wahrscheinlichkeit beim Schweinchenwürfel

Landeposition	Sau	Suhle	Haxe	Schnauze	Backe
Zugehörige W’keit	65%	25%	7%	2%	1%

Um den Wurf mit einem solchen Schweinchenwürfel zu simulieren, eignet sich die Modellierung in der Zufallsmaschine mit dem Bauteil Balken.

Ebenso wie beim vorherigen Bauteil Kreisel werden hier Flächen betrachtet und nicht diskrete Objekte wie bei den Bauteilen Box oder Stapel. Der Ergebnisraum wird wiederum durch die Beschriftung der einzelnen Balken abgebildet. Genaue Prozentwerte oder Anteile können wie in Abbildung 22 zu sehen zusätzlich eingeblendet werden. Dieses Bauteil visualisiert die unterschiedlichen Wahrscheinlichkeiten für die einzelnen Landepositionen des Schweinchenwürfels aus Tabelle 3 sehr gut und hilft somit bei der kognitiven Erfassung der Situation. Durch die visuelle Abbildung der unterschiedlichen

Wahrscheinlichkeiten für die einzelnen Landepositionen ist die Einschätzung der Wahrscheinlichkeiten für den Nutzer auf einen Blick gegeben.

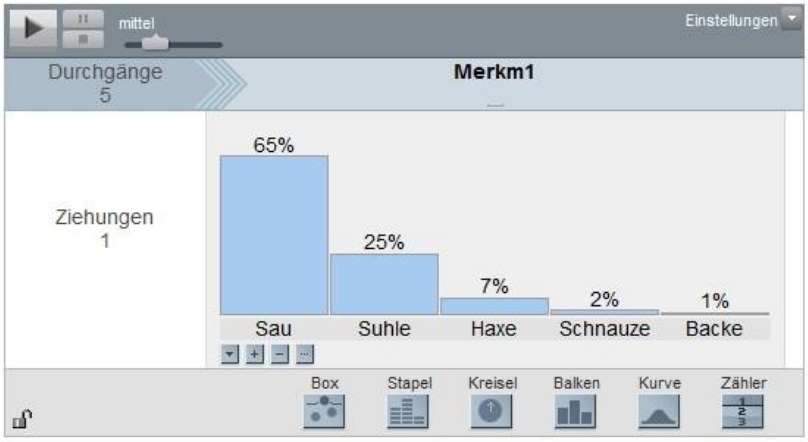


Abbildung 22: Modellierung des „Schweinchenwürfels“ mit dem Bauteil Balken

Für die Simulation einer Situation, die aus unterschiedlichen Ausgängen mit unterschiedlichen Wahrscheinlichkeiten besteht, bietet das Bauteil Balken eine gute Unterstützung. Liegt eine Wahrscheinlichkeitsverteilung in Tabellenform vor (vgl. Tabelle 3), so ist die kognitive Übertragung in ein Säulendiagramm einigermaßen naheliegend. Diese Übertragung wird durch das Bauteil Balken unterstützt. Im Bauteil (Verteilungs-)Balken gibt es für jedes Element einen Balken, der in der Höhe angepasst werden kann. Die jeweilige Prozentzahl oder der Anteil der einzelnen Balken können eingeblendet werden, in der Summe aller Balken müssen sich auch hier 100% bzw. 1 ergeben. Die Wahrscheinlichkeit für einen einzelnen Balken, und somit zu einem der möglichen Ergebnisse, ist proportional zu seiner Fläche. Auch hier sind maximal 10.000 Elemente/Balken möglich. Für die Beschriftung gilt das Gleiche wie bisher. Wie beim Bauteil Kreisel ist die Option mit Zurücklegen voreingestellt und kann nicht verändert werden. Auch hier muss bei der Größeneingabe der einzelnen Balken genau hingeschaut werden. Ändert man die Höhe eines Balkens mit der Maus durch Ziehen, so ändern sich die anderen Balken nicht in ihrer absoluten Höhe, allerdings in ihrem Anteil. Wird eine prozentuale Angabe der Höhe eines Balkens geändert, so ändert sich die Höhe des Balkens rechts daneben dahingehend, dass beide zusammen mit den zusätzlich vorhandenen Balkenhöhen wieder 100 Prozent ergeben. Können aufgrund der zu geringen Höhe des linken Balkens insgesamt nicht 100 Prozent erreicht werden, so wird die Höhe beim gerade geänderten Balken automatisch auf eine niedrigere Zahl angepasst. Bei einer Änderung der Balkenanteile muss der Nutzer also sehr genau darauf achten, welche Balken in ihrem Anteil mitgeändert werden und prüfen, ob dies der gewünschten Verteilung entspricht. Wie beim Bauteil Kreisel ist dadurch die Handhabung des Bauteils für den Nutzer beschwerlich.

Dieses Bauteil ähnelt dem Kreiselbauteil bei anderer optischer Repräsentation. Kommt beim Benutzer die kognitive Vorstellung einer Wahrscheinlichkeitsverteilung in der Repräsentation durch ein Säulendiagramm zum Tragen, so unterstützt dieses Bauteil die visuelle Repräsentation. Auf der anderen Seite kann gerade diese Balkendarstellung von Wahrscheinlichkeiten als Flächen einen kognitiven Vorteil für den Benutzer darstellen, da es sich hierbei auch um eine neue Art der Darstellung einer Wahrscheinlichkeitsverteilung handeln kann.

Allerdings sind hier, wie auch beim Bauteil Kreisel, nur diskrete Ergebnisräume möglich. Zur direkten Unterstützung fehlt außerdem eine Skala, durch die die Höhe der Balken ablesbar wäre. Ohne Einblenden der Prozente oder Anteile ist die zugrundeliegende Verteilung schwer erkennbar und somit für den Nutzer schwer einschätzbar (siehe Abbildung 23). Dies gilt aber in gleicher Weise für das Bauteil Kreisel ohne Einblendung der Anteile. Für Situationen, in denen Zufallsexperimente mit nicht gleichverteilten Grundmengen durch Anteile gegeben sind, eignet sich dieses Bauteil ebenso wie der Kreisel. Es ist nun am Nutzer zu entscheiden, welches Repräsentationsformat für ihn das geeignetere ist.

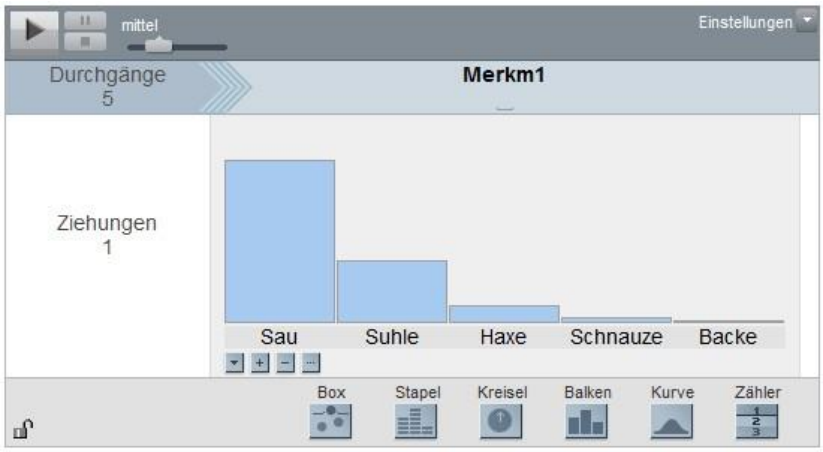


Abbildung 23: Repräsentation des Schweinchenwürfels ohne Prozentangabe bei den Balken

In den Bauteilen Kreisel und Balken werden Wahrscheinlichkeiten über Flächenanteile visualisiert, statt über Anzahlen von Elementen wie bei den Bauteilen Box (Kugeln) und Stapel (Blöcke). Trotzdem handelt es sich bei allen vier Bauteilen um Möglichkeiten, um diskrete Ergebnisräume abzubilden.

Bauteil Kurve

Abbildung 24: Bauteil Kurve

Wie bereits oben erwähnt, kann die Zufallsmaschine von TinkerPlots als „Datenfabrik“ aufgefasst werden. Bei dem folgenden Beispiel ist dies eine hilfreiche Sicht, um die Vorstellung zu fördern, dass Personen mit bestimmten Eigenschaften „hergestellt“ werden. Man kann so „produzierten“ Personen eine Eigenschaft verleihen, zum Beispiel eine Körpergröße. Nimmt man nun die weibliche Bevölkerung, so ist die Körpergröße in ihr annähernd normalverteilt (Abbildung 25).

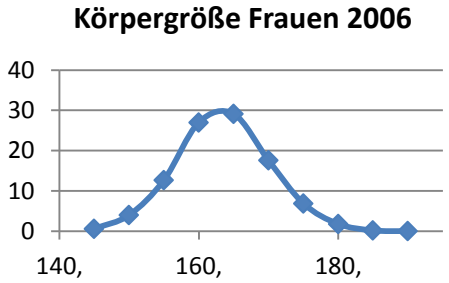


Abbildung 25: Prozentuale Verteilung der Körpergröße in cm der weiblichen Bevölkerung 2006[13]

Möchte man in TinkerPlots beispielsweise fünfzig Personen „herstellen“, deren Verteilung des Merkmals Körpergröße normalverteilt ist, so lässt sich dies mit dem Bauteil Kurve (Abbildung 24) am besten realisieren (Abbildung 26).

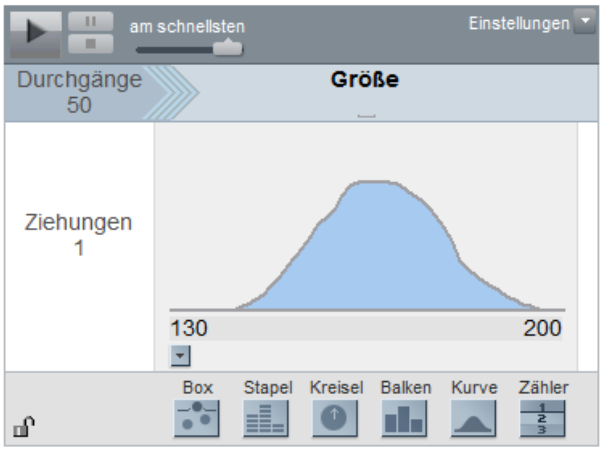


Abbildung 26: Normalverteilung der Körpergröße nachgebildet mit dem Bauteil Kurve

In Abbildung 26 ist mit dem Bauteil Kurve von Hand eine annähernd glockenförmige Verteilung in den Grenzen von ca. 140 bis 200 gezeichnet als Annäherung an die Normalverteilung der Körpergröße von Frauen in cm. Bei der Einzeichnung der Kurve in Abbildung 26 hat eine Orientierung am Mittelwert als Hochpunkt und der Breite der Verteilung, genommen aus Abbildung 25, stattgefunden. Mit dieser Modellierung kön-

[13] Datenquelle: DIW Berlin. (n.d.). Verteilung der Körpergrößen nach Geschlecht im Jahr 2006. In Statista - Das Statistik-Portal. Zugriff am 29. Juni 2018, von http://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/.

nen nun ungefähr normalverteilte, zufällige Größen erzeugt werden, wie in Abbildung 27 für fünfzig Personen zu sehen ist.

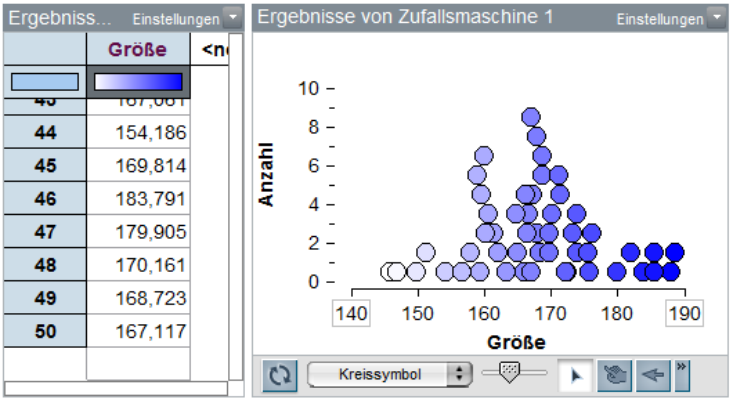


Abbildung 27: Zufällige Größen für 50 Personen

Dabei werden in der Tabelle Werte mit insgesamt sechs Ziffern angezeigt.

Das Bauteil (Verteilungs-)Kurve weicht von den vorherigen ab. Bisher konnten nahezu beliebige (maximal mit zehntausend Elementen) diskrete Ergebnismengen modelliert werden. Mit dem Bauteil Kurve wird nun Bezug genommen auf die Idee der kontinuierlichen Dichtefunktion: Jedem Punkt eines reellen Intervalls [a; b] wird ein Wahrscheinlichkeitsdichtewert zugeordnet. In der Softwareumsetzung allerdings ist das kontinuierlich erscheinende Intervall [a; b] wiederum diskret und operiert auf rationalen Zahlen. Da über das Pixelraster des Monitors mit diesem Bauteil im Endeffekt eine Rechteckverteilung erzeugt wird, kann mit Hilfe der Inversionsmethode als Verfahren zur Erzeugung von Wahrscheinlichkeitsverteilungen die letztlich ausgegebene Zufallszahl erzeugt werden. Die Fläche unter der Kurve, die der Benutzer einzeichnet, beträgt Eins, anteilig repräsentiert die Höhe über jedem Punkt dessen Wahrscheinlichkeit. Wird die Kurve durch den Nutzer verändert, so ändert sich gegebenenfalls die Darstellung, die Fläche unter der Kurve wird jedoch im Hintergrund stets auf Eins skaliert, auch wenn in keinem Fall Skalierungen angezeigt werden. Der Kurvenverlauf kann durch den Benutzer einzig durch Ziehen des Mauszeigers bestimmt werden, exakte Nachbildungen beispielsweise der Normalverteilungen sind damit nicht möglich.

Durch das Bauteil Kurve werden Zufallswerte mit maximal sieben Ziffern generiert, so dass auch dieses Bauteil durch die Ausgabemöglichkeit auf einer diskreten Ergebnismenge operiert. Bei der Eingabe der Grenzen des Intervalls ist zu beachten, dass es sich um rationale Zahlen handeln muss. Wird die linke Zahl größer als die rechte gewählt, so werden die Grenzen automatisch vertauscht. Der Verlauf der Kurve ist beliebig veränderbar, allerdings ist es durchaus mühsam, den gewünschten Kurvenverlauf einzuzeichnen, da dies ausschließlich mit dem Mauszeiger händisch möglich ist.

Aus Nutzersicht ist dies ein durchaus schwer zu bedienendes Bauteil im Vergleich mit den restlichen Bauteilen. Zum einen ist es mühsam, den gewünschten Kurvenverlauf einzuzeichnen, zum anderen ist durch die fehlende Skalierung der Höhe ein genaues Arbeiten mit dem Bauteil kompliziert. Es kann hiermit zum Beispiel eine angenäherte Normalverteilung modelliert werden, allerdings kann diese nicht über die Standardparameter Mittelwert und Standardabweichung eingegeben werden. Durch die freie Zeichnung sind auch sämtliche weiteren denkbaren Kurvenverläufe möglich. Das heißt, dass dieses Bauteil ein großes Maß an Freiheit für den Nutzer bedeutet, gleichzeitig aber auch ein hohes Maß an Ungenauigkeit beinhaltet, da das Kurvenzeichnen durch die Maus nur rudimentär möglich ist. Aus diesem Grund wird das Unterstützungspotential dieses Bauteils als geringer eingestuft als bei den anderen Bauteilen.

Bauteil Zähler:

Abbildung 28: Bauteil Zähler

Das letzte Bauteil, der Zähler (Abbildung 28), nimmt eine Sonderrolle ein. Hierbei handelt es sich nicht um einen Zufallsgenerator, wie bei den vorherigen Bauteilen, sondern um einen systematischen Zähler.

Betrachtet wird als Beispiel zur Funktionsweise dieses Bauteils das Erzeugen der Kombinationsmöglichkeiten zum doppelten Würfelwurf. Es ist zu beachten, dass hiermit keine Zufallszahlen erzeugt werden.

Abbildung 29: Repräsentation der Kombinationsmöglichkeiten beim doppelten Würfelwurf durch das Bauteil Zähler

Beide Bauteile in der Zufallsmaschine in Abbildung 29 (links) sind Zähler mit sechs Feldern beschriftet mit den Zahlen von Eins bis Sechs. Jedes dieser Bauteile stellt ein „Zählrad“ dar und soll hier einen Würfel repräsentieren. Nun werden nacheinander in jedem Zug systematisch die sichtbaren Felder „gezogen“. Nach einer solchen Ziehung geschieht eine systematische Drehung des hintersten Zählers (d. h. desjenigen, der sich am weitesten rechts befindet) um ein Feld, bis dieser einmal durchgelaufen ist. Anschließend wird der vordere Zähler um ein Feld weitergedreht und wieder mit jedem

Feld des hinteren Zählers kombiniert. Dies wird solange fortgeführt, bis die Anzahl der Durchgänge erreicht ist. Bei zwei Zählern mit jeweils sechs Feldern gibt es sechs mal sechs Kombinationsmöglichkeiten, die durch eine genügend große Angabe bei den Durchgängen sicher alle „erdreht“ werden, da hier kein Zufallsprozess am Werke ist. Mit einer geeigneten Darstellung in der Graphik (Abbildung 29, rechts) lassen sich die Kombinationsmöglichkeiten systematisch darstellen.

Auch bei diesem Bauteil können maximal 10.000 Einträge pro Zähler vorgenommen werden, bei beliebiger Beschriftung. Hier ist nur Ziehen mit Zurücklegen möglich.

Mit Hilfe dieses Bauteils allein lassen sich keine Zufallsprozesse abbilden, wohl kann es aber ein Hilfsmittel zur Modellierung bestimmter Prozesse sein (z. B. zur Randomisierung, vgl. Abschnitt 3.5.2.4). Für sich genommen eignet sich dieses Bauteil beispielsweise zur Darstellung von Ergebnisräumen, da sukzessive jedes Feld des Rades erdreht wird, vorausgesetzt, eine entsprechend hohe Anzahl der Durchgänge ist angegeben.

Beispiel: Modellierung typischer Zufallsgeräte: Würfel, Münze und Urne

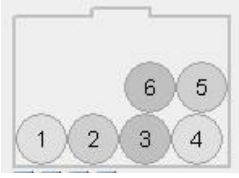
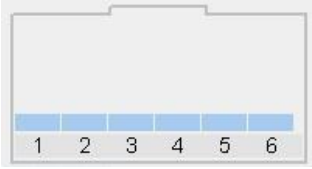
Bis hierhin wurden die Bauteile der Zufallsmaschine vorgestellt. Dabei sind die ersten fünf Bauteile Box, Stapel, Kreisel, Balken und Kurve Bauteile, mit denen Zufallswerte generiert werden können. Hinter dem sechsten Bauteil, dem Zähler, steht kein Zufallsgenerator.

Den Bauteilen Box und Stapel liegt in gewisser Weise die Modellvorstellung einer Urne zugrunde, dem Bauteil Kreisel liegt die Modellvorstellung eines Kreisels oder eines Glücksrads zugrunde und mit dem Bauteil Balken kann eine Repräsentation von Wahrscheinlichkeiten durch die Modellvorstellung eines Säulendiagramms erfolgen. Mit diesen vier Bauteilen lassen sich diskrete Ergebnismengen mit maximal zehntausend Elementen abbilden. Dem Bauteil Kurve liegt die Modellvorstellung einer stetigen Wahrscheinlichkeitsverteilung zugrunde, allerdings ist auch dessen Realisierung in der Software auf eine diskrete Ergebnismenge beschränkt. Viele Zufallsexperimente lassen sich mit unterschiedlichen Bauteilen realisieren. Dabei kann der Wechsel von einem zu einem anderen Bauteil kognitive Vorteile für den Benutzer haben. Dies wird nun an der Modellierung eines Würfels, einer Münze und einer Urne vorgestellt.

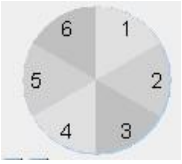
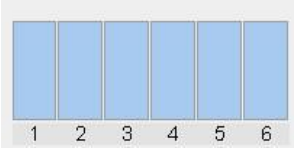
Der Würfel als typisches Zufallsgerät kann in den vier verschiedenen Bauteilen Box, Stapel, Kreisel und Balken realisiert werden. Eine eigene Darstellungsform für das Zufallsgerät des Würfels gibt es in TinkerPlots nicht. Stochastisch liefert jede Repräsentation die gleichen Ergebnisse, nämlich eine zufällige Auswahl aus den Zahlen {1, 2, 3, 4, 5, 6}. Der Unterschied liegt in der Modellvorstellung zu den einzelnen Bauteilen. Kommt beim Nutzer zum Würfelwurf die Vorstellung „Ziehen aus einer Urne mit sechs Kugeln mit Zurücklegen“ zum Tragen, so wird als Bauteil die Box ausgewählt und dieser sechs Kugeln, beschriftet mit 1 bis 6, hinzugefügt. Benutzt man im Bauteil Stapel

sechs Stapel, beschriftet von 1 bis 6 und bestehend aus jeweils einem Element, so wird eine weitere, in gewissem Sinne geordnete, Urnenvorstellung bedient. Da er keine eigene Entsprechung in den TinkerPlots-Bauteilen besitzt und der Würfelvorgang einem Zufallsprozess mit Zurücklegen entspricht, kann sich ein Nutzer zum Würfel ebenso ein mentales Bild eines Glücksrads mit gleichgroßen Segmenten machen. Dies kann in TinkerPlots über das Bauteil Kreisel abgebildet werden, dem sechs gleichgroße Segmente, beschriftet mit 1 bis 6, hinzugefügt werden. Eine weitere Vorstellung, die zu sechs gleichwahrscheinlichen Ergebnissen passt, ist die von gleichgroßen Blöcken, die mit gleicher Wahrscheinlichkeit ausgewählt werden können. Mit dem Bauteil Balken sieht diese Realisierung ähnlich wie durch das Bauteil Stapel aus, mit dem Unterschied, dass hier ein ganzer Stapel zufällig ausgewählt wird, und nicht ein einzelner Block eines Stapels (für nicht gleichverteilte Zufallsgrößen kommt der Unterschied zwischen diesen beiden Bauteilen stärker zum Tragen). Die zwei letzten Bauteile Kurve und Zähler können für dieses Beispiel nicht genutzt werden. Beide Bauteile lassen sich zunächst scheinbar entsprechend anpassen, allerdings gibt das Bauteil Kurve eine reelle Zahl (gerundet auf maximal sechs Nachkommastellen) zwischen den gewählten Grenzen, also 1 und 6, aus.[14] Das Bauteil Zähler lässt sich natürlich auch mit 1 bis 6 beschriften, allerdings handelt es sich hierbei nicht um einen Zufallsgenerator, sondern um einen Zähler, der linear weiter zum nächsten Feld, hier zur nächsten Zahl, gedreht wird. Die Auswahl eines bestimmten Bauteils für die Modellierung des Würfels kann nach den Vorlieben des Benutzers erfolgen. Die folgende Tabelle 4 gibt eine Übersicht zu den Modellierungsmöglichkeiten des Würfels durch die unterschiedlichen Bauteile der Zufallsmaschine.

Tabelle 4: Übersicht zur Modellierung des Würfels

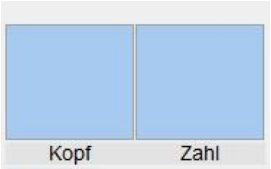
Der Würfel realisiert durch das Bauteil Box zur Vorstellung „Ziehen aus einer Urne“ 	Der Würfel realisiert durch das Bauteil Stapel zur Vorstellung „Ziehen aus einer geordneten Urne“ 

[14] Durch Anpassen des Ergebnismerkmals in der Tabelle, also durch Formeleingabe herbeigeführtes Abschneiden der Nachkommastellen ließe sich sogar mit diesem Bauteil der Würfel realisieren.

Der Würfel realisiert durch das Bauteil Kreisel zur Vorstellung des Würfels als Glücksrad, dem bereits das „Ziehen“ mit Zurücklegen innewohnt 	Der Würfel realisiert durch das Bauteil Balken zur Vorstellung „Repräsentation von Wahrscheinlichkeiten durch ein Säulendiagramm“, bei dem ebenfalls das „Ziehen“ mit Zurücklegen beinhaltet ist 

Auch die Münze als typisches Zufallsgerät hat in den Bauteilen der Zufallsmaschine keine direkte Entsprechung. Wie der Würfel lässt sie sich auch durch die vier Bauteile Box, Stapel, Kreisel und Balken repräsentieren, je nach Vorstellung und Vorliebe des Nutzers. Wie auch beim Würfel wohnt der Münze als Zufallsgerät das „Zurücklegen“ beim Werfen inne, weshalb die Bauteile Kreisel und Balken, obwohl sie vielleicht nicht prototypisch die Vorstellung einer Münze unterstützen, einen gewissen Vorteil bieten. Die nachfolgende Tabelle 5 zeigt die Darstellung der Münze mit den verschiedenen TinkerPlots-Bauteilen.

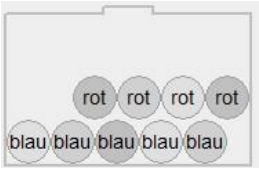
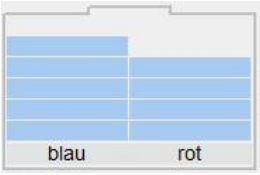
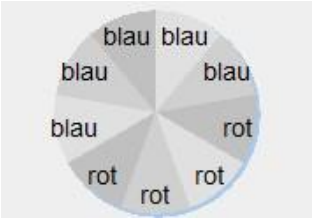
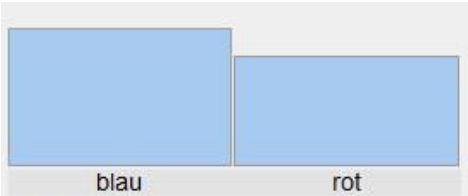
Tabelle 5: Übersicht zur Modellierung einer Münze

Die Münze realisiert durch das Bauteil Box zur Vorstellung „Ziehen aus einer Urne“	Die Münze realisiert durch das Bauteil Stapel zur Vorstellung „Ziehen aus einer geordneten Urne“
Die Münze realisiert durch das Bauteil Kreisel zur Vorstellung als Glücksrad, dem bereits das „Ziehen“ mit Zurücklegen innewohnt	Die Münze realisiert durch das Bauteil Balken zur Vorstellung „Repräsentation von Wahrscheinlichkeiten durch ein Säulendiagramm“, bei dem ebenfalls das „Ziehen“ mit Zurücklegen beinhaltet ist 

Ein weiteres typisches Zufallsgerät ist die Urne. Sie besitzt in dem Bauteil Box ihre direkte Entsprechung und kann nach der Vorstellung des Nutzers mit nahezu beliebig beschrifteten Kugeln gefüllt werden. Einzig Farben können nicht angepasst werden, sondern müssen über Beschriftungen der Kugeln realisiert werden, wodurch die Vorstel-

lung von farbigen Kugeln in einer Urne nicht direkt in TinkerPlots umgesetzt werden kann. Die Vorstellung einer geordneten Urne, in der gleiche Elemente zu Stapeln geordnet sind, wird durch das Bauteil Stapel unterstützt. Auch die anderen beiden Bauteile lassen sich für die Vorstellung einer Urne anpassen, jedoch kann mit ihnen nur das Ziehen mit Zurücklegen realisiert werden, weshalb sie nur bedingt für die Modellierung einer Urne einsetzbar sind. Zur Vorstellung einer Urne bieten sich die Bauteile Kreisel und Balken eher weniger an, aber sie bieten ebenfalls die Möglichkeit. Tabelle 6 zeigt die Modellierung einer Urne mit vier roten und fünf blauen Kugeln in den verschiedenen Bauteilen.

Tabelle 6: Modellierung einer Urne mit vier roten und fünf blauen Kugeln

Die Urne realisiert durch das Bauteil Box  	Die Urne realisiert durch das Bauteil Stapel zur Vorstellung „Ziehen aus einer geordneten Urne“ 
Die Urne realisiert durch das Bauteil Kreisel zur Vorstellung als Glücksrad, dem bereits das „Ziehen“ mit Zurücklegen innewohnt  Eine weitere Umsetzung der Urne durch das Bauteil Kreisel, in der gleiche Elemente durch eine Funktion im Menü zusammengelegt wurden 	Die Urne realisiert durch das Bauteil Balken zur Vorstellung „Repräsentation von Wahrscheinlichkeiten durch ein Säulendiagramm“, bei dem ebenfalls das „Ziehen“ mit Zurücklegen beinhaltet ist Eine weitere Umsetzung der Urne durch das Bauteil Stapel, in der gleiche Elemente durch eine Funktion im Menü zusammengelegt wurden 

Die Vorstellung der unteren beiden Darstellungen in Tabelle 6 zur Beispielurne ist nun eine Verhältnisvorstellung von roten zu blauen Kugeln und nicht mehr die von neun einzelnen Kugeln.

Ein Vorteil der Zufallsmaschine ist es, dass sich in gewissen Grenzen die Inhalte eines Bauteils direkt in ein anderes Bauteil übertragen lassen. Ein Wechsel der Darstellungsart durch den Wechsel der Bauteile ist dadurch gut möglich. Ist eine Modellierung mit einem bestimmten Bauteil erfolgt, so kann einfach ein anderes Bauteil in den Bereich 1b der Zufallsmaschine gezogen werden. TinkerPlots passt die Elemente automatisch an. Beim Wechsel zwischen bestimmten Bauteilen funktioniert dies sehr gut, bei anderen weniger gut. Der Wechsel zwischen den beiden Bauteilen Box und Stapel funktioniert reibungslos, da diese beiden Bauteile sehr ähnlich aufgebaut sind. Beide Bauteile beinhalten Elemente in abzählbarer Häufigkeit, die je nach Bauteil unterschiedlich angeordnet werden. Ebenso gelingt der Wechsel zwischen den Bauteilen Kreisel und Balken sehr gut, da die Elemente bei beiden anteilige Flächen der Gesamtfläche Eins sind (als Kreiselsegmente oder als Balkenflächen). Bei allen anderen Wechseln zwischen Bauteilen treten mehr oder weniger Probleme auf, da die Ergebnismenge jeweils in anderer Weise repräsentiert wird. Von den „Häufigkeits“-Bauteilen Box und Stapel zu den „Flächen“-Bauteilen Kreisel und Balken funktioniert der Wechsel noch gut, auch wenn bei diesem Wechsel gleiche Elemente automatisch zusammengefasst werden, was durchaus eine andere mentale Vorstellung des Nutzers unterstützt. So geschieht beispielsweise der Wechsel vom Bauteil Box in das Bauteil Kreisel für das Beispiel in Tabelle 6 so, dass automatisch die untere Darstellung des Kreiselbauteils erzeugt wird (Abbildung 30). Es wird also eine Vorstellung unterstützt, bei der Kugeln gleicher Farbe entsprechend ihrer Häufigkeit zusammengefasst werden und im Verhältnis zur anderen Kugelfarbe als Segment des Kreisels abgebildet werden. Für den Nutzer hat dies den Vorteil, dass Verhältnisse durch das Zusammenfassen gleicher Ergebnisse direkt sichtbar werden.

Abbildung 30: Wechsel von Bauteil Box in das Bauteil Kreisel

Gleiches gilt für den Wechsel vom Bauteil Stapel in das Bauteil Balken. Schwieriger ist der Wechsel in die andere Richtung, also von den „Flächen“-Bauteilen in die „Häufigkeits“-Bauteile. Sofern in einem Flächenbauteil die einzelnen Elemente gleichverteilt sind, geschieht der Wechsel in eins der anderen Bauteile direkt. Sollten die Elemente jedoch nicht gleichverteilt sein, so versucht das Programm automatisch entsprechend des Anteils eines Elements an der Gesamtfläche auf die Häufigkeit dieses Elements im Verhältnis zu allen anderen Elementen zu schließen. Dies resultiert bei entsprechenden Verhältnissen in einer großen Anzahl an Kugeln oder Stapelelementen in den Bauteilen Box oder Stapel. Abbildung 31 zeigt den Wechsel der Darstellung des Bauteils Stapel mit 57% für rot und 43% für blau in das Bauteil Box, ohne dass weitere Modifikationen vorgenommen wurden. Bei dieser Art von Wechseln wird die Bestückung des Bauteils unübersichtlich für den Nutzer.

Abbildung 31: Wechsel vom Bauteil Balken in das Bauteil Box

Durch den Wechsel eines Bauteils und somit der Darstellungsart der Elemente lassen sich für den Nutzer durchaus sinnstiftende Einsichten gewinnen. Im obigen Beispiel der Urne (Tabelle 6) lässt sich durch den Wechsel des Bauteils Box in das Bauteil Balken das Verhältnis von blauen zu roten Kugeln erkennen. Durch das Einblenden der prozentualen Anteile der Balken wird dieses Verhältnis quantifizierbar. Dies mag bei dem gewählten Beispiel auch vorher schon einsichtig gewesen sein, bei umfangreicheren Urnen kann dies jedoch eine nützliche Funktion sein.

Bereich 2: „Ziehungen (Merkmale)“

Zum zweiten Bereich gehört das Feld „Ziehungen“ zusammen mit den Merkmalsnamen oberhalb des verwendeten Bauteils („Merkm1“ und „Merkm2“ in der Defaulteinstellung) und der Eintrag „Zurücklegen“ aus dem Pfeilmenü unterhalb des Bauteils (Abbildung 32).

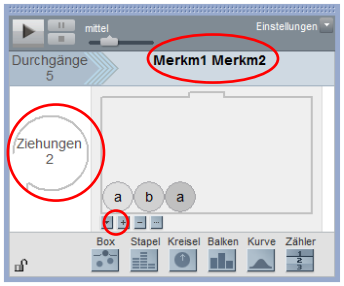


Abbildung 32: Bereich 2 in der Zufallsmaschine

Besteht die Zufallsmaschine aus genau einem Bauteil, so kann die Anzahl der Ziehungen pro Durchgang zwischen 1 und 100 gewählt werden. Wird ein Wert kleiner als Eins eingegeben, so wird die Zahl automatisch auf 1 gesetzt. Wird ein Wert größer als 100 eingegeben, so wird der Wert automatisch auf 100 gesetzt. Unsinnige Eingaben, wie zum Beispiel ein Komma oder Buchstaben, führen wieder automatisch zum Wert 1. Auch dies führt dazu, dass die Zufallsmaschine nutzerfreundlich ist, da unsinnige Eingaben keinen Fehler produzieren, sondern in syntaktisch sinnvolle Eingaben umgewandelt werden. Die Anzahl der Ziehungen ist die Anzahl der einzelnen Merkmale, die erzeugt werden. Diese bekommen die automatisch vergebenen Namen Merkm1 bis Merkm100, welche oben in der Mitte der Zufallsmaschine angezeigt werden. Diese Merkmalsnamen können individuell angepasst werden, wobei nur Buchstaben, Zahlen und Unterstriche verwendet werden dürfen. Außerdem darf ein Merkmalsname nicht mit einer Zahl beginnen.

Wird die Zufallsmaschine aus mehreren Bauteilen gebaut, so gibt es genau so viele Ziehungen wie Bauteilspalten, die Anzahl der Ziehungen kann in diesem Fall nicht variiert werden, da aus jeder Bauteilspalte genau einmal gezogen wird. Entsprechend viele Merkmale werden erzeugt.

Weiterhin gehört zu diesem Bereich der Menüeintrag *Zurücklegen*, über den festgelegt wird, ob mit oder ohne Zurücklegen gezogen wird. Die Standardeinstellung ist mit Zurücklegen. Wird *ohne Zurücklegen* gewählt, so können weitere Einstellungen zum Zurücklegen vorgenommen werden. Die Option *ohne Zurücklegen* steht nur zur Verfügung bei den Bauteilen Box und Stapel.

Aus didaktischer Sicht können mit Hilfe der Funktion „Ziehungen“ folgende stochastische Problemsituationen modelliert werden: einstufige Zufallsexperimente, mehrstufige unabhängige Zufallsexperimente, mehrstufige abhängige Zufallsexperimente. Dabei sind hier Zufallsexperimente mit diskreten, d. h. endlichen oder abzählbar unendlichen Ergebnisräumen gemeint. Als mehrstufiges Zufallsexperiment wird hier zunächst allgemein verstanden, dass mehrere einstufige Zufallsexperimente hintereinander ausgeführt werden. Jedes einzelne einstufige Zufallsexperiment ist dabei eine Stufe des mehrstufigen Zufallsexperiments. Werden nacheinander fünf zusammengehörige Zufallsex-

perimente durchgeführt, so ist dies hier ein fünfstufiges Zufallsexperiment. Dabei ist es zunächst einmal einerlei, ob die einzelnen Zufallsexperimente gleichartig oder verschieden sind und ob die verschiedenen Stufen voneinander abhängen oder nicht. Auf diese Unterscheidung und deren Simulationsumsetzung mit TinkerPlots wird detailliert im vierten Kapitel eingegangen.

Zur Verdeutlichung dient hier der doppelte Würfelwurf als Beispiel für ein zweistufiges, unabhängiges Zufallsexperiment. Theoretisch lässt sich dieser beschreiben durch den Ergebnisraum

$$\Omega = \{(x, y) | x, y \in \mathbb{N}; 1 \leq x \leq 6, 1 \leq y \leq 6\}.$$

Wie bereits oben gezeigt, kann der Würfel mit verschiedenen Bauteilen in TinkerPlots modelliert werden. Ein mögliches mentales Modell zum doppelten Würfelwurf ist das zweifache Ziehen mit Zurücklegen aus einer Urne mit sechs von 1 bis 6 beschrifteten Kugeln. Dies kann in TinkerPlots direkt modelliert werden, indem der Würfel durch die Box modelliert wird und bei Ziehungen 2 eingestellt wird (Abbildung 33).

Abbildung 33: Der doppelte Würfelwurf modelliert durch zwei Ziehungen aus einer Box

Dabei entsprechen mehrere Ziehungen aus dem Bauteil Box (ohne weitere Einstellungen wie z. B. mit Zurücklegen) der Durchführung entsprechend vieler stochastisch unabhängiger Versuche eines Zufallsexperiments.

Bei dem in Abbildung 33 modellierten Zufallsexperiment handelt es sich um ein zweistufiges Zufallsexperiment, dies wird repräsentiert durch die Anzahl der Ziehungen.

Als Unterschied dazu kann auch die Vorstellung zum Tragen kommen, dass zwei Würfel geworfen werden und dementsprechend im mentalen Modell zwei Urnen zwei Würfel repräsentieren. Dann würde mental erst aus der einen Urne und dann aus einer zweiten Urne gezogen werden. Stochastisch ist es das gleiche, ob aus einer Urne zweimal mit Zurücklegen gezogen wird, oder ob aus zwei gleich aufgebauten Urnen je einmal gezogen wird, beiden Experimenten liegt implizit stochastische Unabhängigkeit der

Versuche zugrunde. Für den Nutzer kann es jedoch einen Unterschied in der Vorstellung bedeuten, und genau dies lässt sich mit TinkerPlots abbilden, wie in Abbildung 34.

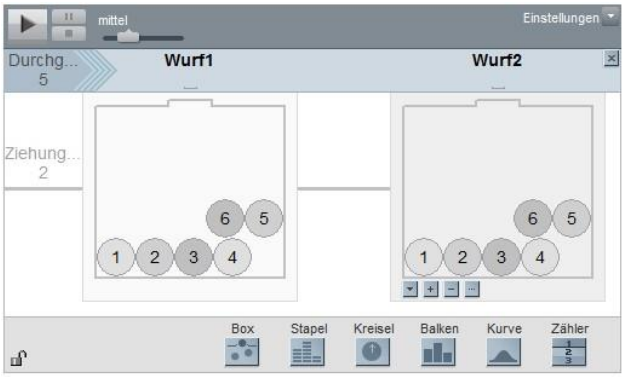


Abbildung 34: Der doppelte Würfelwurf modelliert durch je eine Ziehung aus einer eigenen Box

In Abbildung 34 ist der doppelte Würfelwurf durch zwei Boxen modelliert, die unabhängig voneinander hintereinander angeordnet sind. Die Anzahl der Ziehungen kann in diesem Fall nicht verändert werden, sondern wird durch das Programm automatisch auf zwei festgelegt, da es genau zwei Bauteilspalten, jeweils bestehend aus einem Bauteil, gibt. Natürlich lässt sich dies wie in Tabelle 4 beschrieben auch mit anderen Bauteilen realisieren.

Noch einmal zurück zu dem Urnenbeispiel: Aus einer Urne mit vier roten und fünf blauen Kugeln soll nun zweimal mit Zurücklegen gezogen werden. Die beiden Ziehungen sind durch das Zurücklegen stochastisch unabhängig voneinander, d. h. die Wahrscheinlichkeit für eine bestimmte Kugel beim zweiten Zug ändert sich nicht durch die Ziehung der ersten Kugel. Die Modellierung dieses Zufallsexperiments durch das Bauteil Box ist bereits oben gezeigt (Abbildung 30). Wird das Zufallsexperiment durch das Bauteil Box realisiert, so wird ein großer Kritikpunkt an diesem Bauteil in Verbindung mit dem Interface Zurücklegen deutlich: Es lässt sich nicht direkt in der Zufallsmaschine erkennen, ob beim mehrfachen Ziehen aus der Box mit oder ohne Zurücklegen gezogen wird. Voreingestellt ist das Ziehen mit Zurücklegen, was in diesem Fall zur Beispielsituation passt. Weiß der Nutzer jedoch nicht, wie die entsprechende Einstellung ist, so kann er etwas umständlich im Pfeilmenü unter dem Eintrag „Zurücklegen“ nachsehen (Abbildung 35), ob gerade mit oder ohne Zurücklegen eingestellt ist. Sollte dies nicht passend eingestellt sein, so führt dies recht schnell und unbemerkt zu einer falschen Modellierung einer Situation, sofern es nicht durch den Benutzer aktiv kontrolliert wird.

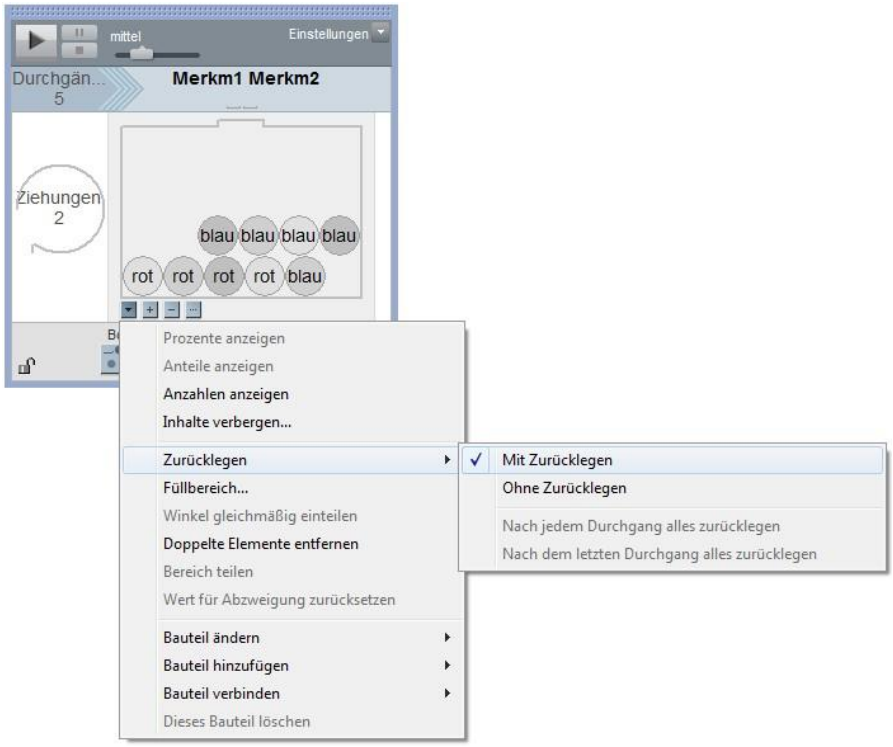


Abbildung 35: Die Einstellung "Zurücklegen" beim Bauteil Box

Liegt die Modellierungsvorstellung zu diesem Beispiel über zwei gleiche Urnen vor (Abbildung 36), so ist es durch das Ziehen mit Zurücklegen ein unabhängiger zweistufiger Zufallsversuch und dieses Problem der Einstellung mit Zurücklegen wird in TinkerPlots umgangen, da aus jeder Urne nur einmal gezogen wird (beim einmaligen Ziehen spielt das Zurücklegen keine Rolle).

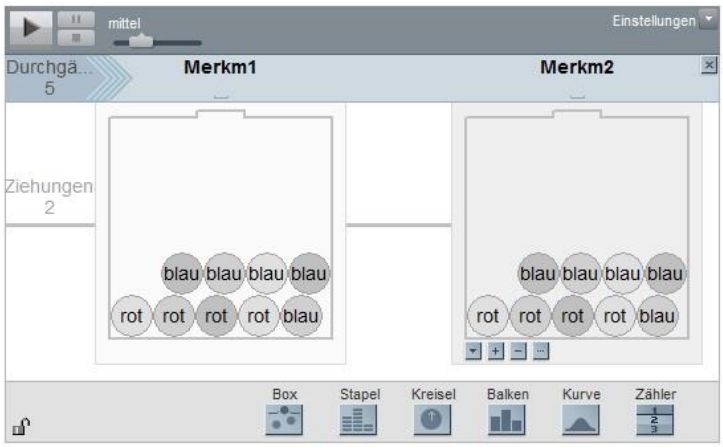


Abbildung 36: Zweimaliges Ziehen aus je einer Urne modelliert durch zwei gleiche Boxen

Es folgt ein Beispiel eines zweistufigen Zufallsversuchs, diesmal als abhängiger Zufallsversuch.

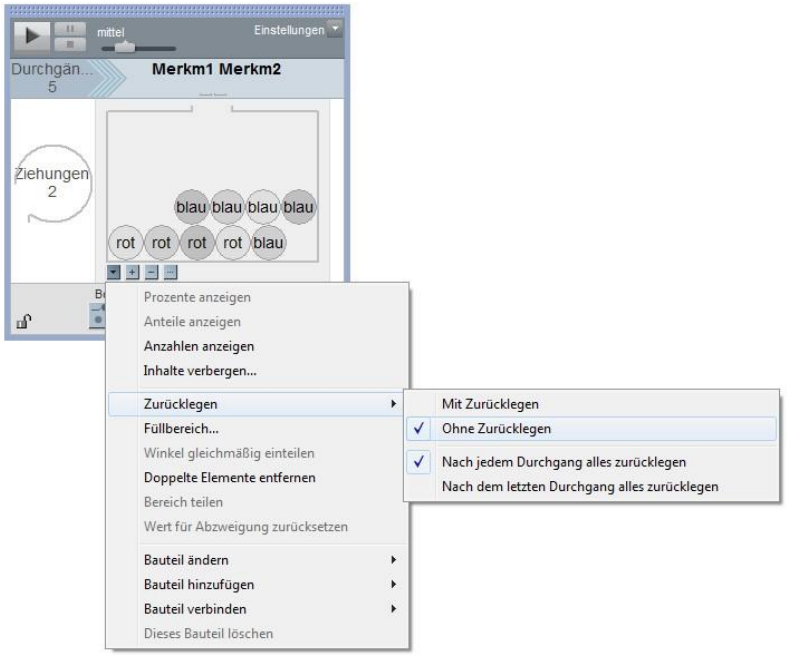


Abbildung 37: Ziehen ohne Zurücklegen

Wird das Beispiel von eben dahingehend geändert, dass aus der Urne mit vier roten und fünf blauen Kugeln zweimal ohne Zurücklegen gezogen wird, so kann die Visualisierung durch zwei Bauteilspalten den Vorteil haben, dass der Nutzer die Situation gut durchdenken muss. Die einfache Modellierung ist diejenige, für die aus einer Box zweimal gezogen wird und ohne Zurücklegen eingestellt wird (Abbildung 37).

Für die Modellierung mit zwei Bauteilspalten, also mit zwei Boxen hintereinander, muss der Nutzer die Situation vollständig durchdenken. Wie sieht die Urne aus, nachdem einmal aus ihr gezogen wurde? Es kann im ersten Zug eine rote oder eine blaue Kugel gezogen werden, je nachdem ist die Urne anschließend unterschiedlich gefüllt. Wurde im ersten Zug eine rote Kugel gezogen, so liegen für den zweiten Zug noch drei rote und fünf blaue Kugel in der Urne. Wurde im ersten Zug eine blaue Kugel gezogen, so liegen für den zweiten Zug noch vier rote und vier blaue Kugeln in der Urne. Für den zweiten Zug werden also quasi zwei Urnenmodelle benötigt, abhängig vom Ausgang der ersten Urnenziehung. Die Zufallsmaschine unterstützt auch diese Vorstellung durch die Möglichkeit, Bauteile abhängig voneinander anzuordnen (Abbildung 38).

Wird in Abbildung 38 für Merkmal1 in der ersten Bauteilspalte eine rote Kugel gezogen, so folgt TinkerPlots automatisch der mit „rot“ beschrifteten Abzweigung in die obere Box der zweiten Merkmalsspalte. Wird hingegen eine blaue Kugel gezogen, so folgt TinkerPlots der mit „blau“ beschrifteten Abzweigung in die untere Box der rechten Merkmalsspalte. Da jede einzelne Box in Abbildung 38 einzeln bestückt werden muss, muss auch für jede einzelne Box durch den Benutzer eine Entscheidung getroffen wer-

den, wie viele Kugeln mit welcher Beschriftung für die jeweilige Box vorhanden sein müssen.

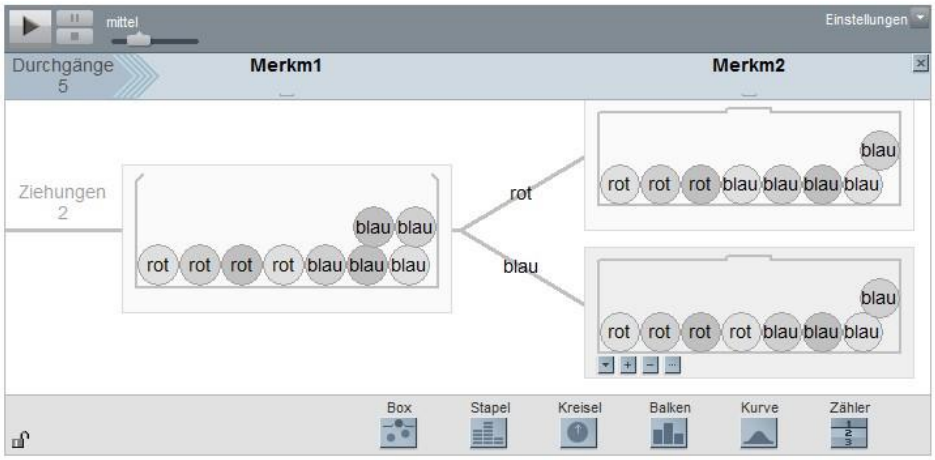


Abbildung 38: Ziehen ohne Zurücklegen mit zwei Merkmalsspalten visualisiert

Wie an diesen Beispielen zu sehen ist, lassen sich mit der Zufallsmaschine mehrstufige Zufallsexperimente gut realisieren. Auf genau diese Art können weitere Bauteilspalten hinzugefügt werden. Die Funktion Ziehungen entspricht dabei den n Stufen des Zufallsexperiments. Werden die Bauteile für mehrstufige Zufallsexperimente hintereinander, unabhängig oder abhängig, kombiniert, so entspricht die Einstellung bei Ziehungen automatisch der Anzahl der Bauteilspalten und somit der Anzahl der Stufen eines Zufallsexperiments. Es können bis zu 100 Bauteilspalten realisiert werden, mit entsprechenden Abzweigungen. Vor allem die Größe des Monitors und dadurch resultierend die Bedienbarkeit der einzelnen Bauteile setzt der Anzahl der Abzweigungen Grenzen. Weiter können ab der dritten Bauteilspalte auch übereinander liegende Urnen miteinander verbunden werden, so wie in Abbildung 39.

Die Abzweigungen von einer zur nächsten Bauteilspalte werden zunächst durch die Software automatisch beschriftet. Entsprechend der Beschriftung auf der Abzweigung wird das für die Simulation zu benutzende Bauteil der nächsten Bauteilspalte ausgewählt. Die automatische Beschriftung erfolgt anhand der Elemente, die das Bauteil beinhaltet, von dem die Abzweigung abgeht. Werden Abzweigungen hinzugefügt, so ändert die Software die Beschriftungen auch automatisch. Natürlich können die Abzweigungen auch beliebig von Hand beschriftet werden, mit einer Ausnahme. Eine sinnvolle Beschriftung entspricht einem Element oder einem Bereich von Elementen des Bauteils, von dem die Abzweigung abgeht. Außerdem sind sinnvolle Beschriftungen der Abzweigungen an einem Bauteil disjunkt. Es ist möglich, eine Abzweigung mit einem Bereich zu beschriften. Liegen beispielsweise in einer Box sechs Kugeln von 1 bis 6 und gehen von dieser Box zwei Abzweigungen ab, so kann eine Abzweigung mit 1-5 und die andere mit 6 beschriftet werden. Dies gilt ebenso für Buchstabenbereiche. Taucht

ein Element als Inhalt der Beschriftung in mehreren Abzweigungen auf, so wählt die Software für dieses Element die oberste Abzweigung. Wird eine Beschriftung gewählt, die in dem vorigen Bauteil nicht vorkommt, so wird die entsprechende Abzweigung nicht verwendet und aus dem entsprechenden Bauteil nicht „gezogen“.

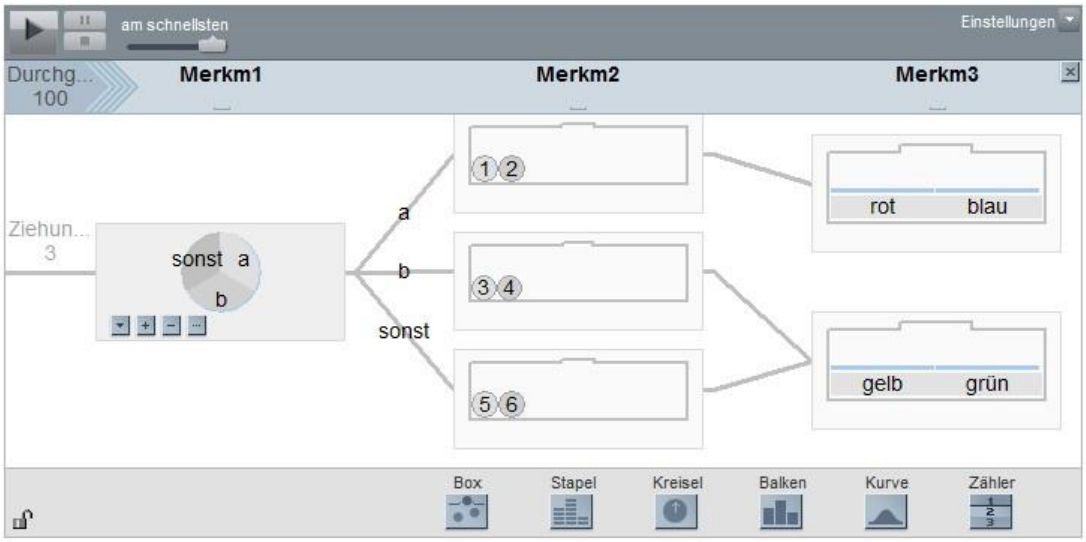


Abbildung 39: Eine Zufallsmaschine mit verzweigten und wieder zusammengeführten Bauteilen

Eine Sonderstellung nimmt die Abzweigung „sonst“ ein. Gibt es im Bauteil, von dem die Abzweigung abgeht, mindestens ein nichtnumerisches Element, so erscheint automatisch beim Hinzufügen einer Abzweigung auf der untersten Abzweigung die Beschriftung „sonst“. Liegen Elemente im Bauteil, die sich in keiner Beschriftung der Abzweigung wiederfinden, und es existiert eine Abzweigung „sonst“, so wird diese für diese Elemente ausgewählt. Trotzdem ist es möglich, dem ersten Bauteil ein Element hinzuzufügen, das „sonst“ heißt, in diesem Falle wird auch die „sonst“-Abzweigung genommen. Allerdings kann keine Abzweigung manuell mit „sonst“ beschriftet werden, denn dann wird automatisch ein Fehler im Programm erzeugt (siehe Abbildung 40).

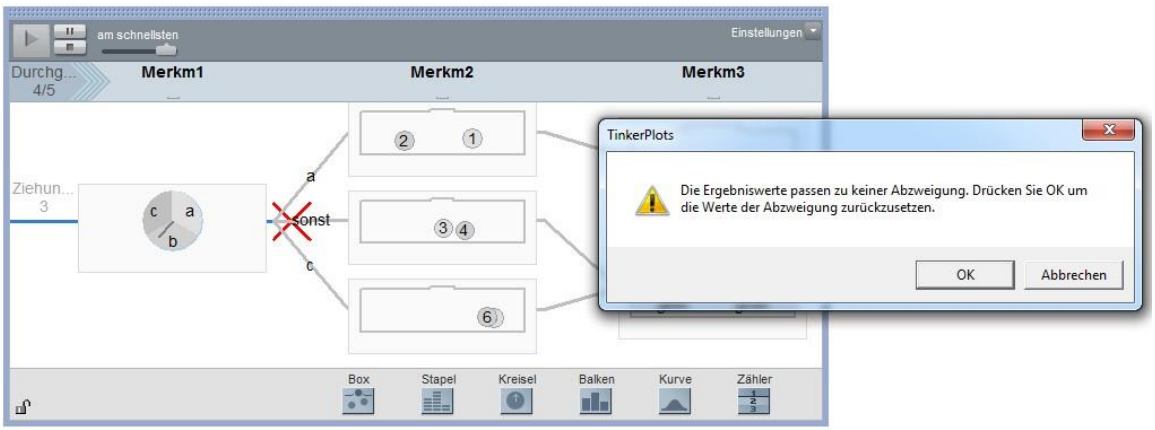


Abbildung 40: Manuelle Eingabe "sonst" an einer Abzweigung

Allerdings wird dieser Fehler durch das Programm nach Betätigen der OK-Taste automatisch behoben und in eine sinnvolle Beschriftung geändert (Abbildung 41).

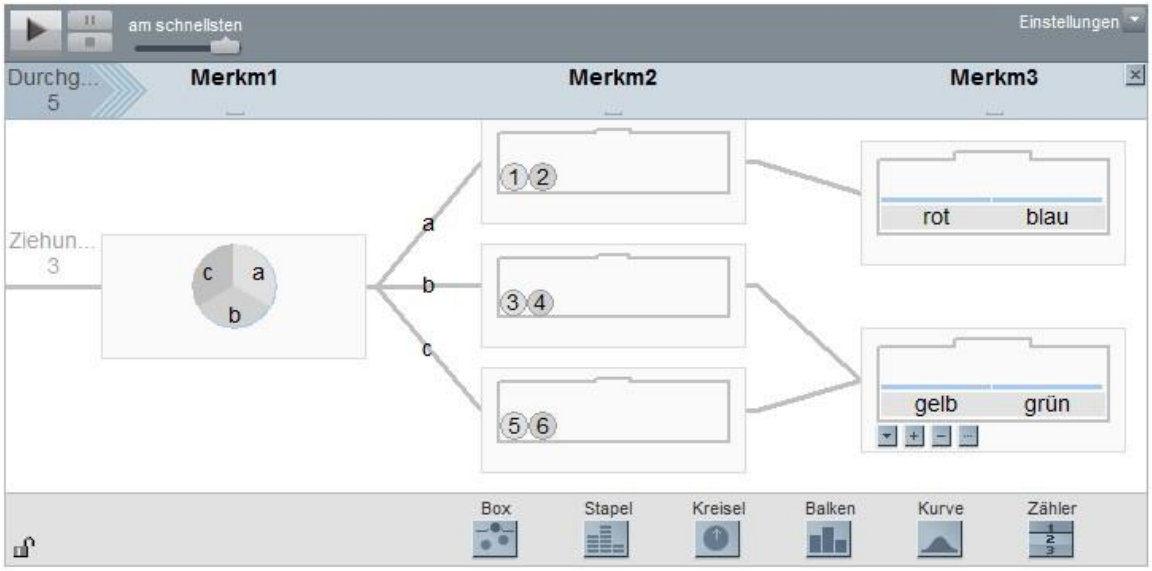


Abbildung 41: Automatische Korrektur der Beschriftung an der mittleren Abzweigung

Die Visualisierung eines mehrstufigen Zufallsversuchs durch die Kombination mehrerer Bauteile hintereinander hat aus didaktischer Sicht den Vorteil, dass der Benutzer den zugrundeliegenden Prozess aktiv darstellen muss. Die Auswirkungen von stochastischer Unabhängigkeit oder Abhängigkeit werden so direkt thematisiert. Bei der Realisierung von mehreren Bauteilspalten können auch unterschiedliche Bauteile zum Einsatz kommen, wodurch die unterschiedliche Repräsentation eines mentalen Models gefördert werden kann.

Die einfache Eingabe der Ziehungsanzahl in Verbindung mit der Einstellung zum Zurücklegen hat den Vorteil der schnellen Umsetzung einer Modellierung. Hierdurch können jedoch wie oben gezeigt Fehlerquellen auftreten, sofern das Thema mit oder ohne Zurücklegen nicht genügend Beachtung beim Nutzer findet. Sollte sogar ein Bauteil gewählt werden, bei dem das Ziehen ohne Zurücklegen nicht einstellbar ist, wie zum Beispiel beim Bauteil Kreisel, so kann dies, durchaus unbemerkt, zu einer falschen Modellierung einer Simulation führen.

Bereich 3: Durchgänge

In Bereich 3 wird die Anzahl der Durchgänge festgelegt (Abbildung 42). Voreingestellt sind fünf Durchgänge. Hier sind Eingaben natürlicher Zahlen bis 100.000 möglich. Jede Eingabe, die davon abweicht, wird durch die Software automatisch auf die Zahl 1 gesetzt. Ein Durchgang beinhaltet so viele Ziehungen, wie in Bereich 2: Ziehungen angegeben sind.

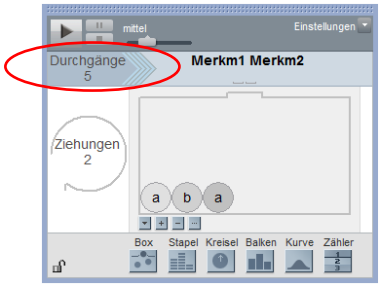


Abbildung 42: Bereich 3 der Zufallsmaschine

Bereich 4: „Steuerungselemente“

Bereich 4 ist der obere dunkelgraue Bereich, in dem die Simulation gestartet, gestoppt und in der Geschwindigkeit variiert werden kann (Abbildung 43). Ebenfalls zu diesem Bereich zählt das Schloss in der linken unteren Ecke, mit dem die Zufallsmaschine mit Passwort gegen Veränderungen geschützt werden kann. Dieser Bereich ist rein technischer Natur und gehört so gesehen zu keinem Interface. Er wird jedoch benötigt, um den Prozess der Zufallserzeugung zu starten, deshalb ist er hier dem Interface Modellierung zugeordnet.

Abbildung 43: Bereich 4 der Zufallsmaschine

Diesem Bereich kommt aus didaktischer Sicht Bedeutung zu, weil durch die Anpassung der Geschwindigkeit der ablaufende Zufallsprozess sichtbar gemacht werden kann. Eine sehr langsame Geschwindigkeit bedeutet, dass zum Beispiel beim Bauteil Box die Kugeln sehr langsam durcheinandergewirbelt werden bis schließlich eine Kugel ausgewählt und deren Wert notiert wird. Somit ist die Generierung der Zufallswerte sehr gut visualisiert und es agiert keine „Blackbox“ im Hintergrund, deren Funktionsweise unklar bleibt.[15] Ist die Visualisierung dieses Prozesses nicht erwünscht oder steht die Geschwindigkeit der Simulation im Vordergrund, so lässt sich auf schnellster Stufe diese Visualisierung abschalten und die Ergebnisse werden sozusagen auf einen Klick erzeugt.

[15] Natürlich agiert in technischer Hinsicht im Hintergrund ein mathematischer Algorithmus, der die Zufallszahlen erzeugt. Nichtsdestotrotz wird durch die Zufallsmaschine dieser Prozess visualisiert und kann dadurch besser nachvollzogen werden.

Exkurs: Zufallsexperimente modellieren in der Tabelle

Es gibt noch eine zweite Möglichkeit der Realisierung von Zufallsexperimenten in TinkerPlots: Die Benutzung einer Tabelle mit entsprechenden Befehlen (Bereich 5). Sollen beispielsweise Zufallszahlen aus speziellen Zufallsverteilungen erzeugt werden, wie zum Beispiel der Binomialverteilung, so kann dies in einer Tabelle mit dem entsprechenden Befehl und dem Hinzufügen von Fällen geschehen. Ebenfalls lassen sich viele Zufallsexperimente über Befehle in einer Tabelle umsetzen, dies geschieht nahezu identisch zu Fathom mit den entsprechenden Möglichkeiten. Diese Art der Zufallserzeugung wurde von Maxara (2009) für Fathom genauestens analysiert und kann auf TinkerPlots übertragen werden, da die entsprechenden Befehle nahezu identisch in TinkerPlots umgesetzt werden können. Maxara nennt die agierenden Befehle zur Zufallserzeugung „Zufallsmaschinen“ (Maxara, 2009, S. 58ff) – hier lässt sich bereits erkennen, dass für einen Großteil dieser Zufallsmaschinen in TinkerPlots ein eigenes Objekt kreiert wurde. Aus diesem Grund wird die Erzeugung von Zufall über Befehle in der Tabelle hier nicht weiter betrachtet. In TinkerPlots wurden zwar die entsprechenden Möglichkeiten von Fathom übernommen, jedoch ist der Standardweg zur Erzeugung von Zufallswerten in TinkerPlots der über das Objekt Zufallsmaschine. Wird der Weg zur Zufallserzeugung über die Tabelle gewählt, dann gehört Bereich 5, die Tabelle, zur Modellierung eines Zufallsexperiments, da in dieser dann entsprechende zufallserzeugende Befehle die Modellierung einer stochastischen Problemsituation übernehmen und das gesamte Objekt der Zufallsmaschine nicht benutzt wird. Gleichzeitig gehört der Bereich zur Umsetzung der Ereignisse und Zufallsgrößen (vgl. Abschnitt 3.6).

Bereich 5a: Ergebnisse der Zufallsmaschine in der Tabelle

Abbildung 44: Bereich 5a: Ergebnisse der Zufallsmaschine

Sobald eine Simulation durch Klicken auf den Playbutton (Bereich 4) gestartet wird, erscheint automatisch eine Tabelle, die die Ergebnisse der Zufallsmaschine protokolliert (Abbildung 44, rechts). In dieser Tabelle werden die durch die Zufallsmaschine erzeugten Werte für jedes Merkmal in einer eigenen Spalte dokumentiert. Jede Zeile der Ta-

belle repräsentiert einen Durchgang der Zufallsmaschine. Dabei hängt die Anzahl der Spalten von der Anzahl der Ziehungen in der Zufallsmaschine ab. Sobald es zwei oder mehr Ziehungen gibt, gibt es zusätzlich zu den einzelnen Merkmalsspalten eine Spalte „Gesamt“, die an vorderster Stelle in der Tabelle steht. Diese Spalte stellt eine Besonderheit dar, da in ihr alle Ergebnisse eines Durchgangs, das heißt alle Ergebnisse der vorhandenen Merkmale, in eine Zelle geschrieben werden. Das Trennungszeichen zwischen den einzelnen Einträgen ist ein Semikolon, dies kann jedoch verändert werden, sofern es gewünscht ist. Eine solche Spalte gibt es weder in Fathom, noch in Tabellenkalkulationsprogrammen, wie beispielsweise Excel. Diese besondere Spalte „Gesamt“ lässt Auswertungsmöglichkeiten zu, die in dieser Form neu sind und einfachere Umsetzungen von Ereignissen und Zufallsgrößen erlauben (vgl. dazu Abschnitt 3.6). Jede Wiederholung der Zufallsmaschine durch Klick auf Play überschreibt die Werte der Tabelle mit neuen Werten. Es kann bei Bedarf eingestellt werden, dass die Werte nicht überschrieben, sondern hinzugefügt werden. Dies kann über *Einstellungen zur Zufallsmaschine → Fälle ersetzen* geändert werden.

Kurz zusammengefasst gilt: Zeilen repräsentieren Durchgänge, Spalten repräsentieren Ziehungen.

Zusammenfassung

Die Zufallsmaschine eignet sich für eine Vielzahl von Modellierungen von Zufallsexperimenten. Wie oben gezeigt, können einstufige, mehrstufige, abhängige und unabhängige Zufallsexperimente mit endlichen oder abzählbar unendlichen[16] Ergebnisräumen in TinkerPlots umgesetzt werden. Dafür stehen die Bauteile Box, Stapel, Kreisel und Balken im Vordergrund. Das Bauteil Kurve verfolgt die Idee der Dichtefunktion, lässt allerdings nur händisch gezeichnete Kurvenverläufe zu und operiert letztlich auf einem diskreten Intervall. Das Bauteil Zähler ist kein Zufallsgenerator. Die vier Bauteile Box, Stapel, Kreisel und Balken ähneln sich in gewissem Sinne. Für viele Zufallsexperimente ergeben sich mehrere Möglichkeiten für die Bauteilwahl. Das hängt im Einzelnen von subjektiven Vorlieben ab oder auch von subjektiven mentalen Modellen. Durch die Bereitstellung der verschiedenen Bauteile ist die Wahl oft stark durch den Nutzer geprägt, allerdings sind trotzdem meistens weitere Realisierungen möglich (auch ein Glücksrad kann z. B. über beschriftete Stapel modelliert werden). Diskrete Verteilungen lassen sich mit den verschiedenen Bauteilen realisieren. Die möglichen Ergebnisse eines Zufallsexperiments werden durch die Beschriftung der Elemente in einer Zufallsmaschine abgebildet. Durch die unterschiedliche Darstellungsart der einzelnen Bauteile lässt sich die zugehörige Wahrscheinlichkeitsverteilung unterschiedlich visualisieren. Durch die große Freiheit bei der Beschriftung der Elemente (bis auf Bauteil Kurve) lassen sich

[16] Siehe dazu Kapitel 3.7.

viele Modellierungen intuitiv durchführen, oft ohne dass weitreichende Abstraktionen notwendig sind.

Zufallsexperimente mit gleichverteilten Ergebnismengen lassen sich mit allen vier genannten Bauteilen gut realisieren, durch die graphische Repräsentation eignen sich hier die Box oder der Kreisel besonders gut. Nicht gleichverteilte Ergebnismengen lassen sich über die Bauteile Kreisel oder Balken gut umsetzen, je nach Problemstellung eignet sich bei einer Angabe in Anteilen das Bauteil Kreisel eher, bei absoluten Anzahlen eher das Bauteil Balken. Wird eine Umsetzung eines Zufallsexperiments ohne Zurücklegen modelliert, so kommen nur die Bauteile Box oder Stapel in Frage, da weder beim Kreisel noch beim Balken das Ziehen ohne Zurücklegen möglich ist. Die Unterschiede zwischen den vier Bauteilen bestehen zum einen in der technischen Bedienbarkeit und leichteren oder schweren Umsetzbarkeit eines Problems in ein bestimmtes Bauteil. Zum anderen unterscheidet sich die optische Repräsentation der einzelnen Bauteile und es werden durch die verschiedenen Angebote unterschiedliche Vorstellungen bedient. Innerhalb einer Zufallsmaschine lassen sich Bauteile beliebig miteinander kombinieren, so dass durch Abzweigungen auch mehrstufige, abhängige oder unabhängige Zufallsexperimente modelliert werden können. Der Würfel oder die Münze als typische Zufallsgeräte besitzen keine visuelle Repräsentation in TinkerPlots.

Der Wechsel zwischen Bauteilen hat aus didaktischer Sicht einen hohen Nutzen, da Zusammenhänge somit sichtbar gemacht werden können, wie zum Beispiel das Verhältnis zwischen roten und blauen Kugeln in einer Urne (vgl. Tabelle 6). Ebenfalls ein hohes Unterstützungspotential bietet die Möglichkeit, mehrstufige abhängige Zufallsexperimente über mehrere Bauteilspalten zu visualisieren, da der Nutzer für diese Realisierung eingehender über die Modellierung entscheiden muss als bei der alternativen Modellierung über ein Bauteil in Verbindung mit der Einstellung ohne Zurücklegen. Generell bietet die Zufallsmaschine mit der Möglichkeit, Bauteile hintereinander anzuordnen, hohes Unterstützungspotential bei der Modellierung stochastischer Zufallsexperimente.

3.5.2 Verschiedene Simulationsarten

Bisher wurde aufgezeigt, wie mit der Zufallsmaschine oder der Tabelle in TinkerPlots Zufallswerte generiert werden können. Zu vielen gegebenen mehrstufigen Zufallsexperimenten hat man nun verschiedene Möglichkeiten der Simulationsmodellierung in TinkerPlots. Die Bauteile der Zufallsmaschine können wie oben gezeigt beliebig hintereinander geschaltet und/oder mit Abzweigungen versehen werden. Dadurch können unabhängige, abhängige, einstufige und mehrstufige Zufallsexperimente realisiert werden und zwar wie oben gezeigt auf mehrere Weisen. Diese Modellierungen wurden systematisch analysiert und hinsichtlich ihres Potentials ausgelotet. Mit unterschiedlichen Modellierungen ist nicht nur gemeint, dass zum Beispiel der Würfelwurf durch unterschiedliche Bauteile in der Zufallsmaschine repräsentiert werden kann (vgl. Tabelle 4),

sondern, dass ein Zufallsexperiment durch unterschiedliche Simulationsarten modelliert werden kann wie beispielsweise das zweimalige unabhängige Ziehen von Kugeln aus einer Urne, repräsentiert durch das zweimalige Ziehen aus der gleichen Box (Abbildung 35) oder das Ziehen nacheinander aus zwei verschiedenen, aber gleich aufgebauten Boxen (Abbildung 36). Die Kategorisierung der verschiedenen Simulationsarten wird nun in diesem Abschnitt präsentiert. Dabei werden die Bezeichnungen simultan und sequenziell verwendet, die bereits Maxara (2009) zur Beschreibung der Simulationsarten mit Fathom verwendet hat, allerdings werden die Begriffe auf den vorliegenden Kontext angepasst und unterliegen dadurch einer Bedeutungsverschiebung. Zu den einzelnen Simulationsarten werden kurze Beispiele gezeigt, die vor allem technische Erläuterungen enthalten. Es werden zunächst die Interfaces Modellierung und Wiederholung bewertet, da diese einen wesentlichen Beitrag zur Unterscheidung der verschiedenen Simulationsarten liefern. Im nachfolgenden Abschnitt zum Interface Ereignisse und Zufallsgrößen wird Bezug genommen auf die verschiedenen Simulationsarten.

Maxara (2009, S. 67) unterscheidet für Fathom die vier Simulationsarten *simultane Simulation, sequentielle Simulation, Simulation durch Stichprobenziehung* und *Simulation durch Randomisierung*. Diese werden im Folgenden aufgegriffen, wobei das Hauptaugenmerk auf der simultanen und der sequentiellen Simulation liegt. Als Hauptunterscheidungsmerkmal definiert Maxara die simultane Simulation als Repräsentation von Teilexperimenten eines Zufallsexperiments nebeneinander in Spalten und die sequentielle Simulation als Repräsentation von Teilexperimenten in Zeilen. Diese Unterscheidung gilt auch für TinkerPlots. Die weiteren unterschiedenen Simulationsarten sind das Simulieren durch Stichprobenziehen, welches in TinkerPlots durch die Machart der Zufallsmaschine im Bauteil Box integriert ist und die Simulation durch Randomisierung, für die das gleiche gilt.

Die Analyse der verschiedenen Simulationsumsetzungen hat folgende Simulationsarten für TinkerPlots erbracht: *simultan-gebündelte Simulation* und *simultan-hintereinandergeschaltete Simulation (*als zwei Spielarten der simultanen Simulationsart), *sequentielle Simulation, Simulation durch Stichprobenziehung* und *Simulation durch Randomisierung*. Dabei baut die Namensgebung auf Maxaras Definitionen der Simulationsarten in Fathom auf und adaptiert sie entsprechend auf TinkerPlots. Wie bei Maxara wird die *Simulation mit Abbruchbedingungen* nicht als eigenständige Simulationsart gewertet, weil sie auf die zuvor definierten Arten angewendet werden kann und diese somit lediglich anreichert. Einige Zufallsexperimente lassen sich durch mehrere Simulationsarten realisieren, für andere (wenige) ist durch die Natur des Zufallsexperiments die Simulationsart vorgegeben.

Für die einzelnen Simulationsarten werden zunächst Zufallsexperimente mit unabhängigen Teilexperimenten und anschließend Zufallsexperimente mit abhängigen Teilexperimenten betrachtet.

3.5.2.1 Simultane Simulationsarten: Simultan-gebündelte und simultan-hintereinandergeschaltete Simulation

Simultane Simulationen unabhängiger Teilexperimente

Als Zufallsexperiment wird nun noch einmal die zweifache Urnenziehung mit Zurücklegen aus einer Urne mit vier roten und fünf blauen Kugeln gewählt. Wie bereits oben gezeigt, kann dies auf die folgenden zwei Arten umgesetzt werden, die beispielhaft für die beiden simultanen Simulationsarten stehen. In der ersten Variante wird die simultan-gebündelte Simulationsart vorgestellt, anschließend die simultan-hintereinandergeschaltete Art.

In Abbildung 45 (links) ist die simultan-gebündelte Variante des Zufallsexperiments zu sehen. Dabei wird aus einer Box mit vier roten und fünf blauen Kugeln zweimal mit Zurücklegen gezogen. Das Zufallsexperiment besteht aus zwei Teilexperimenten, bei denen jeweils aus der Box eine Kugel mit Zurücklegen gezogen wird. Die beiden einzelnen Teilexperimente des jeweiligen Ziehens aus der Box werden hier über die gleiche Box realisiert, also werden die zwei Teilexperimente über die Anzahl der Ziehungen (Bereich 2 der Zufallsmaschine) repräsentiert. In der Ergebnistabelle erscheinen die einzelnen Teilexperimente als Spalten (Abbildung 45, rechts). Zusätzlich gibt es noch die Spalte „Gesamt“, in der die Ergebnisse aller Teilexperimente zusammengefasst in einer Zelle stehen.

Abbildung 45: Zweifache Urnenziehung als simultan-gebündelte Simulation

Zur Abgrenzung wird nun noch die zweite Simulationsart an diesem Beispiel vorgestellt, bevor beide Arten allgemein erklärt werden.

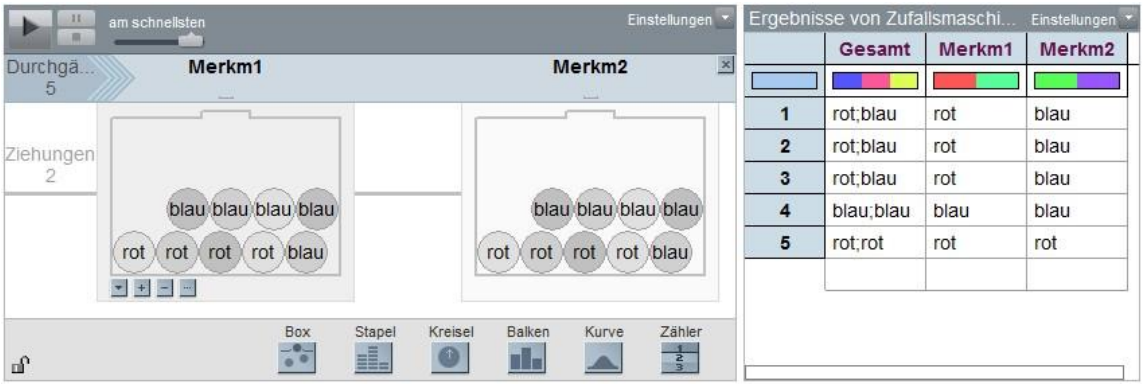


Abbildung 46: Zweifache Urnenziehung als simultan-hintereinandergeschaltete Simulation

Bei der simultan-hintereinandergeschalteten Simulationsumsetzung des zweifachen Ziehens aus der Urne mit vier roten und fünf blauen Kugeln werden die beiden Teilexperimente in jeweils einer eigenen Bauteilspalte realisiert. In Abbildung 46 repräsentiert die erste (linke) Box „Merkm1“ die erste Ziehung und die zweite (rechte) Box „Merkm2“ die zweite Ziehung. Dadurch, dass mit Zurücklegen gezogen werden soll, sind beide Boxen identisch bestückt. Wie bei der simultan-gebündelten Simulation auch entspricht die Anzahl der Ziehungen der Anzahl der Teilexperimente. Die Tabelle (Abbildung 46, rechts) ist gleich aufgebaut wie bei der simultan-gebündelten Simulationsart.

Sowohl die simultan-gebündelte als auch die simultan-hintereinandergeschaltete Simulationsart sind Spielarten des gleichen Simulationstyps, nämlich der simultanen Simulation. Bei diesem Simulationstyp werden die einzelnen Teilexperimente eines Zufallsversuchs in einzelnen Merkmalen repräsentiert, welche in der zugehörigen Ergebnistabelle in einzelnen Spalten abgebildet sind. Die Teilexperimente eines Zufallsexperiments werden bei beiden simultanen Simulationsarten ausschließlich über die Anzahl der Ziehungen realisiert. Dabei bildet die Zufallsmaschine in einem Durchgang alle Teilexperimente ab.

Wird ein Zufallsexperiment mit mehreren Teilexperimenten durch eine mehrfache Ziehung aus dem gleichen Bauteil realisiert, so liegt in Bereich 1b der Zufallsmaschine nur eine Bauteilspalte vor und es handelt sich somit um eine *simultan-gebündelte Simulation* (vgl. Abbildung 45). Werden zur Modellierung in der Zufallsmaschine mehrere Bauteilspalten erzeugt, so nennen wir dies: *simultan-hintereinandergeschaltete Simulation* (vgl. Abbildung 46). In beiden Fällen gibt die Anzahl der Ziehungen die Stufen des Zufallsexperiments an. Bei der simultan-hintereinandergeschalteten Simulation kann die Anzahl der Ziehungen nicht frei gewählt werden, sie entspricht der Anzahl der Bauteilspalten, da aus jeder Spalte genau ein Element ausgewählt wird.

Der Unterschied zwischen den beiden simultanen Simulationsarten besteht für die Realisierung von unabhängigen Teilexperimenten, wie beim obigen Beispiel des Ziehens mit Zurücklegen, in der Repräsentation der Teilexperimente durch die Zufallsmaschine. Dies hat im wesentlichen Einfluss auf die Umsetzung eines mentalen Modells und weniger auf die mathematische Umsetzung. Stochastisch gesehen sind die dahinterstehenden Zufallsexperimente *es wird n mal mit Zurücklegen aus einer Urne gezogen* und *es wird je einmal aus n Urnen mit gleichem Inhalt gezogen* gleich. Allerdings kann für die Modellierung eines Zufallsversuchs, der aus unabhängigen, aber unterschiedlichen Teilexperimenten besteht, nur die simultan-hintereinandergeschaltete Simulationsart verwendet werden.

Simultane Simulationen abhängiger Teilexperimente

Hängen die einzelnen Stufen eines Zufallsexperiments voneinander ab, so ergeben sich zwei Möglichkeiten: Entweder kann man die Situation durch ein Urnenmodell repräsentieren, aus dem mehrfach ohne Zurücklegen gezogen wird, oder es handelt sich um allgemeinere Abhängigkeiten. Liegt der erste Fall vor, so wäre bei der modifizierten Aufgabenstellung von oben (Kapitel 3.5.2) das Ziehen ohne Zurücklegen das neue Beispiel: *In einer Urne liegen vier rote und fünf blaue Kugeln. Aus der Urne werden zwei Kugeln ohne Zurücklegen gezogen.* Auch bei diesem Beispiel kann wie bereits oben gezeigt die Modellierung durch die simultan-gebündelte Simulationsart (Abbildung 47) oder durch die simultan-hintereinandergeschaltete Simulationsart erfolgen (Abbildung 48).

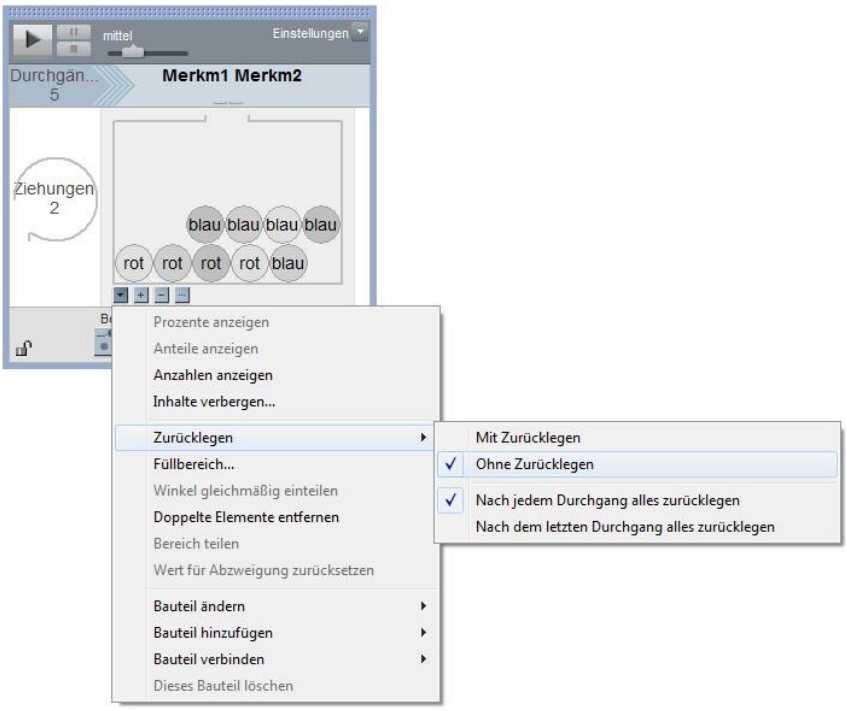


Abbildung 47: Zweifache Urnenziehung ohne Zurücklegen als simultan-gebündelte Simulation

Wie zuvor bei der Umsetzung mehrerer unabhängiger Teilexperimente liegt auch hier der Unterschied zwischen beiden simultanen Simulationsarten nur in der Visualisierung durch die Software. Oben wurde bereits aufgezeigt, welche lerntechnischen Vorteile die simultan-hintereinandergeschaltete Visualisierung hat.

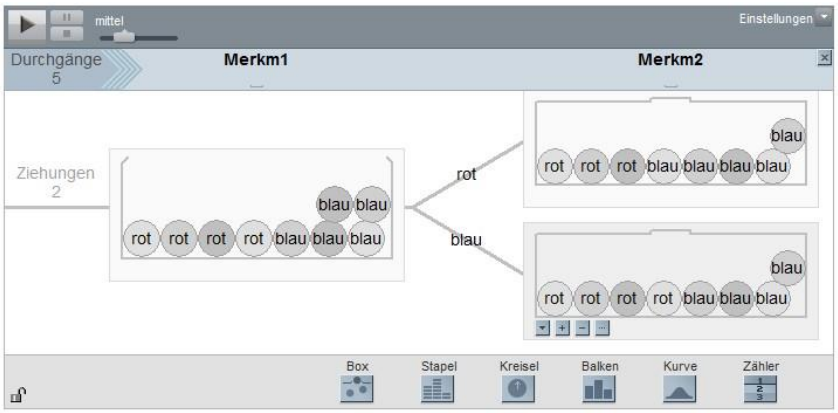


Abbildung 48: Zweifache Urnenziehung ohne Zurücklegen als simultan-hintereinandergeschaltete Simulation

Bis hierher wurden Beispiele verwendet, die sowohl simultan-gebündelt als auch simultan-hintereinandergeschaltet in der Software realisiert werden konnten. Es folgt ein Beispiel, das nur durch die simultan-hintereinandergeschaltete Simulationsart modelliert werden kann.

Kommen wir zu dem zweiten Fall der Modellierung eines Zufallsexperiments mit abhängigen Teilexperimenten. Nun sollen die Teilexperimente allgemeiner voneinander abhängen. Betrachtet wird das folgende prototypische Beispiel (entnommen aus Maxara, 2009, 69): *Es wird zunächst aus einer Urne mit zwei roten, drei blauen und einer gelben Kugel gezogen. Wird eine rote Kugel gezogen, wird beim zweiten Zug aus einer Urne mit zwei mit 1 und 2 beschrifteten Kugeln gezogen, wird im ersten Zug eine blaue Kugel gezogen, so wird anschließend aus einer Urne mit einer lila und einer mit 3 beschrifteten Kugel gezogen, wird im ersten Zug eine gelbe Kugel gezogen, so wird im zweiten Zug aus einer Urne mit zwei mit Smilies beschrifteten Kugeln gezogen.*

Diese Situation lässt sich in TinkerPlots nur durch die simultan-hintereinandergeschaltete Simulationsart modellieren. Dazu werden zwei Bauteilspalten benötigt, die mit den entsprechenden Realisierungen der einzelnen Urnen bestückt werden. Dabei müssen die Abzweigungen nach der ersten Bauteilspalte entsprechend den im Beispiel formulierten Anforderungen genügen.

Bei der Simulationsumsetzung in Abbildung 49 ist wieder das Potential der Zufallsmaschine erkennbar, die es ermöglicht, ein Abbild des mentalen Modells zu erstellen. In der ersten Bauteilspalte in Abbildung 49 ist die erste Urne durch das Bauteil Stapel vi-

sualisiert, von dem drei Pfade abgehen und abhängig von der gezogenen Farbe in Zug1 in entsprechende weitere Bauteile in der zweiten Bauteilspalte führen.

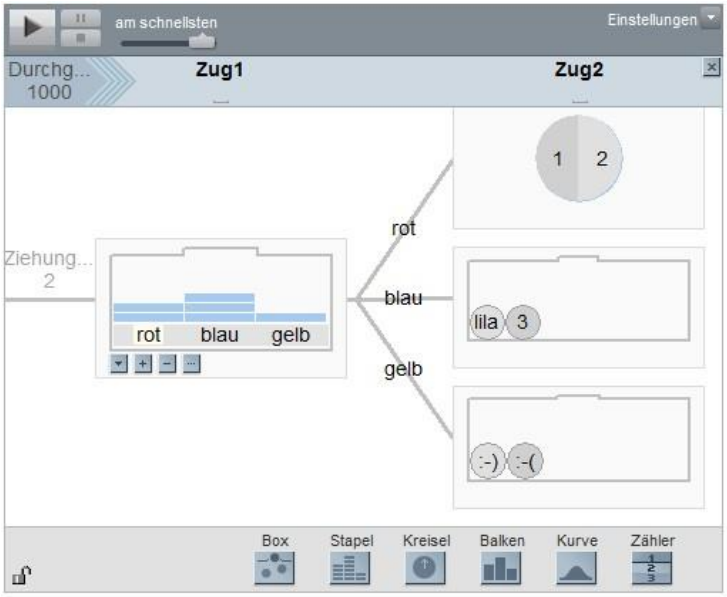


Abbildung 49: Simultan-hintereinandergeschaltete Simulation eines Zufallsexperiments mit abhängigen Teilexperimenten

Auch hier (Abbildung 49) steht die Anzahl der Ziehungen für die Anzahl der Teilexperimente, die zu dem Zufallsexperimente gehören. Diese ist durch die Anzahl der Bauteilspalten festgelegt. Kritisch ist hierbei die Beschriftung der Abzweigungen, denn ein Tippfehler führt hier nicht zu einem Fehler des Programms, sondern dazu, dass durch die fehlende Übereinstimmung zwischen gezogenem Element in Zug1 und nicht passender Abzweigungsbeschriftung das dahinterliegende Bauteil nie ausgewählt würde (vgl. Abbildung 50).

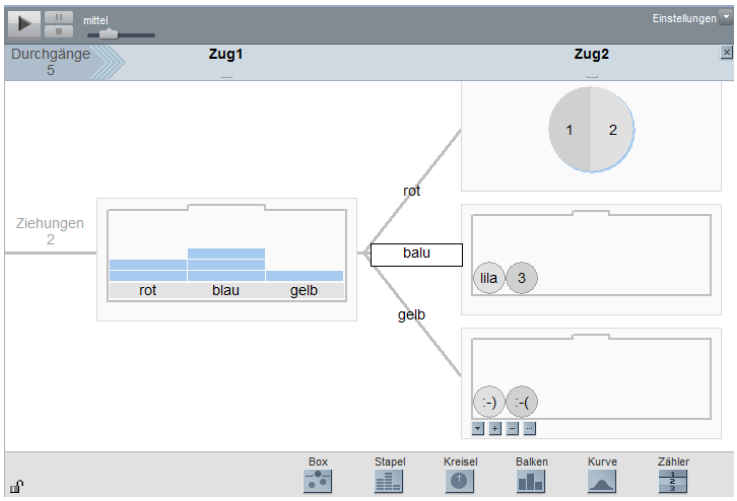


Abbildung 50: Manuell fehlerhaft beschriftete Abzweigung

Durch diese simultan-hintereinandergeschaltete Simulationsart können mit den Bauteilen der Zufallsmaschine abhängige Teilexperiment gut visualisiert werden. Die Abhängigkeiten können durch die Pfadbeschriftung direkt formuliert werden und so sind auch komplexe Zufallsexperimente für den Nutzer ohne komplizierte Funktioneneingaben umsetzbar.

Die Wiederholung des Zufallsexperiments lässt sich über die Anzahl der Durchgänge variieren.

Bewertung

Zur Modellierung mehrstufiger Zufallsexperimente mit einer simultanen Simulationsart werden die Bereiche 1a, 1b und 2 der Zufallsmaschine benötigt. Mit Hilfe der dort vorzunehmenden Einstellungen lassen sich sowohl abhängige als auch unabhängige Zufallsexperimente unkompliziert mit der Zufallsmaschine modellieren. Jedes Teilexperiment wird dazu durch eine eigene Ziehung realisiert und die Ergebnisse in jeweils eigene Merkmalsspalten in der Tabelle geschrieben. Die simultan-hintereinandergeschaltete Simulationsart ist dabei universell, denn durch sie lassen sich sowohl unabhängige als auch abhängige mehrstufige Zufallsexperimente modellieren. Ebenso gilt, dass jede mehrstufige simultan-gebündelte Simulation auch simultan-hintereinandergeschaltet realisiert werden kann. Andersherum gilt dies nicht notwendigerweise. Für unabhängige, gleiche Teilexperimente gilt, dass sie durch beide simultanen Simulationsarten modelliert werden können. Für unabhängige und abhängige verschiedene Teilexperimente muss jedoch die simultan-hintereinandergeschaltete Simulationsart gewählt werden. Dabei gilt, dass die gebündelte Simulationsart vom Aufbau her zeitsparender ist, da viele Teilexperimente einfach durch die Eingabe der Ziehungen und ggf. die Einstellung „ohne Zurücklegen“ realisiert werden können. Dies ist jedoch zugleich ein Schwachpunkt des Systems, da genau diese Einstellung für die Simulation entscheidend ist und auf der Oberfläche der Zufallsmaschine für den Nutzer nicht erkennbar.

Bei beiden simultanen Simulationsarten werden die Teilexperimente des Zufallsexperiments in Spalten der Tabelle realisiert. Die Spalten können dabei sowohl stochastisch unabhängig als auch abhängig voneinander sein, dies wird jedoch in der Zufallsmaschine festgelegt und eine Abhängigkeit ist in den Spalten der Tabelle nicht erkennbar, dort werden in Bereich 5a der Tabelle die Ergebnisse protokolliert.

Der Vorteil der hintereinandergeschalteten Simulationsart gerade bei mehrstufigen abhängigen Teilexperimenten ist der, dass der Nutzer die Abhängigkeiten klar formulieren und die entsprechenden Bauteile für die nachfolgenden Teilexperimente genau bestücken muss. Dies erfordert je nach abzubildendem Zufallsexperiment ein aufwändiges Durchdenken der Möglichkeiten auf den verschiedenen Stufen und kann in der Softwareumsetzung sehr zeitaufwändig sein, wenn es viele Teilexperimente gibt oder wenn

viele Abzweigungen zwischen den einzelnen Stufen benötigt werden. Je nach Größe des Monitors ist auch dies ein limitierender Faktor für die Erstellung einer simultan-hintereinandergeschalteten Simulation, da sich die Darstellungsgröße der Zufallsmaschine nicht anpassen lässt (z. B. sind die Kugeln im Bauteil Box immer gleich groß, unabhängig davon, wie groß das Objekt der Zufallsmaschine ist). Für den Kenner des Systems bietet es sich an, wo es möglich ist, die simultan-gebündelte Simulationsart zu nutzen. Sofern ein Zufallsexperiment mithilfe eines einzigen Bauteils modelliert werden kann, so entspricht die simultan-gebündelte Simulationsumsetzung auch eher dem mentalen Modell eines Nutzers.

Die Wiederholung eines Zufallsexperiments geschieht bei beiden simultanen Simulationsarten einfach über den Playbutton bzw. über die Angabe der Durchgänge, da in jedem Durchgang das gesamte Zufallsexperiment ausgeführt wird. Didaktisch von Vorteil ist die simultan-hintereinandergeschaltete Simulationsart, weil der gesamte Prozess sichtbar ist. Jedes Teilexperiment ist in einer eigenen Spalte visualisiert und Abhängigkeiten können exakt formuliert und visualisiert werden.

3.5.2.2 Sequentielle Simulation

Die sequentielle Simulationsart unterscheidet sich grundlegend von den beiden simultanen Arten. Bei der simultanen Simulation bildet ein Durchgang der Zufallsmaschine das gesamte Zufallsexperiment ab mit der Anzahl der Teilexperimente als Ziehungsanzahl und in der Ergebnistabelle repräsentiert eine Zeile eine Durchführung des Zufallsexperiments. Die sequentielle Simulationsart bildet nun wie bereits von Maxara für Fathom festgestellt (2009, 71) das Zufallsexperiment in den Zeilen der Ergebnistabelle ab. Das bedeutet, die Durchgangszahl repräsentiert die einzelnen Teilexperimente (bei der simultanen Simulationsart war die Durchgangsanzahl bestimmender Faktor für die Anzahl der Wiederholungen des Zufallsexperiments).

Für die sequentielle Simulation wird in TinkerPlots die sogenannte Messgrößenfunktion benötigt. Mit dieser Funktion lassen sich aus einer Graphik heraus von relativen oder absoluten Häufigkeiten und Kennzahlen die Werte mehrerer Simulationsdurchgänge in einer gesonderten Tabelle sammeln und in weiteren Graphen darstellen (vgl. Abschnitt 3.2.3).

Als Einführungsbeispiel, das nur mittels der sequentiellen Simulationsart in TinkerPlots simuliert werden kann, dient das folgende historische Problem von Chevalier de Méré über die Wahrscheinlichkeit von Sechsen beim mehrfachen Würfeln: „*Nach de Mérés Erfahrung konnte man erfolgreich darauf wetten, dass beim mehrmaligen Werfen eines Würfels spätestens bis zum vierten Wurf die Sechs fällt. Nun argumentierte er, dass eine Doppel-Sechs beim Werfen zweier Würfel sechsmal seltener als eine Sechs bei einem Würfel sei und dass folglich beim mehrmaligen Werfen zweier Würfel spätestens bis*

zum vierundzwanzigsten Wurf, denn $6 \cdot 4 = 24$ *, eine Doppel-Sechs fallen müsste"* (entnommen aus Büchter & Henn 2007, 267). Zur Verdeutlichung der sequentiellen Simulationsart bei gleichen unabhängigen Teilexperimenten wird hier das 24malige Werfen zweier Würfel als 24stufiges Zufallsexperiment betrachtet und die Anzahl der Doppel-Sechsen mit Hilfe der Messgrößenfunktion protokolliert.

In Abbildung 51 links oben ist der Wurf zweier Würfel über das Bauteil Box durch die Ziehungsanzahl 2 realisiert. Dieser Doppelwurf soll 24mal stattfinden, was über die Anzahl der Durchgänge umgesetzt ist. Die Modellierung des 24fachen Doppelwurfs ist nur auf diese Weise möglich.[17] In der zugehörigen Ergebnistabelle (Abbildung 51, oben Mitte) stellt eine Zeile einen Doppelwurf als Teilexperiment dar. In der vierten Spalte ist die Zufallsgröße *X: Anzahl der Sechsen* notiert und im Graph (Abbildung 51 rechts oben) dargestellt. Von Interesse ist das Ereignis „Doppel-Sechs", also die Fälle, in denen in beiden Würfen eine Sechs erschienen ist. Das ist bei der abgebildeten Simulation einmal der Fall beim 24fachen Doppelwurf. Diese Anzahl ist nun als Messgröße gesammelt worden (symbolisiert durch das graue Rechteck um den Wert 1 der Ausprägung 2 des Merkmals Anzahl_6) und in der Messgrößentabelle (Abbildung 51 links unten) protokolliert. Die Verteilung der Zufallsgröße, repräsentiert durch das Merkmal Anzahl_Anzahl_6_2 ist in der letzten Graphik (Abbildung 51 rechts unten) zu sehen.

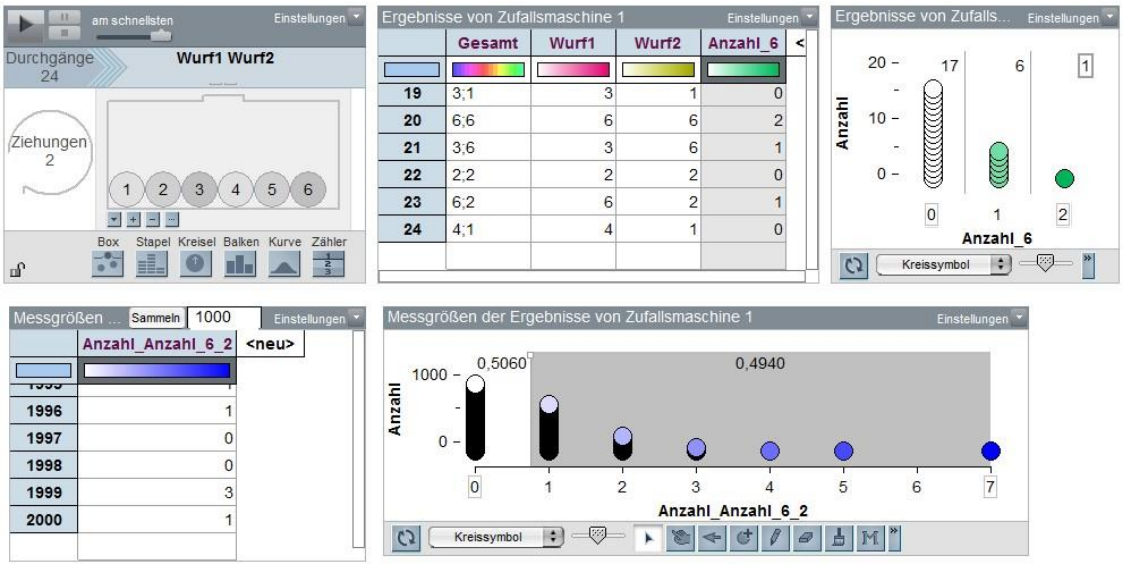


Abbildung 51: Sequentielle Simulation des 24fachen Doppelwurfs

Durch die Beschränkung der Durchführungsanzahl des Experiments auf 24 Durchgänge bei zweimaliger Ziehung, wie durch das Experiment vorgegeben, ist nur eine sequenti-

[17] Natürlich lässt sich der Würfel an sich auch durch ein anderes Bauteil in der Zufallsmaschine repräsentieren (vgl. dazu Tabelle 3), allerdings hat die Würfelrepräsentation keine Auswirkung auf die Einstellungen der Ziehungs- und Durchgangsanzahl, die für diese Simulationsart maßgeblich sind.

elle Simulation möglich. Es ergibt sich bei 2000 Durchführungen des Zufallsexperiments eine relative Häufigkeit von etwas unter 50% für das Auftreten einer Doppel-Sechs.

Bestehen bei einem mehrstufigen Zufallsexperiment die einzelnen Teilexperimente bereits aus Teilexperimenten (wie beim 24fachen Doppelwurf die einzelnen Teilexperimente jeweils aus zwei Würfen, also zwei Teilexperimenten bestehen), so kann dies in TinkerPlots ausschließlich mittels der sequentiellen Simulationsart realisiert werden.

Auf die gleiche Art lassen sich Simulationen umsetzen, die aus abhängigen Teilexperimenten bestehen. Es lässt sich leicht vorstellen, dass in der obigen Situation der doppelte Würfelwurf auch auf hintereinandergeschaltete Weise hätte visualisiert werden können oder bei entsprechend abhängigen Teilexperimenten diese durch die Funktion „ohne Zurücklegen“ oder durch Abzweigungen realisiert worden wären.

Die Zufallsmaschine als Ganzes erzeugt bei dieser Simulationsart in einem Durchlauf Zufallswerte für eine Durchführung des modellierten Zufallsexperiments. Eine Wiederholung des Zufallsexperiments überschreibt in der Standardeinstellung die erhaltenen Werte, d. h. eine Wiederholung findet statt über das Messgrößensammeln der interessierenden Ereignisse und Zufallsgrößen.

Allerdings lassen sich viele Simulationen sowohl simultan als auch sequentiell modellieren. Dazu folgt ein Beispiel, das zunächst mit einer sequentiellen Simulation umgesetzt wird und anschließend erfolgt zum Vergleich die simultane Umsetzung. Als Beispiel dient folgende Situation: „*Test 1 besteht aus 10 Fragen, bei denen der Prüfling entweder ja oder nein ankreuzen kann. Test 2 besteht aus 20 Fragen, bei denen der Prüfling entweder ja oder nein ankreuzen kann. Beide Tests sind bestanden, wenn mindestens 60% der Fragen richtig beantwortet sind. Bei welchem der beiden Tests hat ein Prüfling größere Chancen zu bestehen, wenn er nur rät?“* (Meyfarth 2009, 50).

Exemplarisch wird hier die Situation für den Test mit 10 Fragen vorgestellt, der Test mit 20 Fragen verläuft analog.

Der Test mit 10 richtig oder falsch zu beantwortenden Fragen ist in Abbildung 52 durch das Bauteil Box modelliert. Darin liegen zwei mit „richtig“ und „falsch“ beschriftete Kugeln und in zehn Durchgängen wird je einmal gezogen. Das Ergebnis von zehn Fragen ist in der Ergebnistabelle (Abbildung 52, oben Mitte) zu sehen. Die Tabelle besteht aus einer Spalte mit zehn Zeilen, die für die einzelnen Fragen stehen. Im Hintergrund steht hier eine Bernoulli-Kette der Länge zehn mit Erfolgswahrscheinlichkeit $p = 0{,}5$. Die Zufallsgröße *X: Anzahl der richtigen Antworten* muss für je einen Test aus dem Merkmal1 bestimmt werden. Dies geschieht im Graph oben rechts (Abbildung 52). Im abgebildeten Test wurden vier Fragen richtig beantwortet und die Anzahl der Ausprägung „richtig“ des Merkm1 wurde als Messgröße gesammelt. In der Messgrößentabelle

(Abbildung 52, unten links) ist die Verteilung der Zufallsgröße *X* protokolliert für 1000 Tests bzw. Wiederholungen des Zufallsexperiments und daneben (Abbildung 52, unten rechts) ist die Verteilung graphisch dargestellt und ausgewertet. Es ist zu sehen, dass mindestens sechs Fragen in 35,4% der Fälle richtig beantwortet wurden.

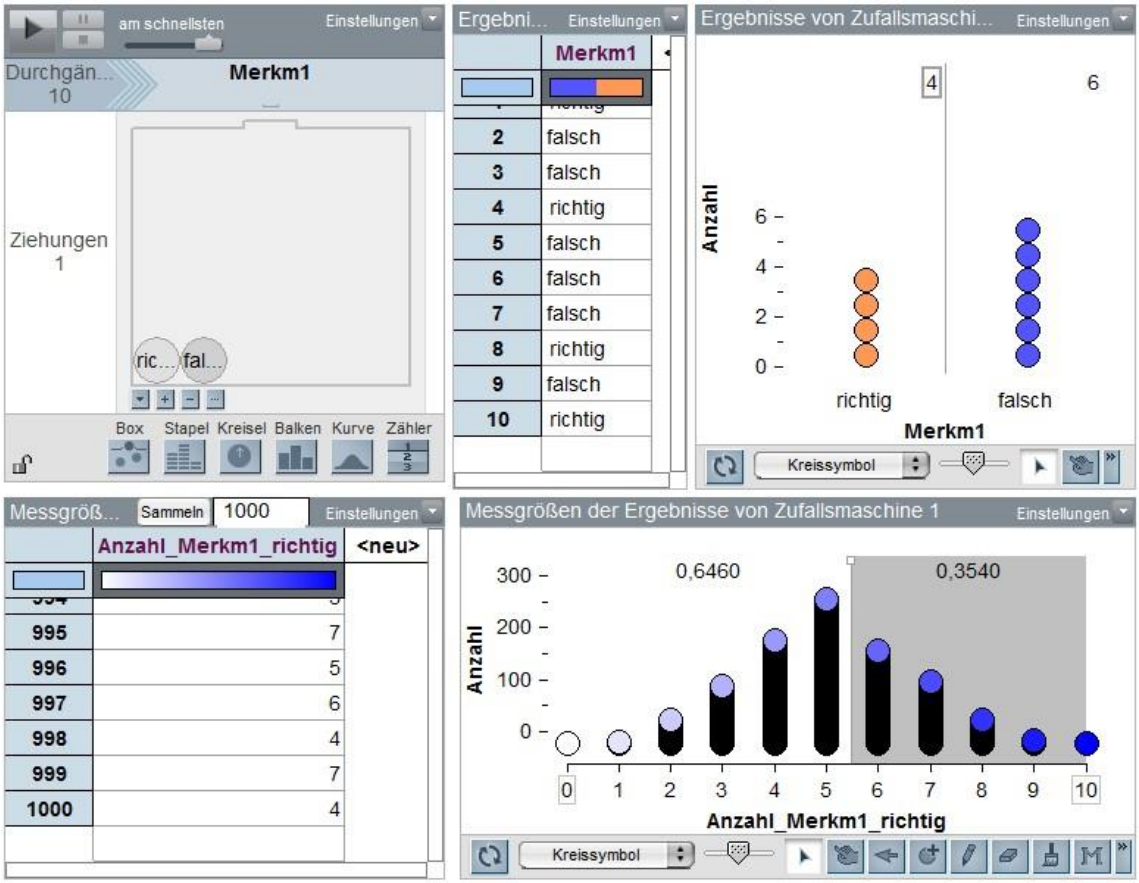


Abbildung 52: Sequentielle Simulation des 10er Tests

Das einfache Auswählen durch Anklicken des Werts, der als Messgröße gesammelt werden soll, ist sehr benutzerfreundlich. Die Messgrößentabelle wird automatisch erzeugt und auch der entsprechende Merkmalsname (im obigen Fall „Anzahl_Merkm1_richtig“) ist für den Nutzer nachvollziehbar aufgebaut durch die Reihenfolge der Worte im Messgrößenmerkmalsnamen *Anzahl_Merkmalsname_Ausprägung*. Die Anordnung der fünf Elemente (Zufallsmaschine, Ergebnistabelle, Graph, Messgrößentabelle, Messgrößengraph) sollte der Nutzer sinnvoll halten, da es bei fünf zwingend erforderlichen Objekten schnell unübersichtlich werden kann.

Es folgt die Umsetzung der gleichen Situation durch eine simultane Simulation.

Auch bei der simultanen Simulationsumsetzung des 10er Tests erfolgt die Modellierung in der Zufallsmaschine durch das Bauteil Box, in dem zwei mit „richtig“ und „falsch“ beschriftete Kugeln für die möglichen Antworten liegen (Abbildung 53, oben links).

Allerdings wird diesmal aus der Box pro Durchgang zehn Mal gezogen und dies 1000 Mal wiederholt. In der zugehörigen Tabelle werden somit elf Spalten automatisch angelegt, für jedes Merkmal eine und zusätzlich die Spalte „Gesamt“, in der alle Merkmalsausprägungen eines Durchgangs in einer Zelle stehen (Abbildung 54). Nun entspricht je eine Zeile einem Test mit zehn Antworten und die Zufallsgröße *X: Anzahl der richtigen Antworten* muss in einer weiteren Spalte der Tabelle definiert werden (Abbildung 53, oben rechts). Hier wird die Anzahl der Ausprägung „richtig“ benötigt, dies kann in TinkerPlots leicht über ein vordefiniertes Merkmal bestimmt werden (Abbildung 53, unten links) (siehe zu vordefinierten Merkmalen Abschnitt 3.6). Auch für dieses Merkmal wird der Merkmalsname automatisch bestimmt und entspricht hier der benötigten Zufallsgröße *X*, die wie oben binomialverteilt ist mit Parametern p = 0,5 und n = 10. Deren Verteilung wird in einem Graphen dargestellt (Abbildung 53, unten rechts) und es ergibt sich hier, dass in 35,8% der Fälle mindestens sechs Fragen richtig beantwortet wurden.

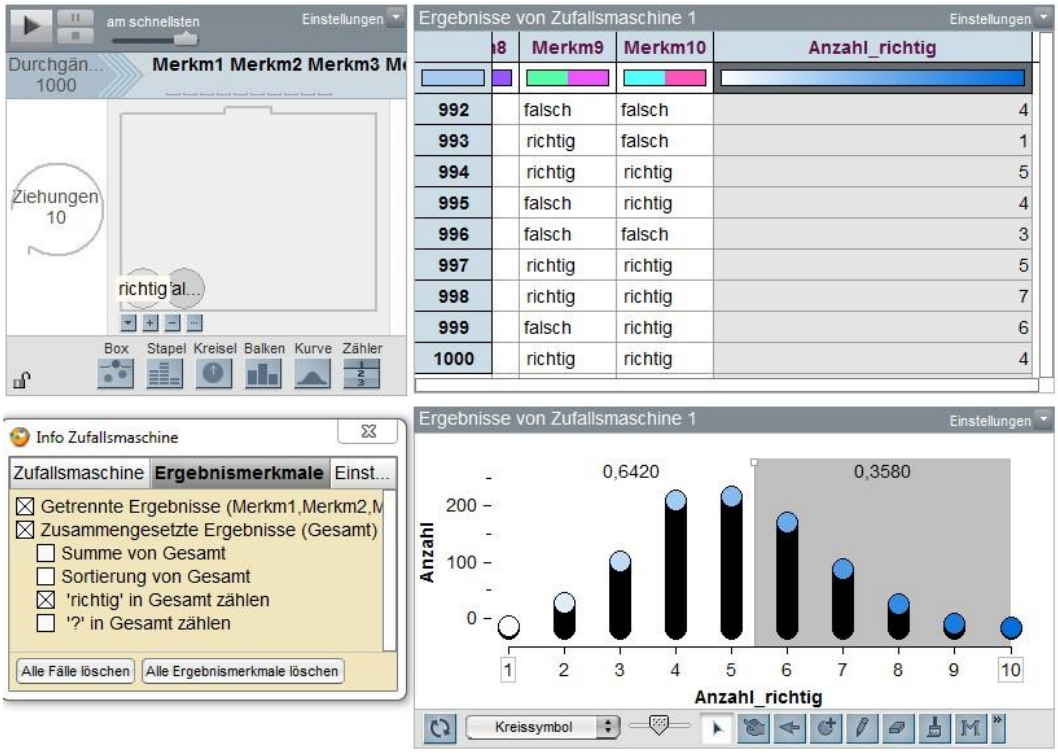


Abbildung 53: Simultane Simulation des 10er Tests

Für diese Simulationsumsetzung werden nur drei Objekte benötigt, dies sind die Zufallsmaschine, die Ergebnistabelle und der Graph. Das zusätzliche Fenster wird nach der Definition des Ergebnismerkmals geschlossen und beansprucht somit keinen Platz auf der Arbeitsfläche.

Wie das Beispiel des 10er Tests können stochastische Situationen, bei denen die Teilexperimente nicht noch einmal aus weiteren Teilexperimenten bestehen, sowohl mit einer

simultanen als auch mit einer sequentiellen Simulation modelliert werden. Der Unterschied liegt in der Betrachtungsweise des Nutzers. Kommt im obigen Beispiel die Vorstellung zum Tragen, dass für zehn Fragen je einmal aus einer Urne mit beiden Antwortmöglichkeiten gezogen wird und zehn dieser einzelnen Ziehungen einen Test ausmachen, so unterstützt die sequentielle Simulation diese Modellierung.

Ergebnisse von Zufallsmaschine 1 Einstellungen

	Gesamt	Merkm1	
995	falsch;richtig;falsch;richtig;falsch;falsch;falsch;richtig;falsch;richtig	falsch	ri
996	falsch;richtig;richtig;richtig;falsch;falsch;falsch;falsch;falsch;falsch	falsch	ri
997	falsch;richtig;falsch;richtig;falsch;richtig;falsch;falsch;richtig;richtig	falsch	ri
998	falsch;richtig;richtig;falsch;richtig;falsch;richtig;richtig;richtig;richtig	falsch	ri
999	falsch;richtig;richtig;falsch;falsch;richtig;richtig;richtig;falsch;richtig	falsch	ri
1000	falsch;falsch;richtig;falsch;falsch;falsch;richtig;falsch;richtig;richtig	falsch	fa

Abbildung 54: Die Merkmalsspalte "Gesamt" der simultanen Simulation des 10er Tests

Kommt hingegen die Vorstellung zum Tragen, dass zehn Ziehungen zusammengehören, wie beispielsweise wenn man aus zehn Urnen gleichzeitig ziehen würde, so unterstützt die simultane Simulation diese Modellierung. Eventuell macht sich ein Nutzer im Vorfeld auch keine Gedanken um die Modellierung der Ziehungen und Durchgänge und wählt ohne besondere Absicht die Ziehungszahl „1“ oder die Ziehungszahl „10“, dann ergibt sich die Simulationsart spontan aus dieser Anzahl der Ziehungen. Ein Nutzer ist an dieser Stelle nicht festgelegt auf eine bestimmte Simulationsumsetzung, was als großer Vorteil der Simulationskapazität von TinkerPlots zu werten ist. An dieser Stelle ist die Bedienung von TinkerPlots intuitiv genug, so dass ein Nutzer fast automatisch die Simulation korrekt weiterführt. Wird bei Ziehung 10 eingegeben, wenn dies intuitiv der Nutzervorstellung entspricht, so geschieht die Wiederholung des Zufallsexperiments über die Anzahl der Durchgänge direkt in der Zufallsmaschine. Wird bei Ziehung 1 eingegeben, wenn das die Nutzervorstellung ist, so ist die Anzahl der Durchgänge festgelegt auf die Anzahl der Fragen (als Stufen des Zufallsexperiments) und die Wiederholung des Zufallsexperiments geschieht über das Sammeln der Messgrößen.

Bewertung

In der Bearbeitung hat die sequentielle Simulationsart den Vorteil, dass eine Durchführung des Zufallsexperiments in einem Graphen dargestellt ist (vgl. Abbildung 52, rechts oben ist ein Test abgebildet). Ein Einzelergebnis ist also für den Nutzer visualisiert und somit ist die Interpretation der einzelnen Elemente des Simulationsprozesses nachvollziehbarer als bei der simultanen Simulation, in der Einzelergebnisse nicht in Graphen visualisiert sind, sondern nur in Tabellen dargestellt sind. Weiterhin ist ein Vorteil der sequentiellen Simulation, dass sie oft ohne Formeln auskommt. An keiner Stelle in der obigen sequentiellen Simulation des 10er Tests musste eine Formel eingegeben werden,

lediglich die als Messgröße zu sammelnde Größe muss durch Anklicken festgelegt werden. Die automatische Benennung der neu generierten Variablen hilft zusätzlich, den Simulationsprozess nicht zu unterbrechen. Gleichzeitig ist an dieser Stelle jedoch der Nutzer gefordert, da die automatische Variablenbeschriftung dazu führen kann, dass der Inhalt des entsprechenden Merkmals nicht durchdacht werden muss, wie es geschehen müsste, wenn ein Merkmalsname händisch sinnvoll vergeben würde. Weiterhin unterscheiden sich die Simulationsarten in der Geschwindigkeit, in der die Zufallswerte generiert werden. Da im sequentiellen Prozess die Ergebnistabelle und die Graphik für jeden Durchlauf neu erzeugt werden muss, ist bei gleicher Wiederholungszahl (Anzahl Durchgänge für die simultane Simulation und Anzahl Messgrößen für die sequentielle Simulation) die simultane Simulationsart schneller.

Im Gegensatz zu Fathom ist in TinkerPlots die Umsetzung einer sequentiellen Simulation technisch nicht anspruchsvoller als eine simultane Simulation. Bei der simultanen Simulation muss (meistens) mit Formeln gearbeitet werden, um Ereignisse und Zufallsgrößen auszudrücken. Dafür werden meistens auch nur drei Objekte, nämlich Zufallsmaschine, Tabelle und Graph benötigt. Bei der sequentiellen Simulation werden (meistens) keine Formeln benötigt, es ist lediglich eine Messgröße (oder mehrere) festzulegen. Die Messgrößentabelle wird automatisch erzeugt, es ist somit eine Auswertungsgraphik für die Messgrößenverteilung zusätzlich nötig, insgesamt hantiert der Nutzer bei der sequentiellen Simulation mit fünf Objekten, nämlich Zufallsmaschine, Tabelle, Graphik, Messgrößentabelle und Messgrößengraphik. Dies erfordert durchaus eine durchdachte Anordnung der Objekte auf dem Bildschirm, damit der Nutzer nicht die Übersicht verliert.

Ein großer Vorteil bei TinkerPlots ist, dass viele Situationen, die sich in Fathom ausschließlich sequentiell realisieren lassen, hier auch durch eine simultane Simulation realisiert werden können. Weitere Unterschiede zwischen den beiden Simulationsarten werden ausführlich im nächsten Abschnitt 3.6 beschrieben, der sich mit dem Interface Ereignisse und Zufallsgrößen beschäftigt. Eine Unterscheidung in unabhängige und abhängige Teilexperimente wird für die sequentielle Simulation nicht gemacht, da sich die Unterschiede wie oben bei der simultanen Simulation ergeben. Die Unterscheidung der simultanen Simulation in eine gebündelte und eine hintereinandergeschaltete Art kann direkt auf die sequentielle Simulation übertragen werden.

3.5.2.3 Simulation als Stichprobenziehung

Maxara (2009, 78ff) hat als weitere Simulationsart die Simulation als Stichprobenziehung hervorgehoben, da dies eine eigene Funktion in Fathom ist. Wie bereits oben erwähnt, kann mit den Mitteln der Zufallsmaschine, insbesondere mit dem Bauteil Box, eine Stichprobenziehung durchgeführt werden. Dazu müssen vorhandene Daten in ein leeres Bauteil hineinkopiert werden und über die Anzahl der Ziehungen kann dann die

Stichprobengröße bestimmt werden. Zur Unterstützung der Modellvorstellung bietet sich wie oben ausgeführt das Bauteil Box an, für das mit oder ohne Zurücklegen eingestellt werden kann. Zu ziehende Merkmale können dabei ein oder mehrere Merkmalsausprägungen haben. Das Stichprobenziehen basiert im Wesentlichen auf der Vorstellung einer Urne und diese weist im Bauteil Box eine nahezu 1:1 Realisierung auf, welche bereits oben ausführlich dargestellt wurde.

Um eine Simulation als Stichprobenziehung auszuführen, ist es hilfreich, wenn die Daten, aus denen die Stichprobe gezogen werden soll, als Liste vorliegen. Diese Liste kann in das leere Bauteil Box der Zufallsmaschine kopiert werden. Sollte die Liste mehrere Merkmale in mehreren Spalten beinhalten, so werden alle Merkmale mitkopiert (maximal 100). Dabei fasst TinkerPlots eine Zeile als einen Fall mit entsprechenden Ausprägungen der Merkmale auf und repräsentiert dies durch eine Kugel in der Box, die mit der Ausprägung des ersten Merkmals beschriftet wird. Oben rechts in der Zufallsmaschine ist dann angegeben, wie viele Fälle/Kugeln mit wie vielen Merkmalen in der Box liegen. Dazu ein Beispiel: In einer Tabelle liegen Daten zu 91 Grundschülern vor, die Angaben zu 29 Merkmalen nach Vorlage eines „Freundschaftsbuchs“ gemacht haben (Abbildung 55).

Freundschaftsbuch_gesamt — Einstellungen

	Fantasiename	Geschlecht	Alter	Körpergröße	Körpergewicht	Monat_Geburtstag	Augenfarbe	Haarfarbe	Schuh
84	Heinrich	männlich	7	133,00	29,00	Juni	mehrere gemischt	braun	
85	Luca1	männlich	7	132,00	25,00	April	blau	blond	
86	Marx	weiblich	8	131,00	26,00	Dezember	mehrere gemischt	blond	
87	Machi	weiblich	8	131,00	23,00	Februar	grün	blond	
88	Hanni	weiblich	7	133,00	28,00	August	blau	blond	
89	Nanni	weiblich	7	130,00	30,50	April	blau	blond	
90	Me	weiblich	8	131,00	24,00	Oktober	grün	blond	
91	Galda	weiblich	8	128,00	23,00	April	braun	braun	

Abbildung 55: Ausschnitt aus einem Datensatz "Freundschaftsbuch"

Die Übertragung in die Zufallsmaschine sieht dann aus wie in Abbildung 56.

Es ist zu sehen, dass sich 91 Fälle, repräsentiert als Kugeln, in der Box befinden (Abbildung 56). Die Beschriftung entspricht der Ausprägung des ersten Merkmals, welches in diesem Beispiel der Fantasiename ist. Die grüne Einfärbung der Kugeln zeigt, dass es sich um „multivariate“ Kugeln handelt, also um Kugeln, die Ausprägungen zu mehr als einem Merkmal haben. Jeder Fall hat 29 Merkmale, was ebenfalls rechts oben in der Zufallsmaschine ablesbar ist. In einem Durchgang wird eine Kugel gezogen. Man kann sich dies so vorstellen, dass eine Kugel zufällig gezogen wird. Auf dieser Kugel stehen die Ausprägungen von allen 29 Merkmalen, die gleichzeitig mitgezogen werden.

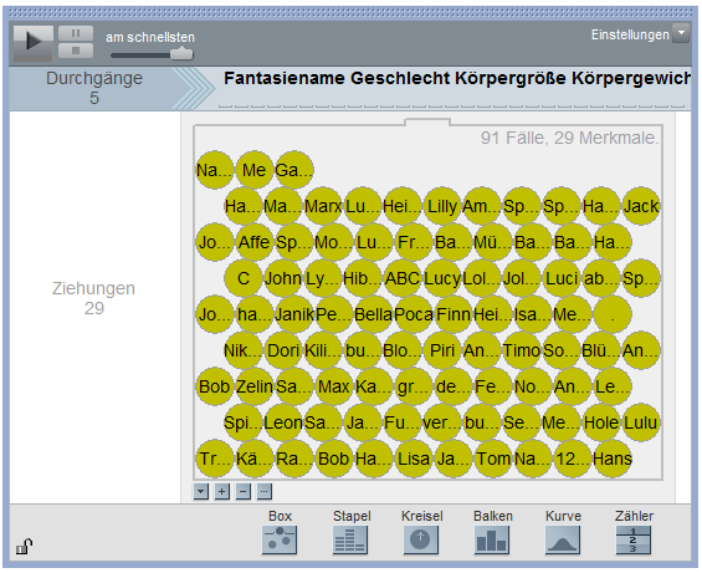


Abbildung 56: Der Datensatz "Freundschaftsbuch" in der Box

Auch wenn bei Ziehungen 29 steht (Abbildung 56), so ist damit keine zufällige Ziehung von 29 Kugeln gemeint, wie es bisher der Fall war, sondern beinhaltet die Ziehung einer Kugel mit allen 29 Merkmalsausprägungen, die in der Ergebnistabelle auch wieder als eigene Merkmale dargestellt werden (Abbildung 57).

Ergebnisse von Zufallsmaschine 1 — Einstellungen

	Fantasiename	Geschlecht	Körpergröße	Körpergewicht	Alter	Monat_Geburtstag	Augenfarbe	Haarfarbe
1	grollerderHeld	männlich	122	31	6	September	braun	braun
2	hapkeinen	männlich	135	31,2	8	Januar	grün	braun
3	Bella	weiblich	141	40	8	März	blau	blond
4	Jonni	männlich	139	20	7	Februar	grün	braun
5	Anna	weiblich	110	20	6	Juni	blau	blond

Abbildung 57: Stichprobenziehung des Freundschaftsbuchdatensatzes aus der Zufallsmaschine in Abb.13

In dieser speziellen Zufallsmaschine (Abbildung 56) bedeutet also ein Durchgang das, was sprachlich mit einer „Ziehung“ ausgedrückt wird. Auf diese Art lässt sich über die Anzahl der Durchgänge eine Zufallsstichprobe des entsprechenden Umfangs aus dem Datensatz ziehen. Eine Randomisierung der einzelnen Merkmalsausprägungen ist mit dieser Methode nicht möglich. Für den Nutzer ist das Stichprobenziehen sehr anschaulich visualisiert durch die Kugeln in der Box. Schwierigkeiten können sich durch die Neuinterpretation der Ziehungen und Durchgänge ergeben.

3.5.2.4 Simulation durch Randomisierung

Weiterhin hat Maxara für Fathom die Simulationsart Simulation durch Randomisierung ausgeführt (Maxara 2009, 81). Auch diese wird hier unterschieden und kann ebenso wie bereits die Stichprobenziehung in TinkerPlots über die Zufallsmaschine realisiert werden. Die Simulation durch Randomisierung ist nur für spezielle Zufallsexperimente geeignet, hierfür bietet TinkerPlots jedoch eine sehr gute Visualisierungsmöglichkeit des zugrunde liegenden Prozesses. Als Verdeutlichung dient das Beispiel aus Maxara (2009, 81) zum Pflanzenwachstum: „*Es liegen die Messdaten von zwei Pflanzenreihen vor. Dazu wurden die Keimlinge von zwei verschiedenen Saaten (a und b) einer Pflanzensorte nach neunzehn Tagen gemessen. Es scheint, dass die Saatsorte b größere Pflanzen hervorbringt als Saatsorte a. Ist es wahrscheinlich, dass die Abweichungen zufällig zustande gekommen sind oder bringt die Saatsorte b wirklich größere Keimlinge hervor?*“ Die Messdaten der Pflanzenreihen liegen in einer Tabelle (Abbildung 58) vor.

Pflanzen Einstellungen

	Samen	Höhe
1	a	13,4
2	a	14,2
3	a	16,0
4	a	15,4
5	a	14,6
6	a	15,7
7	a	16,5
8	b	17,6
9	b	16,0
10	b	16,8
11	b	17,9
12	b	18,1
13	b	15,3
14	b	17,1

Abbildung 58: Ursprungsdaten der Pflanzenreihen

Es sollen nun die Merkmalswerte randomisiert werden, d. h. der jeweiligen Samensorte wird zufällig ein Höhenwachstum zugeordnet. Um zu beurteilen, ob tatsächlich eine Sorte höher wächst als die andere, kann beispielsweise später von beiden Sorten das arithmetische Mittel gebildet und verglichen werden. Die Randomisierung der Merkmalswerte selbst geschieht nun mit der Zufallsmaschine wie folgt. Stellt man sich die Situation so vor, dass in einer Urne 14 Kugeln beschriftet mit „a“ oder „b“ für die Saatsorte liegen und in einer zweiten Urne alle Höhen auf Kugeln stehen, dann kann man zu je einer Kugel aus der „Samenurne“ zufällig eine Kugel mit einer Höhe aus der anderen Urne ziehen. So lässt sich quasi jede Samenkugel mit einer Höhe etikettieren.

Dieser Prozess spiegelt die Situation wieder, in der das Höhenwachstum unabhängig von der Pflanzensorte wäre. In TinkerPlots lässt sich dies gut visualisieren und durch zwei Boxen wird der zufällige Zuordnungsprozess einer Pflanzensorte zu einer Höhe sichtbar.

Die zwei Urnen, in denen die Samensorte bzw. die Höhe notiert ist, sind in Abbildung 59 in der Zufallsmaschine durch zwei Boxen mit entsprechend beschrifteten Kugeln repräsentiert. Die Daten lassen sich spaltenweise in leere Boxen kopieren. In den Einstellungen zu beiden Boxen ist angegeben, dass ohne Zurücklegen gezogen werden soll, so dass bei 14 Durchgängen auch tatsächlich jeder Same, symbolisiert als Kugel, eine Höhe zufällig zugeordnet bekommt. In der Tabelle (Abbildung 59, rechts) ist nach dem Durchlaufen der Zufallsmaschine eine Randomisierung der Ursprungsdaten (Abbildung 59) zu sehen, die nun entsprechend der Fragestellung ausgewertet werden kann.

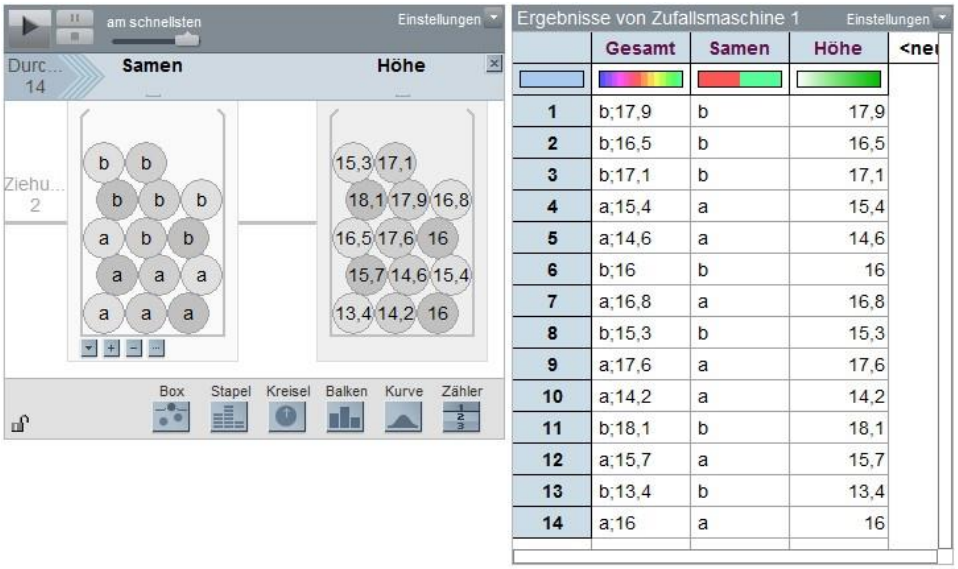


Abbildung 59: Randomisierte Pflanzendaten

Die Auswertung der randomisierten Daten ist nun in TinkerPlots mithilfe des Graphenobjekts leicht möglich. Als Messgröße kann der mit einem Lineal gemessene Abstand zwischen den Mittelwerten beider Samensorten gesammelt werden. Dies wäre in Abbildung 60 das graue Kästchen um 0,69 in der rechten Verteilung.

Die zufällige Zuordnung der Ausprägungen zweier Merkmale, die beim Randomisieren eine wichtige Rolle spielt, kann mit TinkerPlots leicht umgesetzt werden und ist für den Nutzer durch die Visualisierung gut nachvollziehbar. Um den Prozess sichtbar zu machen, eignet sich die Wahl einer langsamen Geschwindigkeit der Zufallsmaschine.

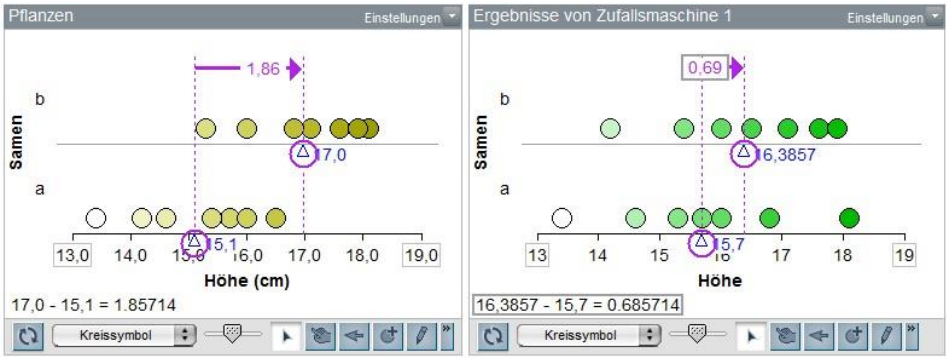


Abbildung 60: Vergleich ursprüngliche Pflanzendaten (links) und randomisierte Daten (rechts)

Hat ein Nutzer ein stärker strukturiertes Modell im Kopf, in dem Sinne, dass zunächst sieben „a“-Samen und anschließend sieben „b“-Samen zufällig mit einer Höhe „etikettiert“ werden sollen, so bietet das Bauteil Zähler weitere Unterstützung an.

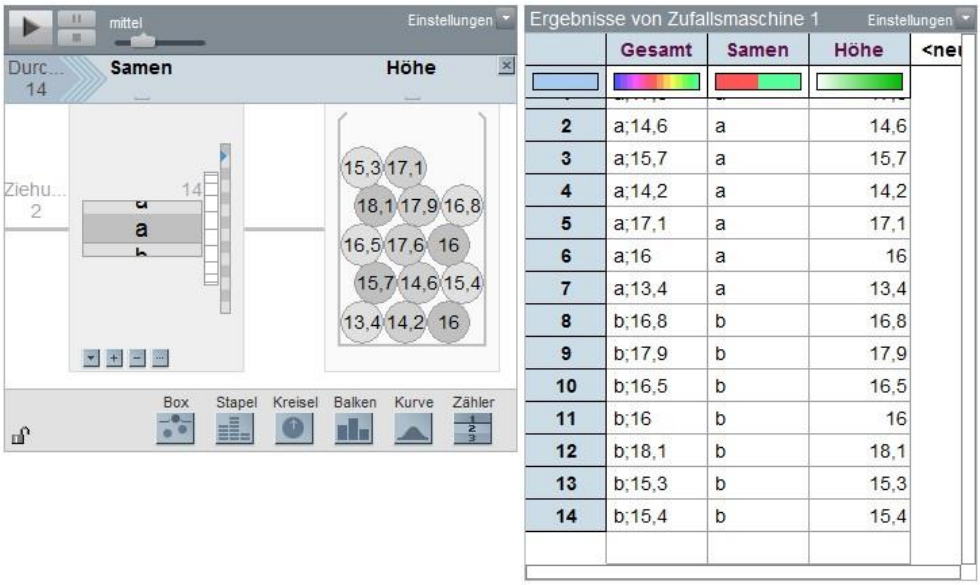


Abbildung 61: Randomisierte Pflanzendaten mit dem Bauteil Zähler

In der Modellierung in Abbildung 61 steht nun zu Beginn das Bauteil Zähler mit 14 Feldern, auf dem die ersten sieben Einträge „a“ lauten und die nächsten sieben Einträge „b“. Pro Ziehung wird der Zähler um ein Feld weitergedreht, so dass im ersten Merkmal nicht zufällig die Samensorte ausgewählt wird und dann zufällig eine Höhe zugelost wird (wieder mit der Einstellung ohne Zurücklegen). In der Tabelle in Abbildung 61 rechts ist dies zu sehen, erst wurden den „a“-Samen Höhen zugelost, dann den „b“-Samen. So lässt sich bezüglich des ersten Merkmals mit diesem Bauteil die Ursprungsordnung der Daten aufrecht erhalten, was bei der Bearbeitung didaktische Vorteile haben kann.

Auch wenn die Simulation durch Randomisierung nur für eine kleine Menge an Zufallsexperimenten eine Rolle spielt, so ist ihr Nutzen doch hoch und TinkerPlots unterstützt durch seine Visualisierung des Prozesses das Verstehen in hohem Maße.

3.6 Interface Ereignisse und Zufallsgrößen

Die Umsetzung von interessierenden Ereignissen und Zufallsgrößen ist neben der Modellierung von Zufallsexperimenten bei Simulationen ein Hauptpunkt. Natürlich können Simulationen auch zur Visualisierung bestimmter Prozesse oder als didaktische Anschauung genutzt werden, ohne dass Ereignisse oder Zufallsgrößen definiert werden, aber eine Hauptanwendung bezieht sich gerade auf die Auswertung von simulierten Daten über Ereignisse und Zufallsgrößen. Deshalb ist es auch bei diesem Punkt wichtig, das Unterstützungspotential der Software detailliert zu untersuchen.

Abhängig von der Simulationsart gibt es in TinkerPlots zwei Stellen, an denen Ereignisse und Zufallsgrößen umgesetzt werden. Zum einen können sie in der Tabelle in Merkmalen definiert werden und zum anderen über die Auswahl bestimmter Werte in Graphen als Messgrößen. Eine Schwierigkeit beim Simulationsprogramm Fathom ist es, dass bereits vor der Modellierung entschieden werden muss, ob Ereignisse und Zufallsgrößen in Merkmalen oder in Messgrößen umgesetzt werden können, da dies die Simulationsart bedingt und häufig nur eine Umsetzung möglich ist (vgl. Maxara 2009, 83). Wie bereits oben gezeigt wurde, ist die Wahl der Modellierung der Zufallsmaschine in TinkerPlots zwar ausschlaggebend für die Simulationsart, allerdings lassen sich viele Ereignisse und Zufallsgrößen sowohl über Merkmale als auch über Messgrößen realisieren, so dass diese Einschränkung für TinkerPlots nicht in gleicher Weise gilt. Ein großes Zutun daran hat die besondere Spalte „Gesamt“ der Ergebnistabelle, da auf ihr einige neue Funktionen operieren, mit denen weiterführende Auswertungen möglich sind.

In TinkerPlots kann nicht auf bestimmte Zellen einer Tabelle explizit zugegriffen werden, eine Angabe wie sie aus Tabellenkalkulationsprogrammen bekannt ist wie z. B. A5 – A10 ist in TinkerPlots nicht möglich. Es kann über den Merkmalsnamen auf bestimmte Spalten zugegriffen werden, der Zugriff auf einzelne Zeilen gestaltet sich als schwierig und ist nur mühsam über verschachtelte Formeln mit dem Befehl *Index* möglich.

Bereich 5b: Merkmale definieren in der Tabelle

Bei simultanen Simulationen werden Ereignisse und Zufallsgrößen als Merkmale in der Ergebnistabelle definiert. Dabei ist ein Ereignis eine Teilmenge der Ergebnismenge Ω, zu denen auch das sichere Ereignis $A = \Omega$ und das unmögliche Ereignis $\varnothing = \{\}$ gezählt werden, welche jedoch für die Simulation in TinkerPlots eher nebensächlich sind.

Auch die Zufallsgröße wird gemäß der gängigen Definition benutzt als Abbildung von einem Ergebnisraum Ω in die Menge der reellen Zahlen

$$Zufallsgröße\ X: \Omega \rightarrow \mathbb{R}.$$

Abbildung 62: Bereich 5b: Definierte Merkmale (graue Spalte)

In der Tabelle können neue Merkmale definiert werden, die auf bereits vorhandene Merkmale zugreifen. Weiß hinterlegte Zellen beinhalten durch Simulation in der Zufallsmaschine generierte Zeichen (oder per Hand eingegebene Zeichen). Grau hinterlegte Zellen in der Tabelle symbolisieren, dass die Einträge zu diesem Merkmal über eine Formel erzeugt wurden. Erkennbar ist dies in Abbildung 62 in der grau hinterlegten Spalte „Anzahl_a“, in der die Anzahl des Zeichens „a“ in der Spalte „Gesamt“ angegeben ist. Auch hier wird im Grunde eine Funktion definiert, die aufgrund einer Formel den Ausprägungen eines Merkmals (oder auch mehrerer Merkmale) eine neue Zeichenkette zuordnet. Es handelt sich dabei um eine Funktion, die zeilenweise operiert unter Bezugnahme auf die Spalte, die durch einen bestimmten Merkmalsnamen spezifiziert wird.

$$f: Merkmalsname \rightarrow \text{Zeichen, mit } f\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \mapsto \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Wobei $z_1, \ldots, z_n$ die Ausprägungen eines Merkmals mit dem Namen *Merkmalsname* entlang der Zeilennummerierung sind und $y_1, \ldots, y_n$ eine darauf operierende beliebige Zeichenzuordnung in einem neu definierten Merkmal darstellt.

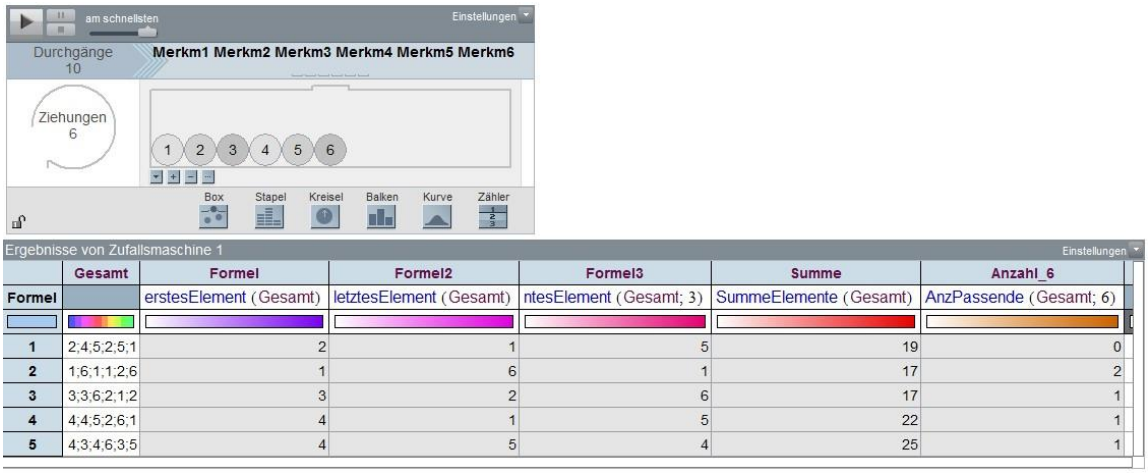
In Fathom werden neue Merkmale üblicherweise über Funktionen definiert, die über einen Formeleditor eingegeben werden müssen. Diese Definitionen sind unabhängig davon, ob mit diesen Formeln Zufallswerte erzeugt werden, wie dies in TinkerPlots durch die Zufallsmaschine geschieht, oder ob durch diese Formeln neue Merkmale definiert werden, die auf bereits vorhandenen Merkmalen operieren. Diese Funktionen existieren in gleicher Form für TinkerPlots. Maxara (2009, S. 84ff) unterscheidet drei Typen von Funktionen, mit denen aus vorhandenen Merkmalen neue Merkmale definiert werden können. Dies sind zum ersten Funktionen, mit denen einzelne Merkmalswerte in

andere Merkmalswerte transformiert werden können, diese Funktionen stammen aus den Bereichen Arithmetik, Bedingungen, Logik, Statistik/Transformation, Textauswertung und Trigonometrie. Die zweite Art von Funktionen verknüpfen zwei oder mehr Merkmalswerte zu einem neuen Merkmalswert. Und die dritte Art von Funktionen transformiert zwei oder mehr Merkmalswerte zu einem neuen Merkmalswert und stammen aus den Bereichen Arithmetik und Logik und die Funktion *verkette*. Dies ist analog in TinkerPlots vorhanden und wird deshalb hier nicht erneut ausgeführt. Für TinkerPlots wurde jedoch ein zusätzliches Merkmal generiert, durch das gegenüber Fathom neue Möglichkeiten eröffnet werden. Dieses Merkmal ist das Merkmal „Gesamt“, das zeilenweise alle Einzelergebnisse der Merkmale als Realisierung der Ziehungen der Zufallsmaschine in einer Zelle auflistet. Zusätzlich wurden in TinkerPlots weitere Funktionen implementiert, die speziell auf Listen operieren. Das Merkmal „Gesamt“ stellt durch seine Darstellungsweise eine Liste dar. Durch diese neuen Funktionen sind die Auswertungsmöglichkeiten im Vergleich zu Fathom erheblich erweitert worden und sind die Möglichkeiten, die die simultane Simulation in TinkerPlots bietet, wesentlich vergrößert bzw. benutzerfreundlicher gestaltet worden.

Zu den vorhandenen drei Typen von Funktionen kommt nun noch ein vierter Typ hinzu, als Funktionen, die auf Zeichenketten operieren, die eine Liste darstellen. Hierdurch werden mehrere Merkmalswerte in einen neuen Merkmalswert transformiert. Aus einer Zeichenkette wird mit Hilfe dieser Funktion ein einzelnes Zeichen. Die neu hinzugekommenen Funktionen sind *erstesElement, AnzElemente, SummeElemente, letztesElement, AnzPassende, MusterVorhanden, ntesElement,* und *ElementeSortieren*, die einige Modellierungen von Ereignissen und Zufallsgrößen vereinfachen.

Mit den Funktionen *erstesElement, letztesElement* und *ntesElement* lassen sich in einer gegebenen Liste Zeichen an bestimmten Stellen der Liste heraussuchen. Durch die Funktion *AnzElemente* lässt sich die Anzahl der Zeichen in einer Liste ausgeben, Merkmalswerte sind hier also natürliche Zahlen und Null. Mit der Funktion *SummeElemente* werden die einzelnen Zeichen einer Liste addiert, sofern es sich um numerische Zeichen handelt. Andernfalls gibt TinkerPlots dem Nutzer einen Hinweis, dass die Zeichen inkompatibel sind. Die beiden Funktionen *AnzPassende* und *MusterVorhanden* überprüfen eine Zeichenkette auf bestimmte Muster. *AnzPassende* überprüft, in welcher Häufigkeit ein bestimmtes Zeichen in einer Zeichenkette vorkommt und gibt diese Häufigkeit als natürliche Zahl oder Null aus. *MusterVorhanden* dagegen überprüft, ob mehrere Zeichen in einer bestimmten Ordnung oder ungeordnet, je nach Angabe, in einer Zeichenkette vorkommen und gibt „wahr“ oder „falsch“ aus.

Dazu dient als Verdeutlichung der sechsfache Würfelwurf, für den verschiedene Zufallsgrößen (Abbildung 63) und verschiedene Ereignisse (Abbildung 64) gezeigt werden.



	Gesamt	Formel	Formel2	Formel3	Summe	Anzahl_6
Formel		erstesElement (Gesamt)	letztesElement (Gesamt)	ntesElement (Gesamt; 3)	SummeElemente (Gesamt)	AnzPassende (Gesamt; 6)
1	2;4;5;2;5;1	2	1	5	19	0
2	1;6;1;1;2;6	1	6	1	17	2
3	3;3;6;2;1;2	3	2	6	17	1
4	4;4;5;2;6;1	4	1	5	22	1
5	4;3;4;6;3;5	4	5	4	25	1

Abbildung 63: Der sechsfache Würfelwurf mit verschiedenen Zufallsgrößen

In Abbildung 63 ist der sechsfache Würfelwurf simultan durch die sechsfache Ziehung aus einer Box modelliert. In der Tabelle (Abbildung 63, unten) ist die erste Spalte „Gesamt“ automatisch erzeugt. Die Merkmalsspalten, die zu den einzelnen Ziehungen erzeugt werden, sind in der Abbildung nicht zu sehen, diese befinden sich weiter rechts in der Tabelle. Die grau hinterlegten Merkmalsspalten sind per Formel definiert, welche sich in der Zeile direkt unter dem Merkmalsnamen findet. Durch diese Formeln sind nacheinander die Zufallsgrößen aus Tabelle 7 realisiert.

Tabelle 7: Realisierungen verschiedener Zufallsgrößen beim sechsfachen Würfelwurf

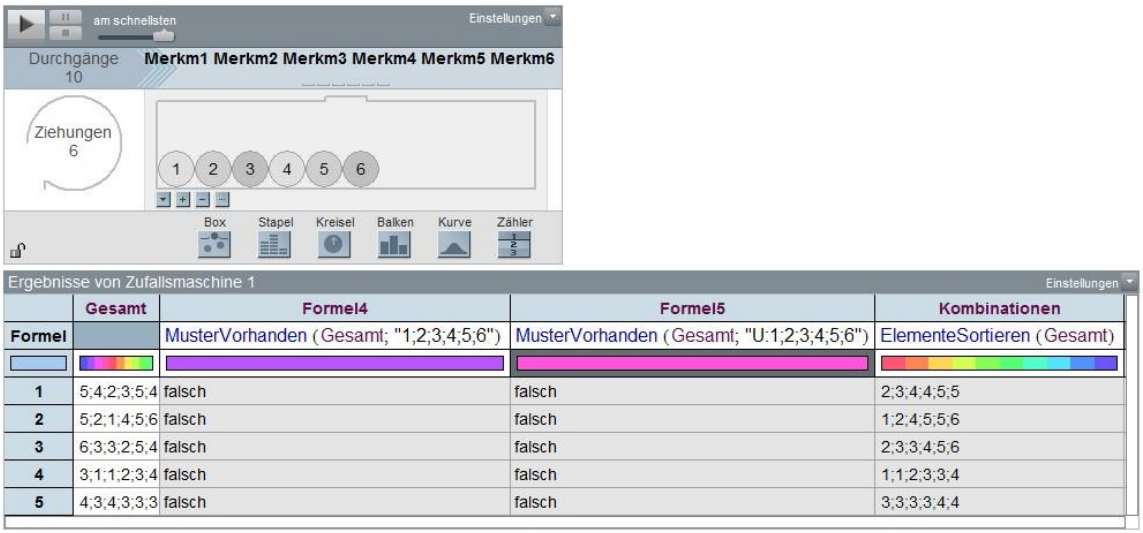
Zufallsgröße	Funktion in TinkerPlots	Merkmalsname in Abbildung 57
„Erste gewürfelte Zahl“	erstesElement(Gesamt)	Formel
„Letzte gewürfelte Zahl“	letztesElement(Gesamt)	Formel2
„Dritte gewürfelte Zahl“	ntesElement(Gesamt; 3)	Formel3
„Augensumme“	SummeElemente(Gesamt)	Summe
„Anzahl der Sechsen“	AnzPassende(Gesamt; 6)	Anzahl_6

Tabelle 7 gibt zum Beispiel des sechsfachen Würfelwurfs (Abbildung 63) verschiedene Umsetzungen von Zufallsgrößen mit den Listenfunktionen aus.

Zudem können durch die Listenfunktionen verschiedene Ereignisse in der Ergebnistabelle realisiert werden. Auch diese erfolgen zum sechsfachen Würfelwurf.

Mit dem Befehl *MusterVorhanden* lässt sich prüfen, ob bestimmte Zeichen oder Zeichenfolgen aufgetreten sind. Im Fall des sechsfachen Würfelwurfs beispielsweise, ob die Folge 1, 2, 3, 4, 5, 6 gefallen ist. Dies geschieht in geordneter Form, für die die Reihenfolge der Ziehungen eine Rolle spielt, mit *MusterVorhanden(Gesamt; „1;2;3;4;5;6“)* (Abbildung 64, Spalte „Formel4) oder in ungeordneter Form *MusterVorhanden(Gesamt; „U:1;2;3;4;5;6“)* (Abbildung 64, Spalte „Formel5“), für die die Reihenfolge der einzelnen Zeichen keine Rolle spielt. Ausgabe bei beiden Funktionen

ist „wahr“, sofern das Muster vorliegt, ansonsten „falsch“. Mit der Funktion *Elemente-Sortieren* werden die einzelnen Werte aufsteigend sortiert (Abbildung 64, Spalte „Kombinationen“).



	Gesamt	Formel4	Formel5	Kombinationen
Formel		MusterVorhanden (Gesamt; "1;2;3;4;5;6")	MusterVorhanden (Gesamt; "U:1;2;3;4;5;6")	ElementeSortieren (Gesamt)
1	5;4;2;3;5;4	falsch	falsch	2;3;4;4;5;5
2	5;2;1;4;5;6	falsch	falsch	1;2;4;5;5;6
3	6;3;3;2;5;4	falsch	falsch	2;3;3;4;5;6
4	3;1;1;2;3;4	falsch	falsch	1;1;2;3;3;4
5	4;3;4;3;3;3	falsch	falsch	3;3;3;3;4;4

Abbildung 64: Der sechsfache Würfelwurf mit verschiedenen Ereignissen

Die Funktion *AnzahlElemente* ist hier im Beispiel nicht aufgenommen, da in jedem Durchgang sechs Ziehungen durchgeführt werden und diese Funktion somit immer sechs ausgibt.

Der sechsfache Würfelwurf wurde hier auf simultan-gebündelte Art simuliert. An dieser Stelle ist das Unterstützungspotential durch die Software TinkerPlots sehr hoch. Dies wird zusätzlich noch weiter erhöht, da für einige der hier gezeigten Realisierungen keine eigene Funktioneneingabe nötig ist. In der Ergebnistabelle gibt es unter Einstellungen die Möglichkeit, vordefinierte Ergebnismerkmale zu benutzen (Abbildung 65).

Abbildung 65: Aufrufen der voreingestellten Ergebnismerkmale

Dadurch öffnet sich ein Fenster, in dem bestimmte Ergebnismerkmale zur Auswertung ausgewählt werden können (Abbildung 66).

Abbildung 66: Fenster zu den vordefinierten Ergebnismerkmalen

Dieses Fenster erlaubt es dem Benutzer, bestimmte Auswertungsmerkmale zu benutzen, ohne die entsprechenden Funktionen in TinkerPlots eingeben zu müssen oder auch nur zu kennen. Dies gilt für die Funktion *SummeElemente*, die sich hinter dem Namen „Summe von Gesamt“ verbirgt und durch einen Klick auf das Kästchen automatisch als Ergebnismerkmal mit Namen „Summe“ der Tabelle hinzugefügt wird (vgl. Abbildung 63, Spalte „Summe“); für die Funktion *ElementeSortieren*, die sich hinter „Sortierung von Gesamt“ verbirgt und durch einen Klick auf das Kästchen automatisch ein Ergebnismerkmal mit Namen „Kombinationen“ erzeugt (vgl. Abbildung 64, Spalte „Kombinationen“), und für die Funktion *AnzahlElemente*, für die sich durch einen Klick auf das Kästchen bei „‘?‘ in Gesamt zählen“ automatisch ein weiteres Fenster öffnet, in welchem eingegeben werden kann, welches Element gezählt werden soll (Abbildung 67). Nach der Eingabe in dieses Fenster wird der Tabelle automatisch eine Merkmalsspalte mit Namen „Anzahl_*Eingabe*“ hinzugefügt (vgl. Abbildung 63 Spalte „Anzahl_6“).

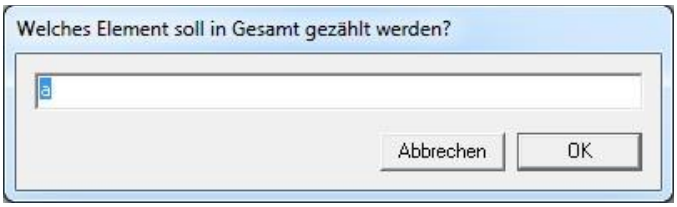


Abbildung 67: Abfragefenster zur Funktion *AnzahlElemente*

Weiter können über das Fenster zu den Ergebnismerkmalen (Abbildung 66) bezogen auf Merkmal1, also die erste Ziehung in der Zufallsmaschine, zusätzliche Auswer-

tungsmerkmale durch Anklicken realisiert werden. Dies sind die Runlänge, ausgehend vom ersten Zeichen des Merkmals Merkm1; die Abweichung, absolute Häufigkeit und die prozentuale Häufigkeit zu einem gewählten Zeichen in Merkmal1. Weiterhin lässt sich die Wiederholungsnummer protokollieren, dies entspricht der Nummer der Ziehung, und die Durchgangsnummer, dies entspricht der Nummer, zum wievielten Male in der Zufallsmaschine auf Start geklickt wurde. Mit Hilfe des letzten Eintrags in Abbildung 66 „Merkmal durch Formel (?)“ öffnet sich der Formeleditor (Abbildung 68), in welchem eine Funktion von Hand eingegeben oder aus dem Verzeichnis ausgewählt werden kann.

Dies ist der aus Fathom übernommene Formeleditor, der um die neuen Listenfunktionen erweitert wurde. Sofern das Fenster zu den Ergebnismerkmalen benutzt wird (Abbildung 66), werden die neu erzeugten Spalten automatisch benannt. Dies geschieht sinnstiftend bei allen Einträgen (z. B. „Summe“ für die Funktion *SummeElemente*) bis auf den Eintrag „Merkmal durch Formel (?)“, für den fortlaufend nummeriert „Formel“ als Merkmalsname erscheint.

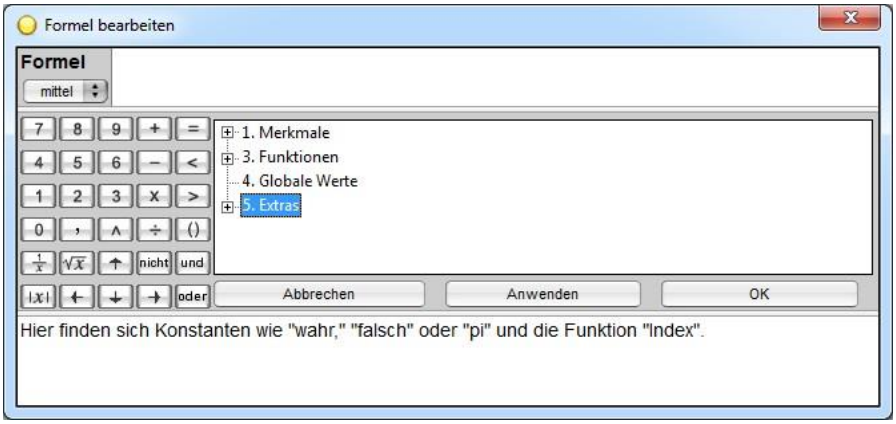


Abbildung 68: Der Formeleditor

Sowohl in Abbildung 63 als auch in Abbildung 64 ist die Formelzeile extra eingeblendet, dies ist nicht unbedingt nötig. Durch Benutzen des Fensters zu den Ergebnismerkmalen (Abbildung 66) werden die Spaltennamen automatisch vergeben, die Formelzeile hingegen wird nicht automatisch eingeblendet. Es sind also nur der Merkmalsname und die erzeugten Werte in der Tabelle zu sehen, die zugrunde liegende Funktion ist nicht sichtbar. Dies ist zweischneidig, zum einen erleichtert es das Benutzen der Software, weil eine Namensvergabe für ein neues Merkmal nicht zwingend erforderlich ist, sondern das Programm automatisch dafür sorgt, und der Benutzer muss sich nicht weiter um die Funktion kümmern, die das Programm im Hintergrund verwendet. Andererseits jedoch muss der Benutzer nicht darüber nachdenken, welche Funktion gerade ausgeführt wird und wie diese am Besten im Merkmalsnamen zu beschreiben ist. Das Umbenennen vorhandener Merkmalsnamen (und auch weiterer Variablennamen) kann in Tin-

kerPlots für den Benutzer sehr sinnvoll sein, um sinnstiftende Bezeichnungen für die einzelnen Variablen zu vergeben. Auf der anderen Seite beschleunigt es den gesamten Prozess, wenn sich der Benutzer um diese Dinge nicht kümmern muss.

Merkmale können außer über das oben beschriebene Ergebnismerkmale-Fenster auch - für den Benutzer etwas schwieriger - direkt in der Tabelle erzeugt und über Funktionen selbst definiert werden. Dazu muss die Formelzeile in der Tabelle eingeblendet werden und über diese kann dann der Formeleditor aufgerufen werden. Auf diesem Wege muss jedoch jeder neuen Spalte aktiv ein Merkmalsname durch den Benutzer gegeben werden, ein zwingender Vorgang, auf den TinkerPlots sonst verzichtet. Soweit es geht, ist die Variablenbenennung automatisiert, so dass der Nutzer diese nicht vornehmen muss. Die Beschriftung kann überall geändert werden. Der große Vorteil der Software ist jedoch, dass dies nicht zwingend erforderlich ist und somit eine Quelle von möglichen Schwierigkeiten direkt ausgeschlossen ist. Ein zusätzlicher Vorteil der vordefinierten Merkmale des Ergebnismerkmale-Fensters ist der, dass sämtliche Funktionen, die zugrunde liegen, im Hintergrund bleiben, für den Nutzer also nicht sichtbar sind. Die beiden Funktionen *AnzElemente* und *SummeElemente*, die die Anzahl eines bestimmten Zeichens oder einer bestimmten Zeichenkette in einem Durchgang ausgeben bzw. die Summe der Elemente in einem Durchgang berechnen, können sehr universell eingesetzt werden und helfen bei vielen Modellierungen von Zufallsgrößen und Ereignissen in Verbindung mit dem Graphen weiter.

Durch die Kombination des neuen Merkmals „Gesamt“, der neuen Listenfunktionen und des Ergebnismerkmal-Fensters ist das Unterstützungspotential von TinkerPlots zur Umsetzung von Ereignissen und Zufallsgrößen sehr hoch. Sehr benutzerfreundlich wird durch diese Komponenten das Definieren von Merkmalen und somit die Umsetzung von Ereignissen und Zufallsgrößen gestaltet.

Bei aller intuitiven Nutzungsmöglichkeit der Zufallsmaschine ist zum sinnvollen Weiterarbeiten mit Messgrößen und Ereignissen jedoch ein gewisses Maß an Abstraktion bei der Modellierung hilfreich bzw. notwendig. Zur Verdeutlichung wird noch einmal das Beispiel des Glücksrads aus Abschnitt 3.5.1 aufgegriffen. Mithilfe des Bauteils Kreisel wurde das Glücksrad mit der in Tabelle 8 angegebenen Wahrscheinlichkeitsverteilung in der Zufallsmaschine modelliert (Abbildung 69).

Tabelle 8: Wahrscheinlichkeitsverteilung des Glücksradbeispiels

k (Gewinn in €)	0	5	10
P(X=k)	0,5	0,25	0,25

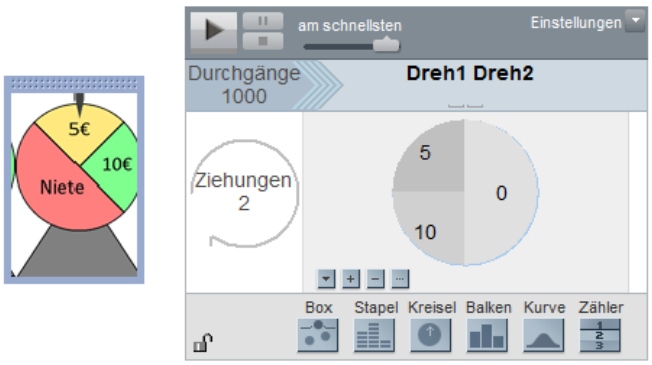


Abbildung 69: Links: Zu modellierendes Glücksrad; Rechts: Umsetzung in der Zufallsmaschine

Soll nun die Zufallsgröße X: „Gewinnsumme nach zweimaligem Drehen“ dargestellt werden, so lässt sich dies durch das vordefinierte Merkmal *Summe* direkt realisieren (Abbildung 70). Die Einträge des Merkmals „Gesamt“ aus der Tabelle in Abbildung 70 werden als Gewinn aufsummiert.

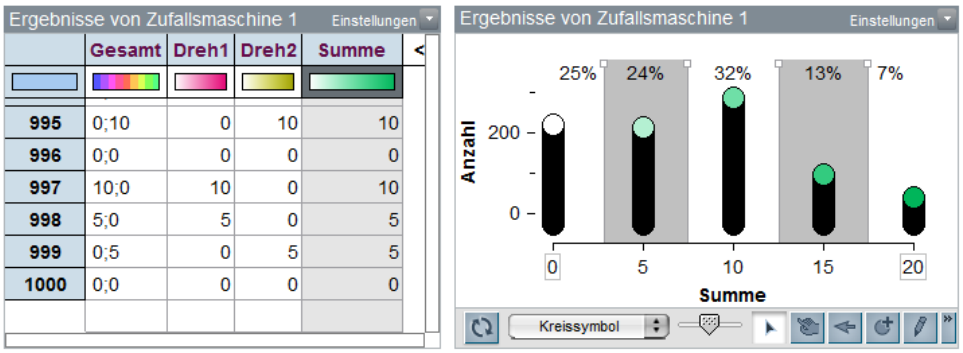


Abbildung 70: Darstellung der Zufallsgröße "Summe nach zweimaligem Drehen des Glücksrads"

Wie in Abschnitt 3.5.1 erwähnt, lässt sich mithilfe der Zufallsmaschine das Glücksrad auch direkt abbilden mit der Beschriftung 5€, 10€ und Niete wie in Abbildung 71 zu sehen.

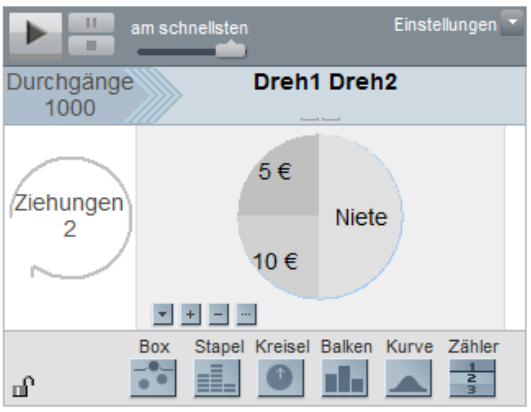


Abbildung 71: Modellierung des Glücksrads auf niedrigstem Abstraktionsniveau

Diese Modellierung weicht nun kaum von der obigen in Abbildung 69 ab, kann jedoch aus Nutzersicht näherliegend sein, weil ein direktes Abbilden der Beispielgraphik möglich ist. Allerdings wird das Definieren der Zufallsgröße hier zur Herausforderung für den Nutzer, weil das vordefinierte Merkmal *Summe* nicht benutzt werden kann aufgrund der nicht-numerischen Beschriftung der Segmente. Ebenso wenig können die Ausprägungen in einem neuen Ergebnismerkmal addiert werden, da es keine numerischen Werte sind. Von Hand können jedoch zwei weitere Merkmale definiert werden, die die Ausprägungen der beiden Drehungen in numerische Werte mittels der Funktion *transform* übertragen. Diese beiden neuen Merkmale können dann in einem dritten Merkmal, das die Zufallsgröße darstellt, addiert werden (Abbildung 72).

Diese Vorgehensweise widerspricht dem einfachen Modellierungscharakter der Software. Hat ein Benutzer die direkte Modellierung in Abbildung 71 gewählt und stellt fest, dass es für ihn schwierig ist, die benötigte Zufallsgröße zu definieren, lässt sich natürlich die Beschriftung der Segmente leicht ändern, so dass die Abstraktion während des Arbeitens mit der Zufallsmaschine stattfinden kann und somit ein einfacheres Arbeiten beim Realisieren der Zufallsgröße möglich ist.

Ergebnisse von Zufallsmaschine 1 — Einstellungen

	Gesamt	Dreh1	Dreh2	Dreh1_Euro	Dreh2_Euro	Gewinn
Formel				transform (Dreh1) { ("5 €") : 5; ("10 €") : 10; sonst : 0	transform (Dreh2) { ("5 €") : 5; ("10 €") : 10; sonst : 0	Dreh1_Euro + Dreh2_Euro
996	10 €;Niete	10 €	Niete	10	0	10
997	Niete;Niete	Niete	Niete	0	0	0
998	10 €;Niete	10 €	Niete	10	0	10
999	5 €;5 €	5 €	5 €	5	5	10
1000	Niete;10 €	Niete	10 €	0	10	10

Abbildung 72: Merkmalsdefinitionen zur Realisierung der Zufallsgröße „Gewinn“ bei der direkten Glücksradmodellierung

An diesem Beispiel wurde gezeigt, dass wie oben dargestellt zwar viele intuitive Modellierungsmöglichkeiten mit dem Objekt der Zufallsmaschine gegeben sind, die weiteren Interfaces jedoch im Blick behalten werden sollten, um das Definieren von Ereignissen und Zufallsgrößen nicht unnötig zu erschweren.

Bereich 7: Merkmale durch Messgrößen auswerten

Das Auswerten von Merkmalen geschieht in TinkerPlots ausschließlich über Graphen. Hierbei gibt es nun zwei Möglichkeiten. Entweder liegt bereits das benötigte Merkmal als Realisierung einer Zufallsgröße oder eines Ereignisses in einer Tabelle vor, wie das bei der simultanen Simulation der Fall ist. Dann können deren Verteilung und zusätzlich benötigte Kennzahlen wie zum Beispiel das arithmetische Mittel, Quantile oder bestimmte Häufigkeiten in einem Graphen dargestellt werden. Oder die interessierenden

Ereignisse und Zufallsgrößen lassen sich nicht als Merkmal in der Tabelle realisieren, weil die Simulation sequentiell ausgeführt wurde. Dabei ist es nicht entscheidend, ob die Simulation nun zwingend sequentiell ausgeführt wurde oder weil es dem subjektiven Nutzerprofil entspricht, aber es führt dazu, dass sich Ereignisse und Zufallsgrößen nicht über Merkmale der Ergebnistabelle realisieren lassen. An dieser Stelle werden Messgrößen benötigt, über die die gesuchten Ereignisse und Zufallsgrößen realisiert werden. Das gilt ebenso für die Simulation durch Randomisierung, auch hier müssen Messgrößen zur Definition der interessierenden Ereignisse und Zufallsgrößen genutzt werden.

Wir betrachten hier noch einmal den sechsfachen Würfelwurf, zu dem die Anzahl der gefallenen Sechsen als Zufallsgröße interessiert. Oben wurde dies bereits über die simultane Simulation in Verbindung mit der Listenfunktion *AnzahlPassende* ausgeführt. Nun soll die gleiche Simulation sequentiell erfolgen.

Abbildung 73: Sequentielle Simulation des sechsfachen Würfelwurfs

In Abbildung 73 wird aus der Box einmal gezogen und der sechsfache Würfelwurf über sechs Durchgänge realisiert. Somit existiert nur ein Merkmal (Merkm1), welches zeilenweise den sechsfachen Würfelwurf repräsentiert. Durch diese Simulationsart ist in der Tabelle nur ein sechsfacher Wurf realisierbar. Im Graphen in Abbildung 73 (rechts) ist noch einmal der sechsfache Würfelwurf repräsentiert und so dargestellt, dass die Anzahl der gefallenen Sechsen ablesbar ist.

Um nun dies als Zufallsgröße zu realisieren, muss die „Anzahl der 6en im sechsfachen Würfelwurf“ als Messgröße definiert werden. Die zwei möglichen Wege zur Auswahl einer Messgröße sind entweder über das Auswahlmenü der rechten Maustaste (Abbildung 74) oder über das Betätigen des M-Buttons unten in der Graphik (Abbildung 75) und die anschließende Wahl der zu sammelnden Größe.

Beide Wege zum Messgrößensammeln über den Graphen sind intuitiv. Der große Vorteil am Messgrößenkonzept in TinkerPlots ist der, dass es ohne Formeln auskommt.

Abbildung 74: Definieren einer Messgröße über das Menü

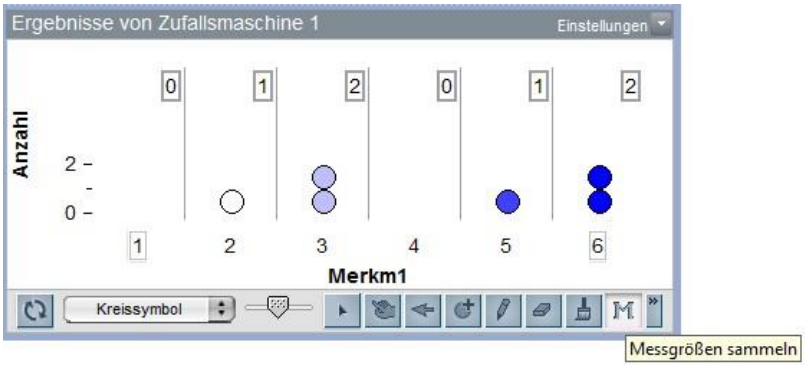


Abbildung 75: Definieren einer Messgröße über den Messgrößen-Button und Anklicken der zu sammelnden Größe

Ist eine Messgröße implizit in einer Graphik definiert, wird sie automatisch in einer Messgrößentabelle als eigenes Merkmal abgelegt (Abbildung 76). Auch bei einer Messgröße wird in TinkerPlots der Merkmalsname automatisch vergeben.

Abbildung 76: Automatisch erzeugte Messgrößentabelle

Da im obigen Beispiel eine absolute Häufigkeit gesammelt wird, beginnt der Messgrößenname in Abbildung 76 mit „Anzahl_“. Anschließend folgt der Merkmalsname, von dem die Messgröße gesammelt wird, gefolgt von der Ausprägung, zu der die Anzahl gesammelt wird, so dass sich insgesamt, automatisch vergeben, der Name „Anzahl_Merkm1_6“ ergibt. Diese Struktur ist für jegliche Sammlung von Messgrößen vorgegeben. Relative Häufigkeiten werden beginnend mit „Anteil_“ bezeichnet, Messgrößen von Mittelwerten beginnen entsprechend mit „arithmetisches Mittel_“ oder „Median_“ usw. Auch hier ist eine Umbenennung des Messgrößenmerkmals möglich, jedoch durch die intelligente automatische Beschriftung nicht nötig. Dies trägt wiede-

rum zum hohen Unterstützungspotential der Software bei, da hierdurch das Arbeiten wesentlich erleichtert wird. Wie bereits oben angesprochen ist auf der anderen Seite auch kein Nachdenken über eine sinnvolle Beschriftung nötig, was durchaus zu einer didaktischen Verringerung führen kann.

Messgrößen werden in TinkerPlots standardmäßig über Graphen definiert. Eine Messgrößendefinition in einer Tabelle ist nicht möglich. In Graphen können statistische Kennzahlen, absolute oder relative Häufigkeiten, etc. als Messgrößen ausgewählt werden. Auch hier ist die Software benutzerorientiert gestaltet, denn durch den Button „M“ im Graphen lassen sich direkt alle Werte erkennen, die zur Messgrößendefinition geeignet sind.

In Abbildung 78 sind beispielhaft Ergebnisse des sechsfachen Würfelwurfs zu sehen. Als Messgrößen bezogen auf ein Merkmal können Mittelwerte, Kennzahlen des Hut- oder Boxplots, mit Hilfe der Lineal-Funktion gemessene Abstände und relative und absolute Häufigkeiten gesammelt werden. Alle diese Werte haben nach Betätigen des M-Buttons (Abbildung 78) eine graue Kastenumrandung. Dies bezieht sich auf ein Merkmal.

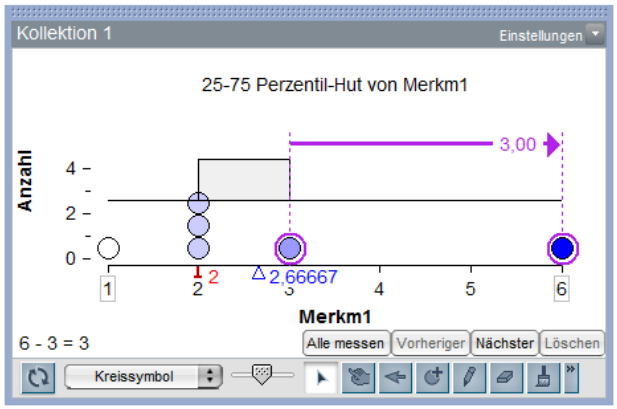


Abbildung 77: Graphik ohne gedrückten M-Button

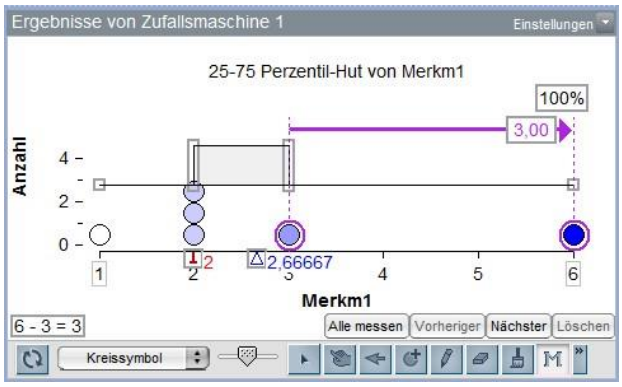


Abbildung 78: Kastenmarkierung um Werte, die als Messgröße gesammelt werden können

Bezogen auf zwei Merkmale können zusätzlich sämtliche mit der Lineal-Funktion ermittelten Abstände als Messgröße gesammelt werden. Hier hört die Unterstützung durch die Software allerdings auf.[18] Allerdings können in der Messgrößentabelle (Abbildung 76) nun weitere Merkmale definiert werden, die auf den gesammelten Messgrößen operieren. Dazu stehen die Funktionen aus dem Formeleditor zur Verfügung, die zur statistischen Auswertung zweier Merkmale implementiert sind. Es sei an dieser Stelle noch einmal bemerkt, dass über die Messgrößentabelle nicht per Funktionen auf Merkmale der Zufallsmaschine zugegriffen werden kann.

Bereich 8: Anzahl Messgrößen

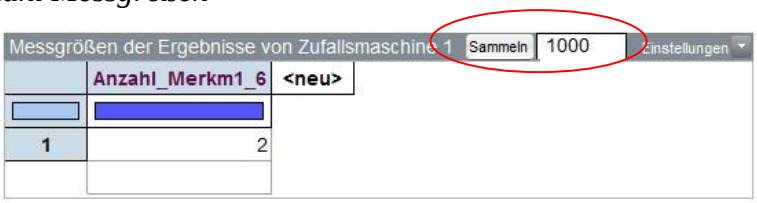


Abbildung 79: Bereich 8: Messgrößen sammeln

Das Interface Wiederholung des Zufallsexperiments bei sequentieller Simulation wird über Bereich 8, das Hinzufügen von Messgrößen, bedient. Bereich 8 umfasst das Feld zur Eingabe zusätzlich zu sammelnder Messgrößen. Hier ist die Anzahl auf 5000 zu sammelnde Messgrößen begrenzt. Da die Messgrößen jedoch bei jedem Vorgang den vorhandenen Messgrößen hinzugefügt werden, können über mehrere Sammelvorgänge auch leicht höhere Anzahlen als 5000 gesammelt werden. Für jede Messgröße werden durch die zugehörige Zufallsmaschine neue Zufallswerte generiert, von denen jeweils der Wert der Messgröße in der Messgrößentabelle gesammelt wird. Zu einem solchen Vorgang können auch mehrere Messgrößen definiert werden, so dass verschiedene Umsetzungen von Zufallsgrößen zu einer Modellierung möglich sind.

Von Bedeutung aus Nutzersicht ist noch die Geschwindigkeit, mit der das Messgrößensammeln geschieht. Im Gegensatz zur simultanen Simulation benötigt die sequentielle Simulation durch das Sammeln von Messgrößen mehr Rechenleistung und damit mehr Zeit. Dabei gilt, je schneller das zugrunde liegende System des verwendeten PCs ist, desto schneller ist auch der Sammelvorgang. Dennoch kann es bei komplexen Zufallsmaschinen, damit ist gemeint, dass eine Zufallsmaschine mehrere Abzweigungen und

[18] Es gibt für den Experten des Systems noch einen weiteren Weg, Messgrößen zu definieren. Bei aktiviertem Messgrößentabellenobjekt kann über den Menüeintrag „Info Kollektion“ unter Objekt das aus Fathom importierte Fenster zur Definition von Messgrößen geöffnet werden. Hier ist nun auch eine Definition von Messgrößen über Funktionen möglich. Dies ist jedoch für den normalen Benutzer von TinkerPlots, der zudem in den seltensten Fällen auch noch ein Kenner des Systems Fathom sein wird, äußerst komplex und im Sinne der einfach zu bedienenden Software folgt dies auch nicht mehr dem eigenen Anspruch der Software.

dadurch auch mehrere Ziehungen hat, dazu kommen, dass der Sammelvorgang bei einer entsprechend großen Anzahl zu sammelnder Messgrößen von ein paar Sekunden bis hin zu einigen (wenigen) Minuten dauert. Der Vorgang des Messgrößensammelns kann jedoch im Vorfeld beschleunigt werden durch zwei Schritte. Erstens sollte auf jeden Fall die Animation zum Messgrößensammeln ausgeschaltet werden. Diese Animation hat keinen weiteren Einfluss auf die Simulation, es ist auch an keiner Stelle eine zusätzliche Visualisierung irgendeines Vorgangs zu sehen. Deshalb sollte die Animation ausgeschaltet werden unter *Einstellungen zu Messgrößen → Markierung entfernen* bei *Animation an*. Das zweite, was einen erheblichen Einfluss auf die Geschwindigkeit beim Messgrößensammeln hat, ist die Darstellungsleistung, die für jeden Durchgang der Zufallsmaschine benötigt wird. Bei jeder zu sammelnden Messgröße lässt TinkerPlots die Zufallsmaschine neu laufen, dadurch werden alle Einträge der Ergebnistabelle neu erzeugt und die Darstellung in der zugehörigen Graphik wird auch jedes Mal neu generiert. Es wirkt beschleunigend, wenn alle diese Objekte minimiert werden, wie in Abbildung 80 zu sehen.

Abbildung 80: Minimierte Objekte: Zufallsmaschine (link), Ergebnistabelle (mittig) und Graphik (rechts)

Abgesehen von der Zeit, die das Messgrößensammeln in Anspruch nimmt, ist eine sequentielle Simulation im Wesentlichen nur anders und nicht unbedingt anspruchsvoller an den Benutzer als eine simultane Simulation.

Exkurs: Messgrößensammeln als didaktisches Werkzeug

Das Messgrößensammeln kann aus didaktischer Sicht auch benutzt werden, um abseits von der Definition von Ereignissen und Zufallsgrößen, die Schwankungen von Ergebnissen darzustellen. Dies ist gerade für weniger erfahrene Nutzer ein Vorteil, da sich durch das einfache Sammeln von Messgrößen ein Gefühl dafür entwickeln lässt, wie sich zu einer gegebenen Wiederholungsanzahl die Schwankung eines Ergebnisses verhält. Dazu ein einfaches Beispiel: Der sechsfache Würfelwurf soll hundert Mal simuliert werden. Es interessiert, wie häufig es vorkommt, dass keine Sechs fällt.

Als Ergebnis erhält man in der Simulation aus Abbildung 81, dass in 37 von 100 Fällen keine Sechs beim sechsfachen Würfelwurf aufgetreten ist.

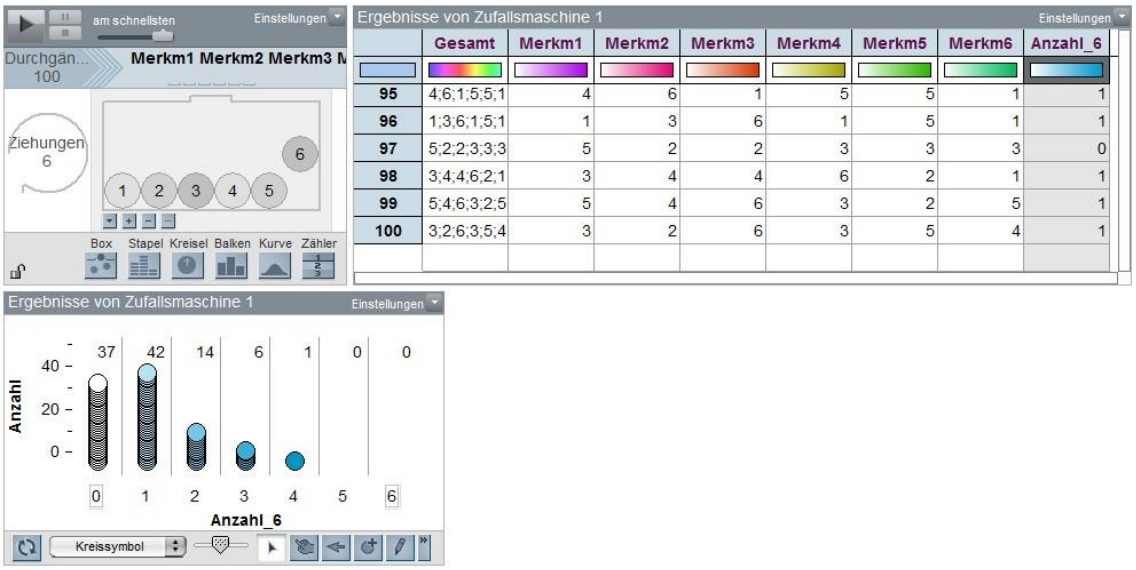


Abbildung 81: Simulation zu der Anzahl an Sechsen beim sechsfachen Würfelwurf

Durch mehrfache Wiederholung der Simulation über den Playbutton lässt sich sehen, dass diese Anzahl schwankt (Abbildung 82).

Abbildung 82: Drei Beispielverteilungen zur Anzahl der Sechsen zur Simulation des sechsfachen Würfelwurfs

Nun kann das Messgrößenkonzept als Protokollfunktion eingesetzt werden, um die Häufigkeit des Ereignisses „es ist keine Sechs aufgetreten“ zu dokumentieren (Abbildung 83).

Ohne dass nun tiefergehende stochastische Hintergründe thematisiert werden müssen, lässt sich anhand von ein paar Durchläufen ein erstes Gefühl für die Schwankungen bei einer Wiederholungszahl von n = 100 entwickeln. Als Beispiel wurden in Abbildung 83 zehn Durchgänge der Zufallsmaschine durch Messgrößen protokolliert. Dies lässt sich natürlich analog übertragen für beliebige n.

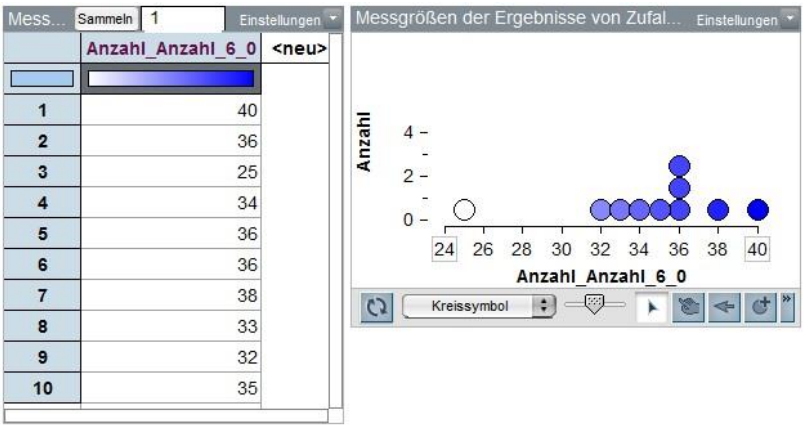


Abbildung 83: Zehn Messgrößen zum Ereignis "Es ist keine Sechs aufgetreten"

Vergleich und Bewertung

Abhängig von der Simulationsart können Zufallsgrößen und Ereignisse in Merkmalen oder in Messgrößen modelliert werden. Dabei ist die Definition von Zufallsgrößen und Ereignissen in Merkmalen in der Tabelle für die simultane Simulationsart über Funktionen möglich. Sofern eine Zufallsgröße oder ein Ereignis über das Ergebnismerkmale-Fenster definiert werden kann, wird keine aktive Funktioneneingabe durch den Benutzer nötig. Hier bietet TinkerPlots hohes Unterstützungspotential an. Allerdings lassen sich auf diesem komfortablen Weg nur einige wenige Zufallsgrößen und Ereignisse realisieren. Die zeilenweise Summe aller durch Simulation erzeugten Merkmale und die Anzahl bestimmter Zeichen, z. B. die Anzahl der Sechsen beim mehrfachen Würfelwurf unter allen durch Simulation erzeugten Merkmalen gehören hierzu. Gerade die Möglichkeit, per einfachem Klick die Anzahl bestimmter Ereignisse zu bestimmen, ist für viele Modellierungen sehr nutzerfreundlich. Für alle anderen Realisierungen von Zufallsgrößen und Ereignissen müssen „von Hand“ Funktionen eingegeben werden. Dabei wiederum bieten die Listenfunktionen in Kombination mit dem Merkmal „Gesamt“ gute Unterstützung, da durch sie Ereignisse und Zufallsgrößen durch simultane Simulation realisiert werden können, die andernfalls nur auf sequentielle Simulationsart realisiert werden könnten. Gerade die vielen Möglichkeiten der Funktion *MusterVorhanden* ermöglichen die Umsetzung verschiedener Muster oder Zeichenabfolgen als Ereignis, wie z. B. Full House, oder eine Straße beim Kniffel, was sich bei Fathom als Schwierigkeit herausgestellt hat (vgl. Maxara 2009, 92f). Allerdings ist hier eine gute Formelkenntnis von Nöten, was je nach Nutzer eine Herausforderung darstellen kann.

Da das Messgrößenkonzept in TinkerPlots auf Formeln zur Definition von Messgrößen verzichtet, ist es kaum fehleranfällig. Dies und die einfache Auswahl der zu sammelnden Größe machen das Konzept für den Benutzer leicht zugänglich, so dass auch die sequentielle Simulation eine gut benutzbare Simulationsart in TinkerPlots darstellt. Der Vorteil an der Umsetzung von Ereignissen und Zufallsgrößen über Messgrößen ist der,

dass dies ohne Eingabe von Formeln geschehen kann. Allerdings sind hier am gesamten Simulationsprozess mehr Objekte beteiligt, zusätzlich zu der Zufallsmaschine, der Tabelle und dem Graphen, die bereits für die simultane Simulation benötigt werden, kommen noch die Messgrößentabelle und der Graph für die Messgrößenverteilung zur Auswertung hinzu. Eine gute Strukturierung der Arbeitsfläche durch den Benutzer ist hier nötig, um die Übersicht zu behalten.

Das Umsetzen von Ereignissen und Zufallsgrößen in Merkmalen ist in TinkerPlots sehr umfassend möglich, und sofern dies über das Ergebnismerkmale-Fenster geschehen kann auch mit hohem Unterstützungspotential durch die Software gegeben. Die Funktioneneingabe über den Formeleditor für alle weiteren Merkmalsdefinitionen ist ebenfalls gut, allerdings nicht mehr so intuitiv. Für die Umsetzung von Ereignissen und Zufallsgrößen in Messgrößen spricht, dass dies ein Vorgang ist, der nahezu ohne Funktioneneingabe auskommt. Individuell zu bewerten ist noch der Zeitfaktor, denn bei großen Wiederholungszahlen benötigt die Umsetzung in Messgrößen deutlich mehr Zeit als bei der Umsetzung in Merkmalen.

3.7 Simulieren mit Abbruchbedingungen

Eine Erweiterung der bisherigen Simulationsarten stellen stochastische Wartezeitprobleme dar. Diese können durch Simulationen mit Abbruchbedingungen modelliert werden. Abbruchbedingungen können sowohl für simultan als auch für sequentiell ausgeführte Simulationen definiert werden. Hierzu gibt es bei den Einstellungen zur Zufallsmaschine ein eigenes Feld, in das eine explizite Abbruchbedingung eingegeben werden kann. Ein gleiches Feld findet sich unter den Einstellungen zu Messgrößen im selben Fenster.

Für einige Modellierungen von Wartezeitproblemen ist das Unterstützungspotential der Software hoch, da die gewünschten Bedingungen und dazu benötigten Messgrößen ohne Funktioneneingabe auskommen. Andere Realisierungen hingegen sind komplizierter auszuführen, da eine gute Formelkompetenz benötigt wird. Messgrößen werden bei Wartezeitproblemen immer benötigt.

Als Beispiel eines Wartezeitproblems wird zunächst das Warten bis zur ersten Sechs bei einem Würfelwurf ausgeführt.

Der Würfel kann dazu in der Zufallsmaschine entsprechend der eigenen Vorliebe modelliert werden. In Abbildung 84 wurde eine Modellierung des Würfels mithilfe des Kreiselbauteils gewählt. Bei Ziehungen wird Eins eingetragen, da insgesamt ein Würfel gewürfelt wird. Unter Einstellungen lässt sich das Info-Fenster zur Zufallsmaschine öffnen, in welchem nun eingestellt werden kann, was die Stoppbedingung für die Zufallsmaschine ist. Standardmäßig ist dies der unterste Punkt „Wiederholen … Mal“, wobei die dort stehende Zahl der Anzahl der eingegebenen Durchgänge entspricht. Für die

Modellierung von Wartezeitproblemen von Interesse sind jedoch die beiden darüber stehenden Punkte „Wiederholen bis …“. Der obere Punkt ist vor allem dann interessant, wenn ein zugrundeliegendes Zufallsexperiment aus mehreren Teilexperimenten besteht. Aber auch für dieses Beispiel lässt sich dies verwenden: Interpretiert man das Auftreten einer Sechs beim Würfelwurf als Muster, so kann in dem Feld „Wiederholen bis folgendes Muster auftritt“ eine 6 eingetragen werden. Durch Starten der Zufallsmaschine über den Start-Button werden nun so viele Durchgänge durchgeführt, bis zum ersten Mal das Muster „6“ aufgetreten ist, also bis zum ersten Mal eine Sechs gefallen ist. Diese Eingabemöglichkeit bezieht sich auf alle möglichen Ergebnisse, die auftreten können. Getrennt durch Semikolon, wobei dieses Trennzeichen im gleichen Fenster oben gewählt werden kann, können nun beliebige Muster, geordnet oder ungeordnet über das Ankreuzen von „Beliebige Reihenfolge“ als Abbruchbedingung der Zufallsmaschine definiert werden. Hierzu ist es nur nötig, das gewünschte Muster zu kennen und semikolongetrennt einzugeben.

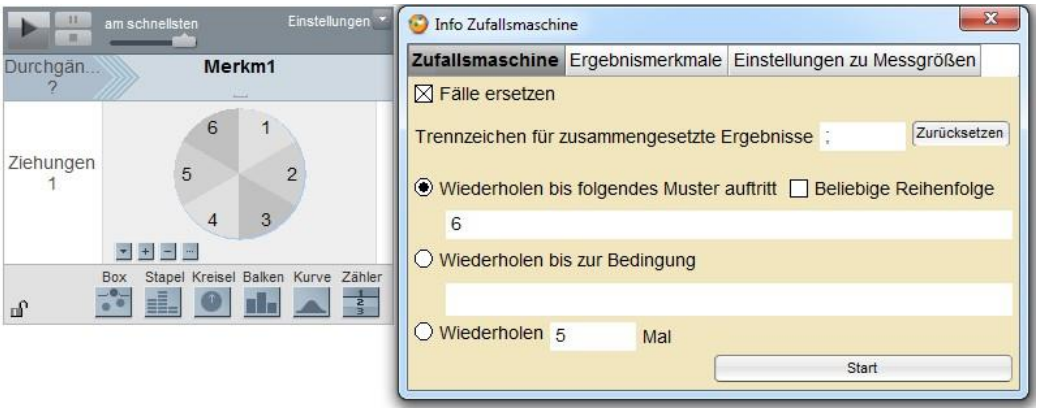


Abbildung 84: Die Zufallsmaschine mit zugehörigem Info-Fenster

Alternativ kann eine Abbruchbedingung über eine Funktion definiert werden. Dazu steht das zweite Feld zur Verfügung „Wiederholen bis zur Bedingung“ (Abbildung 85).

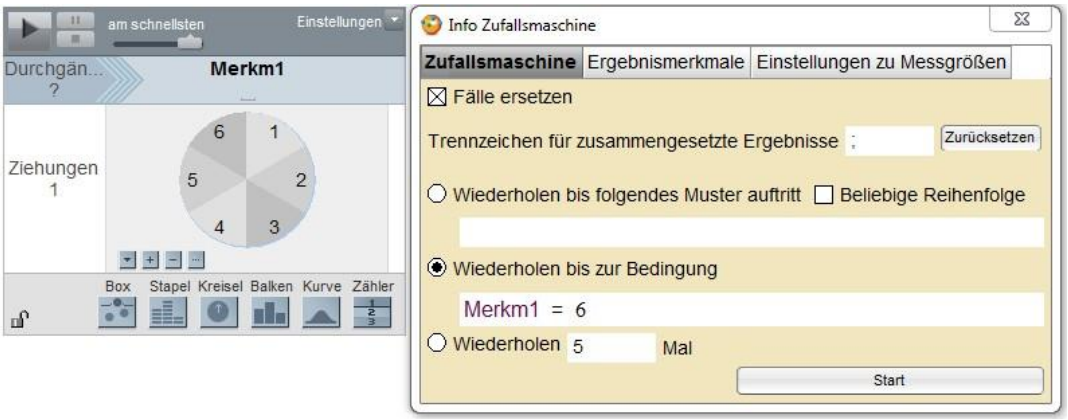


Abbildung 85: Die Zufallsmaschine mit Info-Fenster und eingegebener Bedingung

In Abbildung 85 ist das zweite Feld genutzt worden, um die Stoppbedingung für die Zufallsmaschine festzulegen. Hier benötigt der Benutzer nun Funktionen, da die Abbruchbedingung über den Formeleditor definiert werden muss. In diesem Fall soll die Zufallsmaschine solange laufen, bis bei Merkmal1 eine Sechs kommt.

Im Hintergrund läuft bei beiden Umsetzungen (Abbildung 84 und Abbildung 85) die Zufallsmaschine, bis eine 6 gekommen ist. Allerdings kann in beiden Umsetzungen nicht auf die zugrundeliegenden Formeln zugegriffen werden, die Simulationswerte erscheinen ohne Angabe einer Formel in weiß hinterlegten Zellen in Bereich 5a der Tabelle (Abbildung 86).

Abbildung 86: Warten bis zur ersten Sechs

Die Anzahl der Durchgänge entspricht nun der gesuchten Zufallsgröße *X: Anzahl Würfe bis zur ersten Sechs*. Diese Zufallsgröße kann über eine Messgröße realisiert werden. Auch hierzu gibt es nun wieder zwei Wege. Entweder wird die Anzahl der Fälle in einem Graphen dargestellt, als Anzahl eingeblendet und gesammelt (Abbildung 87).

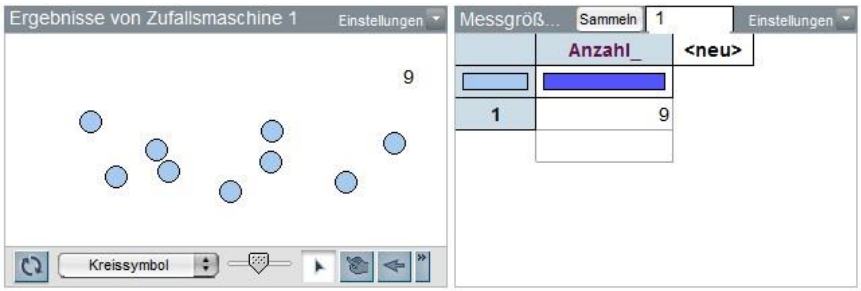


Abbildung 87: Anzahl Fälle bis zur ersten Sechs als Messgröße ausgewählt

Oder mit Hilfe des Info-Fensters zur Zufallsmaschine wird über die Einstellungen zu Messgrößen die Durchgangsanzahl als Messgröße definiert (Abbildung 88).

Mit Hilfe der verschiedenen Unterstützungsmöglichkeiten lässt sich ein solches Wartezeitproblem gut in TinkerPlots umsetzen.

Abbildung 88: Durchgangsanzahl als Messgröße ausgewählt

Als nächstes Beispiel wird der dreifache Würfelwurf betrachtet, bei dem gewürfelt werden soll, bis drei Sechsen erscheinen als Beispiel für ein Wartezeitproblem mit mehreren Teilexperimenten.

Dabei erfolgt die Simulation zunächst simultan. Es ist unerheblich, ob der dreifache Würfelwurf simultan-gebündelt oder simultan-hintereinandergeschaltet modelliert wird, Vorteile liegen hier in der Betrachtungsweise durch den Nutzer (s. o.). Im Wesentlichen erfolgt die Simulation wie oben für das einstufige Experiment.

Der dreifache Würfelwurf wird in Abbildung 89 wie oben über den Kreisel und die Anzahl 3 der Ziehungen modelliert. Das Ereignis „es fällt dreimal eine Sechs“ wird interpretiert als Muster „6;6;6“ und entsprechend in das Info Fenster der Zufallsmaschine in das Musterfeld eingegeben. Rechts in Abbildung 89 ist eine Simulation zu sehen, bei der im 80. Durchgang das Ereignis eingetreten ist.

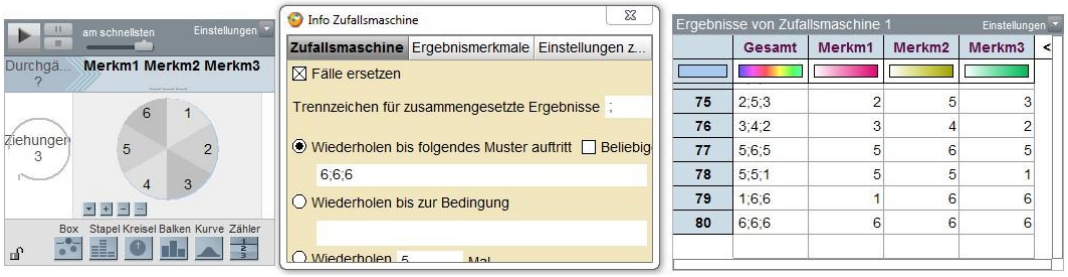


Abbildung 89: Dreifacher Würfelwurf bis dreimal eine Sechs erscheint

Auch für diese Simulation wird als Messgröße die Anzahl der Durchgänge benötigt (Abbildung 90).

Abbildung 90: Messgrößendefinition zum Ereignis "6;6;6" beim dreifachen Würfelwurf

Zur Benutzung der Option „Wiederholen bis folgendes Muster auftritt…“ als Abbruchbedingung gibt es noch zwei wichtige Anmerkungen. Anstelle eines bestimmten Ergebnisses kann auch ein Platzhalter verwendet werden. Das Symbol * nimmt dabei die Funktion einer beliebigen Ausprägung ein. Soll beispielsweise aus einer Urne mit drei roten und zwei blauen Kugeln solange gezogen werden, bis im zweiten Zug eine blaue Kugel gezogen wurde, so ist die Farbe der Kugel in der ersten Ziehung uninteressant. Dies lässt sich folgendermaßen modellieren (Abbildung 91):

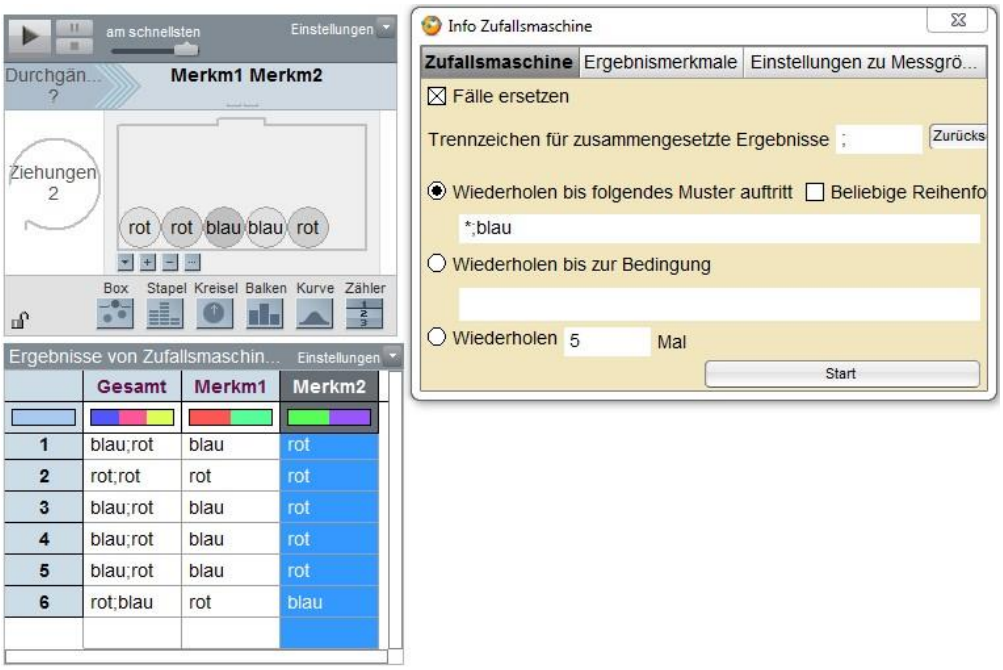


Abbildung 91: Modellieren einer Abbruchbedingung mit Platzhalter *

Die zweite Anmerkung zur Benutzung der Abbruchbedingung über die Eingabe eines Musters betrifft das Auftreten mehrerer gleicher, aber nicht festgelegter Zeichen. Es lassen sich über V1, V2, … Vn Variablennamen definieren. Soll zu der obigen Urne mit drei roten und zwei blauen Kugeln solange zweimal gezogen werden, bis zwei gleiche Farben gezogen wurden, so lässt sich dies mithilfe von Variablen festlegen (Abbildung 92).

Das Beispiel von oben zum dreifachen Würfelwurf mit dem interessierenden Ereignis „Es fällt dreimal eine Sechs“ als Abbruchbedingung lässt sich nicht sequentiell modellieren, weil TinkerPlots die Möglichkeit Messgrößen von Messgrößen zu sammeln nicht zur Verfügung stellt.[19]

[19] Für ausgewählte Simulationen lassen sich Messgrößen von Messgrößen sammeln, allerdings ließ sich auch in Zusammenarbeit mit dem Entwickler Cliff Konold nicht systematisch analysieren, für welche Situationen diese Möglichkeit gegeben ist und für welche nicht.

Wartezeitprobleme, denen wie dem obigen Beispiel eine explizite Abbruchbedingung innewohnt, lassen sich gut in TinkerPlots realisieren. Am einfachsten ist dies, wenn die Simulation simultan (gebündelt oder hintereinandergeschaltet) ausgeführt wird. Dazu wird die Abbruchbedingung in den Einstellungen zur Zufallsmaschine definiert. Sofern eine bestimmte Reihenfolge der Ergebnisse die Abbruchbedingung definieren kann (wie im obigen Beispiel die Reihenfolge „6;6;6“ als Abbruchbedingung für das Durchführen des dreifachen Würfelwurfs, bis dreimal eine Sechs erschienen ist), ist aus Nutzersicht die entsprechende Eingabe bei „Wiederholen bis folgendes Muster auftritt …“ die einfachste Variante. An dieser Stelle müssen keine Formeln verwendet werden, es kann per Auswahl definiert werden, ob die Reihenfolge oder nur das Vorkommen der Ergebnisse eine Rolle spielen soll und die Abbruchbedingung kann so angegeben werden, wie sie als Ergebnis vorliegen wird.

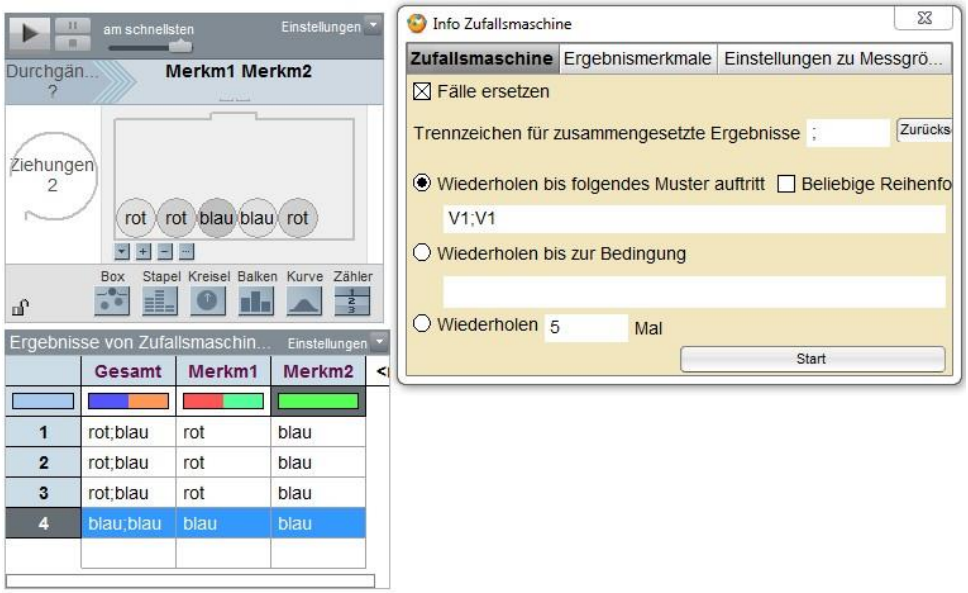


Abbildung 92: Modellieren einer Abbruchbedingung mit Variablen

Hier folgt TinkerPlots wiederum der Strategie, die Simulation stochastischer Problemstellungen so einfach wie möglich umzusetzen. Das mentale Modell, wie ein Ergebnis aussieht, bei dem der Zufallsversuch beendet wird, kann direkt in TinkerPlots eingegeben werden. Die Transparenz für den Nutzer ist an dieser Stelle nicht ganz so hoch wie bei bisherigen Umsetzungen, da die im Hintergrund arbeitende Funktion nicht aufgedeckt werden kann, allerdings hat das wenig praktische Relevanz. Trotzdem ist die Realisierung eines Wartezeitproblems, bei dem die Abbruchbedingung durch das Auftreten bestimmter Ergebnisse bestimmt ist, nutzerfreundlich gestaltet. Dabei haben der Platzhalter * und die Nutzungsmöglichkeit von Variablen einen wichtigen Stellenwert.

Über die zweite Variante, eine Abbruchbedingung durch eine Funktion zu definieren durch die Option „Wiederholen bis zur Bedingung…“ , können diese und weitere War-

tezeitprobleme modelliert werden. Es kann in der Zufallsmaschine wie bisher bei den entsprechenden Bauteilen mit oder ohne Zurücklegen eingestellt werden, so dass sich auch Zufallsexperimente mit abhängigen Teilexperimenten als Wartezeitproblem modellieren lassen. Die Eingabe einer Abbruchbedingung mithilfe einer Funktion erfordert beim Benutzer eine höhere Kompetenz als die reine Angabe eines Musters, die Bedienung einer Formeleingabe ist jedoch über den Formeleditor gelöst, so dass keine neue Syntax gelernt werden muss.

Die Funktion für eine Simulation eine Abbruchbedingung zu definieren, steht für simultane, aber nicht für sequentielle Simulationen zur Verfügung, es können auch für Messgrößen Abbruchbedingungen formuliert werden. Auch beim Stichprobenziehen und beim Randomisieren können in diesem Sinne Abbruchbedingungen formuliert werden.

Insgesamt erfordert das Simulieren mit Abbruchbedingungen eine gute Kenntnis des Systems.

3.8 Interface Auswerten

Jegliche Darstellung in TinkerPlots geschieht über Graphen.[20] Hierbei stellt TinkerPlots etwas Besonderes dar, denn es gibt keine voreingestellten Diagrammtypen, sondern die Daten liegen zunächst immer ungeordnet vor.[21] Jeder Datenpunkt wird dabei durch einen Kreis symbolisiert (Abbildung 93).

Abbildung 93: Ungeordnete Standarddarstellung von Daten

Es müssen alle Diagramme selbst erstellt werden. Dazu gibt es drei Grundoperationen, die mit *Trennen, Stapeln* und *Ordnen* bezeichnet sind (Abbildung 94).

---

[20] Auswertungen über Formeln in der Tabelle sind zwar möglich, werden aber von keinem mir bekannten Nutzer verwendet und in keiner mir bekannten Literatur erwähnt. Beispielsweise lässt sich der Mittelwert eines Merkmals durch die Funktion *aMittel(Merkmalsname)* in einer eigenen Spalte, also als eigenes Merkmal, darstellen. Jeder Eintrag zu diesem Merkmal besteht dann aus dem errechneten Mittelwert.

[21] An dieser Stelle soll keine Werkzeuganalyse zur Datenanalyse gemacht werden, weiterführende Informationen dazu finden sich zum Beispiel ausführlich bei Frischemeier (2017).

Abbildung 94: Die drei Grundoperationen zum Organisieren von Graphen

Es können sowohl konventionelle Graphen erstellt werden, wie zum Beispiel Kreisdiagramme oder Histogramme, aber auch eigene Darstellungsarten „erfunden“ werden.

Bei der graphischen Auswertung von Daten unterscheidet TinkerPlots zwischen kategorialen und numerischen Merkmalen. Für kategoriale Merkmale können die verschiedenen Ausprägungen auf einer Achse in unterschiedlichen Gruppierungen angeordnet werden, für numerische Merkmale ist eine Darstellung in frei wählbaren Klassenbreiten bis hin zu einer vollständig separierten Achse möglich. Der didaktische Gedanke dabei ist, eine subjektiv passende Graphik erstellen zu können und durch die verschiedenen Grundoperationen etwas über den Aufbau von Graphen zu lernen.

Abbildung 95: Graphische Menüleiste zur Organisation von Graphen

Über die symbolische Menüleiste (Abbildung 95) können verschiedene Funktionen für Graphen verwendet werden. Dazu gehört das Einzeichnen von Hilfslinien, das Einteilen einer numerischen Achse in frei wählbare Bereiche über sogenannte Einteiler, das Messen von Abständen zwischen Datenpunkten, das Einblenden von Hüten, die eine Vorstufe zu Boxplots darstellen, das Einzeichnen von Linien, das Einblenden von absoluten oder relativen Häufigkeiten (Anzahlen), das Einblenden verschiedener Mittelwerte, das Nutzen eines Fabmessgeräts, das Etikettieren der Datenpunkte und das Einblenden einer Legende. Hinter jeder Funktion verbergen sich weitere Einstellmöglichkeiten, durch die eine Vielzahl von Darstellungen und Auswertungen ermöglicht werden.

Jedes Merkmal erhält in TinkerPlots seine eigene Farbe. Kategoriale Merkmale erhalten für jede unterschiedliche Ausprägung disjunkte Farben, numerische Merkmale erhalten zu einer Farbe einen Farbverlauf von hell nach dunkel, entsprechend der Ausprägung (hell entspricht einem geringen Wert, dunkel entspricht einem hohen Wert). Es können bis zu drei Merkmale in einem Graphen dargestellt werden, wobei zwei Merkmale auf den Achsen liegen und das dritte Merkmal über seine Farbe Informationen widerspiegelt.

Es ist nicht möglich, Funktionen in Graphen einzuzeichnen, so dass einige Auswertungen nur rudimentär möglich sind. Es gibt jedoch die Möglichkeit, mit einem Stift in einem Graphen zu zeichnen, so dass händisch ein Funktionenverlauf eingezeichnet werden kann. Abstände dazu, im Sinne beispielsweise einer Passung einer Funktion an gegebene Daten, können dabei nicht ermittelt werden.

Ein Graph ist dynamisch mit der zugehörigen Tabelle verbunden, so dass durch das Auswählen eines Falles der entsprechende Fall in der Tabelle markiert wird. Dies funktioniert auch für die Auswahl mehrerer Datenpunkte und auch umgekehrt werden zu einer Auswahl in der Tabelle die entsprechenden Fälle im Graphen markiert.

Das Auswertungsinterface in TinkerPlots ist über Graphen benutzerfreundlich gestaltet und hat durch den Zwang, Diagramme selbst erstellen zu müssen, einen hohen didaktischen Wert. Die drei Grundoperationen *Stapeln, Trennen* und *Ordnen* tragen zum Verständnis der Datenorganisation bei und durch die weiteren Funktionen können viele Auswertungen vorgenommen werden. Dadurch ist die Anzahl an möglichen Darstellungen fast unbegrenzt, eine Gefahr liegt jedoch immer darin, dass ein unerfahrener Nutzer eine Darstellungsart wählt, die kaum interpretierbar ist. Ein großes Manko liegt darin, dass keine Funktionen eingezeichnet werden können.

Da dies hier eine Werkzeuganalyse im Hinblick auf die Simulationskapazität und nicht im Hinblick auf die Datenanalyse ist, sei an dieser Stelle auf verschiedene Literatur verwiesen, die sich mit dem Potential von TinkerPlots zur Datenanalyse beschäftigt (z. B. Frischemeier 2017).

3.9 Weitere Tools

Es gibt in TinkerPlots einige zusätzliche Objekte. Außer den Objekten *Datenkarten, Tabelle, Graph, Zufallsmaschine* und *Textfeld*, die durch ein Symbol in der oberen Iconleiste direkt ersichtlich und zugänglich sind, gibt es noch drei versteckte Objekte. Mit versteckt ist gemeint, dass diese nicht durch ein Symbol in der oberen Iconleiste abgebildet sind, sondern nur über den Menüeintrag Objekt → neues Objekt zugänglich sind. Dazu gehören die Objekte *Bild, Regler* und *Flashobjekt.* Über das Objekt Bild kann ein Bild in die Software importiert werden und über das Flashobjekt lässt sich zum Beispiel ein Lernvideo aus dem Internet einbinden. Über das Objekt Regler lassen sich Parameter einstellen. Allerdings können Regler nicht im Objekt der Zufallsmaschine eingesetzt werden, so dass sie zur Erzeugung von Zufall nicht genutzt werden können.[22] Für den Einsatz in Messgrößen gilt das gleiche, da Messgrößen in TinkerPlots nicht über eine Formel definiert werden, lassen sich auch dort keine Parameter einpflegen.[23] Somit bleibt der Einsatz von Parametern in Graphen vorbehalten. In Graphen lassen sich Pa-

[22] Natürlich lassen sich Parameter über Regler nutzen, wenn zur Zufallserzeugung nicht das Objekt der Zufallsmaschine verwendet wird, sondern Funktionen in der Tabelle. Wie bereits vorher erwähnt wurde, ist dieser Weg zur Erzeugung von Zufall eine große Ausnahme in TinkerPlots. Aus diesem Grund wird hier nicht weiter auf den Nutzen von Reglern zur Zufallserzeugung eingegangen, sondern auf Maxaras Ausführung zur Verwendung von Reglern in Fathom verwiesen, da dies direkt in TinkerPlots übertragen werden kann (Maxara, 2009, S. 108ff).

[23] Auch hier gilt wieder, dass es für den Experten sehr wohl möglich ist, eine Messgröße über eine Formel zu definieren. Allerdings ist dies eine derart versteckte Funktion, dass nur die Experten unter den Experten diese Möglichkeit kennen.

rameter ausschließlich in Filtern einsetzen, da wie oben erwähnt eine Funktioneneingabe im Graphen in TinkerPlots nicht möglich ist. Hier kann ein Regler vor allem Visualisierungszwecken dienen, z. B. um den Growing-Sample-Effekt zu zeigen. Allerdings lässt sich dies durch eine entsprechende Einstellung der Zufallsmaschine ebenfalls realisieren. Alle anderen Nutzarten von Reglern sind hinfällig, da in TinkerPlots das Objekt Zufallsmaschine bereits so designt wurde, dass Regler kaum noch zum Einsatz kommen.

4 Detailanalyse des Simulationspotentials von TinkerPlots

In diesem Kapitel wird ausgehend von stochastischen Problemstellungen die Unterstützungskapazität der Software TinkerPlots an Beispielen analysiert. Dabei wird nach den in Kapitel 3 gewonnenen Erkenntnissen und unterschiedenen Simulationsarten differenziert und das Potential der Software für den Nutzer herausgearbeitet. Weitgehend werden zur Analyse die Struktur und Beispiele aus Maxara (2009) verwendet, teilweise kritisch modifiziert. Maxara hat anhand einer umfangreichen Literaturrecherche vor allem in Schulbüchern und Lernmaterial typische stochastische Problemstellungen klassifiziert und daraus eine Struktur zur Analyse von stochastischer Software erarbeitet. Diese wird hier verwendet und für die Software TinkerPlots angepasst, um das Unterstützungspotential dieser Software aus stochastischer Sicht zu beleuchten.

Im vorherigen dritten Kapitel lag der Blick auf der Software und die Möglichkeiten, die diese bietet, wurden analysiert und vorgestellt. In diesem Kapitel liegt der Blick nun auf ausgewählten stochastischen Problemstellungen und anhand dieser wird untersucht, inwieweit TinkerPlots Unterstützung zu deren Simulation bietet. Einige Problemstellungen lassen sich auf unterschiedliche Arten simulieren, die verschiedenen Simulationsumsetzungen werden in diesem Kapitel vorgestellt und aus Nutzersicht bewertet. Bei anderen Problemstellungen wiederum ergibt sich, dass nur eine Simulationsart möglich ist. Durch die Machart der Zufallsmaschine wird in diesem Kapitel nicht zwischen den Simulationsarten simultane Simulation und Simulation durch Stichprobenziehung unterschieden, da bereits im dritten Kapitel ausgeführt wurde, dass die Simulation durch Stichprobenziehung in TinkerPlots als Unterart der simultanen Simulation angesehen werden kann. Besonders bewertet werden die zwei Interfaces *Modellieren* und *Ereignisse und Zufallsgrößen*, da diese beiden für den Nutzer den Hauptunterschied in der Bedienung ausmachen. Das Interface *Auswerten* ist bei jeder Simulation ähnlich, deshalb wird auf dessen Bewertung kein Fokus gelegt.

Im ersten Abschnitt findet sich eine Zusammenfassung der Klassifizierung stochastischer Problemstellungen nach Maxara (2009). Es werden einstufige und mehrstufige Zufallsexperimente unterschieden, wobei die mehrstufigen weiter differenziert werden. Zusätzlich werden Wartezeitprobleme erschlossen. Diese Klassifizierung wird dann im zweiten Abschnitt aufgegriffen, um mit Beispielen aufzuzeigen, wie entsprechende stochastische Situationen mit TinkerPlots simuliert werden können, welche Möglichkeiten sich für den Nutzer ergeben und wie diese zu bewerten sind.

Am Ende des Kapitels finden sich im letzten Abschnitt noch zwei Ergänzungen. Zum einen wird eine Lernumgebung vorgestellt, die das Gesetz der großen Zahl veranschaulicht und zum anderen wird eine Funktion vorgestellt, mit der sich die Inhalte der Zu-

fallsmaschine verbergen lassen und somit weitere Fragestellungen mit der Software eröffnet werden. Ein Fazit rundet die gesamte Werkzeuganalyse ab.

4.1 Klassifizierung stochastischer Problemstellungen

Grundlage für die Feinanalyse in diesem Kapitel bildet die ausgearbeitete Klassifizierung stochastischer Probleme, die Maxara (2009, 118ff) für die Analyse des Simulationspotentials der Software Fathom vorgenommen hat. Unterschieden werden

1.) Einstufige Zufallsexperimente,
2.) Mehrstufige Zufallsexperimente,
    a. mit stochastisch unabhängigen, verschiedenen Teilexperimenten,
    b. mit stochastisch abhängigen Teilexperimenten (verschieden oder auch gleich),
    c. Stochastische Prozesse/Markowketten,
3.) Wartezeitprobleme.

Mit dieser Einordnung wird ein großer Teil von Zufallsexperimenten abgedeckt, es wird jedoch nicht die Gesamtheit aller möglichen stochastischen Fragestellungen abgedeckt, da dies den allgemeinen Charakter dieser Arbeit übersteigen würde und sich in zu vielen Details verlieren würde.

Markowketten lassen sich in TinkerPlots nicht realisieren, deshalb wird in den folgenden Ausführungen nicht darauf eingegangen.

Bei allen im Folgenden betrachteten Vorgängen werden ideale Zufallsexperimente zugrunde gelegt. Prinzipiell ist dabei der genaue Ausgang eines Zufallsexperiments nicht vorhersagbar, aber alle möglichen Ausgänge sind im Vorfeld bekannt. Ebenso lässt sich zumindest theoretisch das Zufallsexperiment beliebig oft unter gleichen Bedingungen wiederholen. An dieser Stelle wird explizit die Wiederholbarkeit von Zufallsexperimenten gefordert und somit andere Lesarten des Begriffs Zufallsexperiment ausgeschlossen. Erst durch die Wiederholbarkeit lässt sich per Simulation ein Zugang zu Wahrscheinlichkeiten über relative Häufigkeiten herstellen.

4.1.1 Einstufige Zufallsexperimente

Typische Beispiele für einstufige Zufallsexperimente sind der einmalige Würfelwurf, der einmalige Münzwurf, das einmalige Drehen eines Glücksrads, das einmalige Ziehen aus einer Urne mit verschiedenfarbigen Kugeln, etc. Dabei sind einstufige Zufallsexperimente die Grundlage für mehrstufige Zufallsexperimente, so wie auch die Simulation einstufiger Zufallsexperimente die Grundlage für die Simulation mehrstufiger Zufallsexperimente bildet.

Die Menge der n möglichen Ergebnisse ω_i, i = 1, …, n, eines Zufallsexperiments wird

mit Ω bezeichnet und die betrachteten Ereignisse als Teilmengen von Ω werden mit Großbuchstaben A, B, C, etc. bezeichnet. Dabei wird von Zufallsexperimenten mit endlichen oder abzählbar unendlichen Ergebnisräumen ausgegangen. Zufallsexperimente mit stetigen Ergebnisräumen werden hier nicht betrachtet. Bei n möglichen Ausgängen eines einstufigen Zufallsexperiments ergibt sich folgendes:

$$\Omega = \{\omega_1;\ \omega_2; \dots;\ \omega_n\}$$

mit zugehörigen Wahrscheinlichkeiten $P(\omega_i) = p_i \in [0; 1]$, i = 1,…, n, mit $\sum p_i = 1$.

4.1.2 Mehrstufige Zufallsexperimente

Es können nun mehrstufige, stochastisch unabhängige Zufallsexperimente von mehrstufigen, stochastisch abhängigen Zufallsexperimenten unterschieden werden. Dabei können die Teilexperimente entweder verschiedenartig oder gleichartig sein.

Mehrstufige, stochastisch unabhängige Zufallsexperimente

Liegt ein k-stufiges, stochastisch unabhängiges Zufallsexperiment vor, bei dem die Teilexperimente gleich sind, so lässt sich der Ergebnisraum beschreiben als

$$\Omega = \Omega_1 \times \Omega_1 \times \dots \times \Omega_1 \text{ (k-mal } \Omega_1\text{)}$$

mit einem k-Tupel als Ergebnis $(\omega_1, \omega_2, \dots, \omega_k)$, ω_i aus Ω_i, i = 1, …, k

und der Wahrscheinlichkeitsverteilung

$$P(\omega) = P(\omega_1) \cdot P(\omega_2) \cdot \dots \cdot P(\omega_k).$$

Herrscht hierbei eine Gleichverteilung bei den Einzelexperimenten, so liegt ein Laplace-Experiment vor. Jedes Ω_1 besteht dann aus j Elementen, die jeweils mit Wahrscheinlichkeit $\frac{1}{j}$ auftreten. Für diesen Spezialfall vereinfacht sich die Wahrscheinlichkeitsverteilung zu

$$P(\omega) = \frac{1}{j^k},\ \omega \in \Omega.$$

Sind nun die k Teilexperimente eines mehrstufigen, stochastisch unabhängigen Zufallsexperiments verschieden, d. h. die Teilexperimente haben unterschiedliche Ergebnismengen und/oder Wahrscheinlichkeitsverteilungen, so können diese allgemein beschrieben werden durch

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_k$$

mit einem k-Tupel als Ergebnis $(\omega_1, \omega_2, \dots, \omega_k)$, wobei $\omega_i \in \Omega_i, i = 1, \dots, k$.

Für die zugehörige Wahrscheinlichkeitsverteilung gilt

$$P(\omega) = P_1(\omega_1) \cdot P_2(\omega_2) \cdot \dots \cdot P_k(\omega_k).$$

Dabei können die Wahrscheinlichkeitsverteilungen der einzelnen Teilexperimente gleichverteilt sein oder auch nicht.

Mehrstufige, stochastisch abhängige Zufallsexperimente

Wenn bei einem mehrstufigen Zufallsexperiment die Ergebnisse von Teilexperimenten von den Ergebnissen vorheriger Teilexperimente beeinflusst werden, so liegt ein mehrstufiges Zufallsexperiment vor, das stochastisch abhängig Teilexperimente beinhaltet.

Ein Spezialfall, der jedoch für viele Situationen adaptiert werden kann, liegt durch ein mehrstufiges Zufallsexperiment vor, das als k-maliges Ziehen von n Kugeln aus einer Urne ohne Zurücklegen modelliert werden kann. Bei diesem ändert sich die Wahrscheinlichkeit, eine Kugel zu ziehen, für jedes Teilexperiment, also auf jeder Stufe des Zufallsexperiments. Für den ersten Zug liegt die Wahrscheinlichkeit bei $\frac{1}{n}$ für das Herausziehen einer Kugel, auf der zweiten Stufe bei $\frac{1}{n-1}$ und auf der k-ten Stufe bei $\frac{1}{n-k+1}$, wobei $k \leq n$ gelten muss, da der letzte Ausdruck ansonsten unsinnig ist (und ansonsten mehr Ziehungen gemacht würden als Kugeln vorhanden wären).

Weiterhin können mehrstufige Zufallsexperimente unterschieden werden, die aus einer abhängigen Kombination verschiedener Teilexperimente bestehen, z. B. wird erst aus einer Urne mit r roten und s schwarzen Kugeln gezogen und anschließend, je nach gezogener Kugelfarbe entweder ein Würfel oder eine Münze geworfen. Solche Zufallsexperimente lassen sich nicht wie eben beschrieben durch *eine* Urnenziehung ohne Zurücklegen beschreiben. Hiermit ist gemeint, dass die Teilexperimente möglicherweise verschiedene Ergebnisräume haben, trotzdem jedoch voneinander abhängen.

4.1.3 Wartezeitprobleme

Wartezeitprobleme können nach unterschiedlichen Wartebedingungen klassifiziert werden. Ebenso können die Problemsituationen stochastisch abhängige oder auch stochastisch unabhängige Teilexperimente besitzen, die wiederum gleichartig sein können oder auch nicht. Maxara reduziert sinnvoll die Probleme für die Analyse von Fathom auf drei Arten von Wartezeitproblemen Maxara (2009, 128f), die hier ebenfalls betrachtet werden, da diese typische Situationen beschreiben, die in Lehrwerken verwendet werden.

Warten auf den ersten Treffer

Dieser Klasse von Wartezeitproblemen liegen stochastisch unabhängige Teilexperimente zugrunde mit gleichbleibender Trefferwahrscheinlichkeit p. Als Ergebnismenge ergibt sich die abzählbar unendliche Menge

$$\Omega = \{1, 2, 3, \ldots\} = \mathbb{N},$$

da theoretisch ein Treffer nie eintreten kann.

Mit der Zufallsvariablen

X: „Anzahl Versuche bis zum ersten Treffer“

und der Trefferwahrscheinlichkeit p lässt sich die Wahrscheinlichkeitsverteilung beschreiben durch

$$P(X = k) = (1 - p)^{k-1} \cdot p, \quad k \in \mathbb{N}.$$

Warten auf den r-ten Treffer

Weiterhin lässt sich bei dem vorherigen Wartezeitproblem auf den r-ten Treffer warten, der im k-ten Versuch auftritt. Somit müssen in den ersten k-1 Teilexperimenten r-1 Treffer aufgetreten sein, damit im k-ten Teilexperiment der r-te Treffer vorkommt. Aus der Kombinatorik ergibt sich für die zugehörige Wahrscheinlichkeitsverteilung

$$P_r(X = k) = \binom{k-1}{r-1}(1 - p)^{k-r} \cdot p^r, \quad k = r,\ r + 1,\ r + 2,\ \ldots$$

Warten auf die vollständige Serie

Für diese Klasse von Wartezeitproblemen ist z. B. von Interesse, wie lange im Mittel gewartet werden muss, bis jedes Element von Ω mindestens einmal aufgetreten ist. Dabei wird der Ergebnisraum beschrieben durch $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, und der zugehörige Zufallsversuch wird solange wiederholt, bis jedes $\omega_i,\ i = 1, \ldots, n$ mindestens einmal vorgekommen ist. Die interessierende Zufallsgröße kann als

X: „Anzahl der Versuchsdurchführungen“

festgelegt werden.

4.2 Simulation verschiedener Zufallsexperimente

Anhand der im letzten Abschnitt klassifizierten Problemstellungen werden nun für jede Art ein oder mehrere Beispiele ausgeführt, um die Unterstützung der Software für die jeweilige Klasse von Fragestellungen zu analysieren und zu bewerten. Dabei wird immer nach einem festgelegten Schema vorgegangen. Zu Beginn steht ein Beispiel, anschließend werden die Möglichkeiten für die Simulation erwähnt, daran anknüpfend erfolgt die Umsetzung der Möglichkeiten und zuletzt steht eine Bewertung der Möglichkeiten.

4.2.1 Simulation einstufiger Zufallsexperimente

Beispiel Als typisches Beispiel eines einstufigen Zufallsexperiments dient der einmalige Würfelwurf. Für diesen wird betrachtet, ob die geworfene Augenzahl größer als Vier ist.

Möglichkeiten Einstufige Zufallsexperimente können ausschließlich simultan simuliert werden.

Umsetzung Der Würfelwurf kann, wie bereits im vorigen Kapitel beschrieben, durch die Bauteile Box, Stapel, Kreisel oder Balken realisiert werden. Dabei spielt die Repräsentation durch ein bestimmtes Bauteil nur für den Nutzer eine Rolle. Der Würfel kann beispielsweise durch das Bauteil Box modelliert werden (Abbildung 96, links), und über die Anzahl der Ziehungen wird die Anzahl der Würfe festgelegt. Diese muss für den einfachen Würfelwurf ‚1' lauten. Die Anzahl der Durchgänge kann beliebig gewählt werden und entspricht der Häufigkeit, mit der das Zufallsexperiment durchgeführt werden soll. Das Merkmal wurde in Abbildung 96 in ‚Augenzahl' umbenannt, um den Bezug zum Würfelwurf deutlich herauszustellen. Zur Auswertung kann das Merkmal Augenzahl in einem Graph dargestellt werden und beispielsweise der Einteiler hilft dabei zu entscheiden, ob eine Zahl größer als Vier gewürfelt wurde.[24]

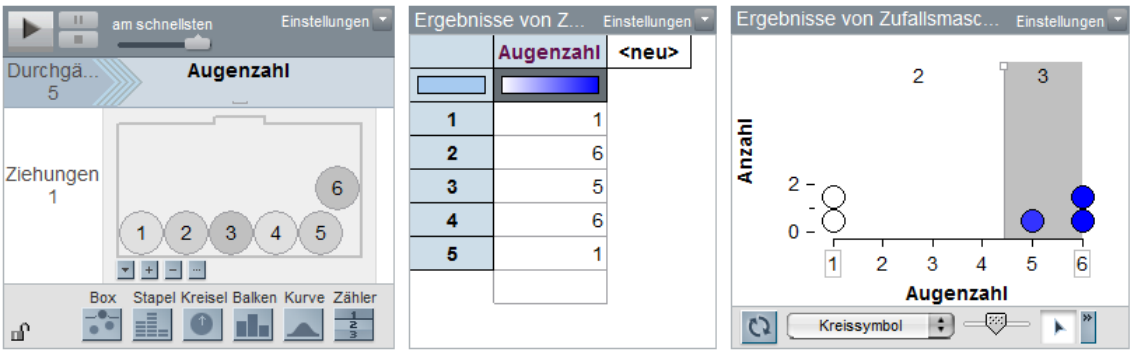


Abbildung 96: Simulation des Würfelwurfs und Definieren des Ereignisses „Augenzahl größer als 4" mit Hilfe des Einteilers

Alternativ kann in der Tabelle ein weiteres Merkmal definiert werden, das prüft, ob die gezogene Augenzahl größer als vier ist. Hierzu kann die Formel *Augenzahl>4* benutzt werden (Abbildung 97).[25]

Unabhängig von der Auswertung ist dies eine simultane Umsetzung des Würfelproblems. Auf die zuerst beschriebene Weise benötigt der Nutzer keine Formel, die Auswertung geschieht im Graph, bei der zweiten Art muss eine gewisse Formelkenntnis vorhanden sein. Eine Unterscheidung in simultan-gebündelt oder simultan-hintereinander (vgl. Kapitel 3.5.2.1) ist bei einem einstufigen Zufallsexperiment nicht gegeben.

Eine sequentielle Simulation ist bei einem einstufigen Zufallsexperiment nicht möglich, da hierfür sowohl für die Anzahl der Ziehungen als auch für die Anzahl der Durchgänge ‚1' gewählt werden müsste. Da Messgrößen immer im Graph ausgewählt werden, der in diesem Fall aber immer nur aus einem Fall, eben der gezogenen Zahl, bestehen würde,

[24] Die Auswertung im Graphen kann auch anders geschehen, beispielsweise durch Klassenbildung.

[25] An dieser Stelle sind weitere Formeln möglich, z. B. *ersterWert(Augenzahl)>4*, *letzterWert(Augenzahl)>4*. Diese liefern bei einem Wert immer dasselbe Ergebnis, bedienen jedoch unterschiedliche Vorstellungen beim Nutzer.

ist eine Messgrößendefinition nicht sinnvoll möglich und somit wird die sequentielle Simulation unmöglich.

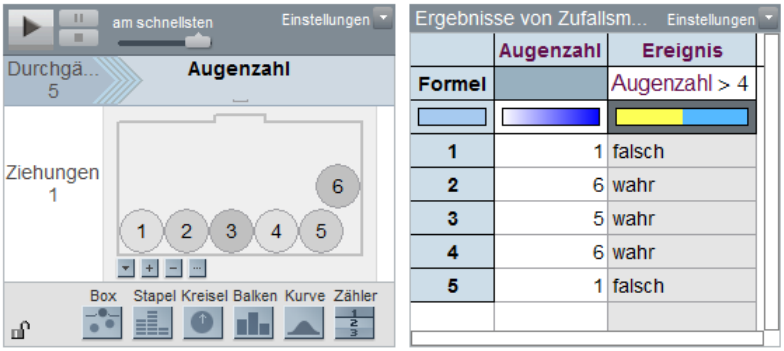


Abbildung 97: Definition des Ereignisses „5 oder 6“ mit Hilfe einer Formel

Die Simulation als Stichprobenziehung ist hier als Simulationsumsetzung ebenfalls möglich, durch die Verwendung des Bauteils Box unterscheidet sie sich jedoch nicht von der oben beschriebenen simultanen Umsetzung. Diese Unterscheidung wird in den weiteren Beispielen nicht mehr getroffen.

Bewertung Für die Simulation einstufiger Zufallsexperimente bietet sich nur die simultane Simulationsart an. Dies ist aus Nutzersicht ein Vorteil, denn die Wahl der Ziehungsanzahl wird intuitiv auf ‚1‘ gesetzt und somit muss über die Simulationsart nicht im Vorfeld nachgedacht werden. Im Modellierungsinterface bleibt die geeignete Wahl eines Bauteils für die Modellierung des Zufallsgeräts dem Nutzer entsprechend seinen Vorlieben überlassen, was ebenfalls als Vorteil gewertet werden kann. Als einziges zu beachten gilt dabei für den Nutzer, dass er ein Bauteil wählt, das die stochastischen Vorgaben des Zufallsexperiments umsetzen kann, beispielsweise bei nicht gleichverteilten Ergebnisräumen das Bauteil Kreisel. Die Angabe der Ziehungen ergibt sich aus dem Zufallsexperiment und beträgt im einstufigen Fall 1, die Wiederholung geschieht über die Anzahl der Durchgänge. Wie oben gezeigt, kann das Interface Ereignisse und Zufallsgrößen im Graphen bedient werden, so dass dort die Definition von Ereignissen und Zufallsgrößen vorgenommen wird. Durch diese Art der Festlegung von interessierenden Ereignissen kann gegebenenfalls auf Formeln verzichtet werden, wodurch erneut dieser Aspekt für den Nutzer einfach realisiert wird und ebenfalls vorteilhaft ist. Oft ist auch eine Definition der interessierenden Ereignisse über Formeln in der Tabelle möglich, dazu ist dann jedoch zwingend Formelkenntnis nötig. Die zusätzliche Spalte in der Tabelle kann für den Nutzer Vor- und Nachteile haben. Als Vorteil lässt sich für ihn sofort für jeden Fall erkennen, ob das entsprechende Ereignis eingetreten ist (im Beispiel ‚wahr‘) oder nicht (im Beispiel ‚falsch‘). Als Nachteil steigt die kognitive Last durch die zusätzliche Spalte, die zusätzlich interpretiert und ausgewertet werden muss. Dies ist bei einstufigen Zufallsexperimenten jedoch zu vernachlässigen. Die entsprechende

Auswertung geschieht dann ebenfalls wieder in einer Graphik. Sofern für die Modellierung eines einstufigen Zufallsexperiments das Bauteil Box gewählt wird, ist die Simulation als Stichprobenziehung gleich der simultanen Simulation. Dem Nutzer bleibt gemäß seiner Vorstellung oder seinen Fähigkeiten überlassen, an welcher Stelle er die interessierenden Ereignisse oder Zufallsgrößen definiert.

4.2.2 Simulation mehrstufiger, stochastisch unabhängiger Zufallsexperimente

Mehrstufige, stochastisch unabhängige Zufallsexperimente wurden in Abschnitt 4.1.2 unterteilt in Zufallsexperimente, die aus gleichartigen Teilexperimenten bestehen und solche, deren Teilexperimente unterschiedlich sind. Maxara hat auf Basis einer Literaturrecherche in Schulbüchern und Aufsätzen zur Stochastik möglichst typische Beispiele identifiziert, anhand derer das Simulationspotential von Fathom untersucht wurde Maxara (2009, 134). Da sich diese Beispiele in ihrer Analyse bewährt haben, werden sie hier soweit möglich und sinnvoll übernommen, um für den geneigten Leser eine Vergleichsmöglichkeit herzustellen. An Stellen, an denen beide Programme voneinander abweichen, werden die Beispiele entsprechend angepasst oder erweitert.

Mehrstufige, stochastisch unabhängige Zufallsexperimente mit gleichartigen Teilexperimenten

Nach Maxara (2009, 135) lassen sich grundsätzlich zwei Arten von Beispielen für diesen Bereich unterscheiden, zum einen Beispiele, in denen die Abfolge der Ergebnisse der Teilexperimente im Fokus des Interesses liegt (beispielsweise, ob beim fünffachen Münzwurf die Folge WWZWW aufgetreten ist), oder Beispiele, bei denen die Auswertung des Gesamtergebnisses von Interesse ist (beispielsweise die Anzahl der Wappen beim fünffachen Münzwurf). Zu beiden Arten wird nun jeweils ein Beispiel vorgestellt und die Simulationsunterstützung von TinkerPlots untersucht.

Zufallsexperimente, deren Teilexperimente gleichartig sind, lassen sich in TinkerPlots durch eine simultane Simulation, durch eine sequentielle Simulation und durch eine Simulation durch Stichprobenziehung realisieren. Solange das Bauteil Box für das Modellierungsinterface verwendet wird, sind die Simulationsarten simultane Simulation und Simulation durch Stichprobenziehung nicht unterscheidbar und aus diesem Grund wird hier die Simulation durch Stichprobenziehung nicht extra aufgearbeitet.

Beispiel Das erste Beispiel nimmt Bezug auf die Abfolge der Ergebnisse der Teilexperimente und steht als ein typisches „Urnenbeispiel“ für eine ganze Klasse an Zufallsexerpimenten. „*In einer Urne liegen zwei rote und drei blaue Kugeln. Aus der Urne werden drei Kugeln mit Zurücklegen gezogen. Es interessiert erstens, ob mindestens zwei blaue Kugeln hintereinander gezogen werden und zweitens ob die Folge „BRB“ gezogen wird.*“ Maxara (2009, 135f). Interessant ist an dieser Stelle nicht nur die Frage,

ob die Folge blau-rot-blau gezogen wurde, sondern auch mit welcher Wahrscheinlichkeit diese gezogen wird, deshalb wird das Beispiel um diese Frage angereichert.

Möglichkeiten Ein Zufallsexperiment dieser Art lässt sich sowohl durch eine simultane, als auch durch eine sequentielle Simulation in TinkerPlots umsetzen.

Umsetzung Im Modellierungsinterface muss zunächst ein Bauteil gewählt werden. Bei dem Ziehen von Kugeln aus einer Urne bietet sich das Bauteil Box an, da es direkt die Vorstellung der Urne repräsentiert. Gemäß der Situation werden der Box fünf Kugeln hinzugefügt, die entsprechend der Farbe beschriftet werden (Abbildung 98). Das dreimalige Ziehen mit Zurücklegen aus der Urne lässt sich nun simultan-gebündelt durch die Anzahl der Ziehungen realisieren, da die drei Ziehungen unabhängig voneinander sind, das heißt mit Zurücklegen. Die Anzahl der Durchgänge hat dabei die Bedeutung der Anzahl der Wiederholungen des gesamten Zufallsexperiments.

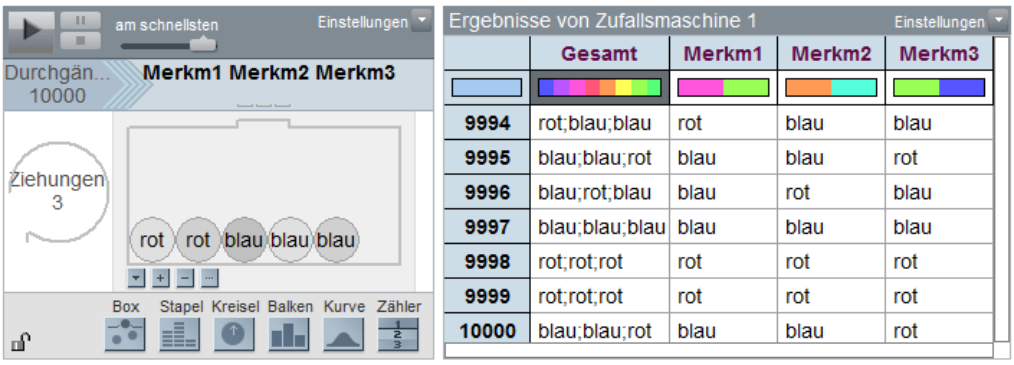


Abbildung 98: Simultane Simulation der Urnenziehung

Ebenfalls ließe sich das Beispiel simultan-hintereinandergeschaltet umsetzen, in diesem Falle hätte man dreimal hintereinander die gleiche Box aus Abbildung 98 in der Zufallsmaschine. Die Vorstellung mit Zurücklegen wird dann intuitiv ersichtlich, allerdings wird dadurch eher die Vorstellung ‚Ziehen einer Kugel aus drei Urnen' gestützt. Das weitere Vorgehen entspricht dem der simultanen-gebündelten Simulation.

Die Ergebnisse werden nach dem Starten der Zufallsmaschine automatisch in einer Tabelle protokolliert (Abbildung 98, rechts). Bei der hier durchgeführten simultanen Simulation werden die Ergebnisse jeden Teilexperiments in einer eigenen Spalte dokumentiert (Merkm1, Merkm2, Merkm3 als automatisch erzeugte Merkmale bei der Ziehungsanzahl 3) und zusätzlich werden sie in der Spalte ‚Gesamt' zusammengefasst in eine Zelle geschrieben. Der Nutzer hat hier also die Übersicht über die Ergebnisse bezogen auf die Teilexperimente und bezogen auf das gesamte Ergebnis.

Wie bereits zuvor bei einstufigen Zufallsexperimenten hat der Nutzer nun wieder die Möglichkeit, die interessierenden Ereignisse entweder im Graphen oder in der Tabelle zu definieren. Bei zwei Ausprägungen und drei Zügen gibt es insgesamt acht verschie-

dene Möglichkeiten für ein Ergebnis. Fasst man in einer Graphik zum Merkmal ‚Gesamt' nun alle interessierenden Ereignisse zusammen (also alle diejenigen, bei denen zweimal hintereinander blau gezogen wurde) und alle nicht interessierenden ebenso, so erhält man eine Darstellung wie in Abbildung 99.

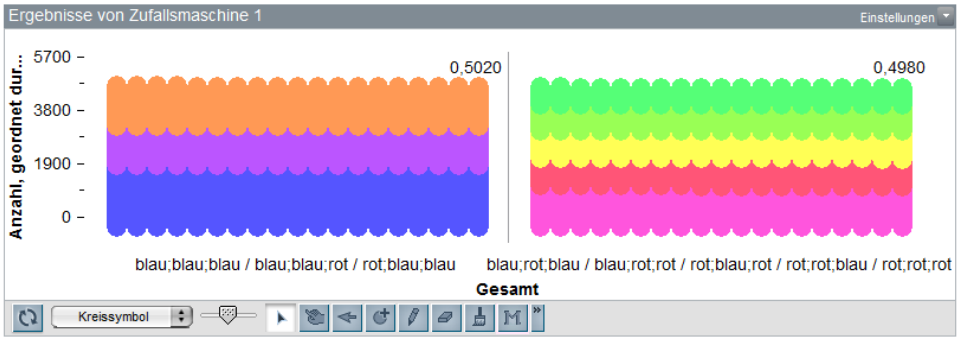


Abbildung 99: Darstellung eines interessierenden Ereignisses („Mindestens zwei blaue Kugeln hintereinander", links) als Anteil aller Ereignisse

Aus der Graphik in Abbildung 99 lässt sich die Wahrscheinlichkeit auf 0,502 schätzen, dass mindestens zwei blaue Kugeln hintereinander gezogen wurden.

Arbeitet der Nutzer lieber mit der Tabelle, so kann das interessierende Ereignis auch mit der Formel *beinhaltet* definiert werden. In Abbildung 100 ist dies in der Spalte ‚Formel' geschehen. Hier wird für jeden Durchgang überprüft, ob die Folge blau;blau direkt hintereinander aufgetreten ist.

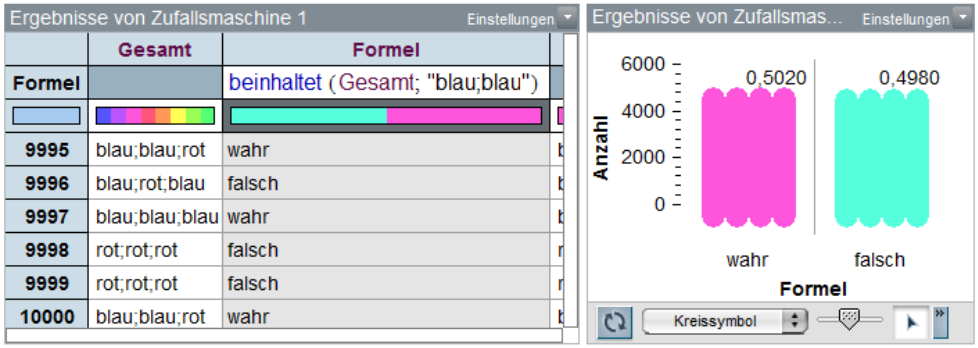


Abbildung 100: Definition eines interessierenden Ereignisses bei der Urnenziehung in der Tabelle der simultanen Simulation

Der rechte Graph in Abbildung 100 zeigt die Auswertung des interessierenden Ereignisses. Wie bereits zuvor lässt sich die Wahrscheinlichkeit auf 0,502 schätzen, dass mindestens zweimal hintereinander eine blaue Kugel gezogen wurde.

Für die Beantwortung der zweiten Frage (ob die Folge blau-rot-blau gezogen wird) kann die Modellierung wie zuvor genutzt werden (Abbildung 98), lediglich das interessierende Ereignis muss neu definiert werden. Wie eben kann dies entweder im Graphen

geschehen, indem das interessierende Ereignis von allen anderen separiert wird (Abbildung 101) oder über die Tabelle bei entsprechender Formelkenntnis durch den Nutzer (Abbildung 102).

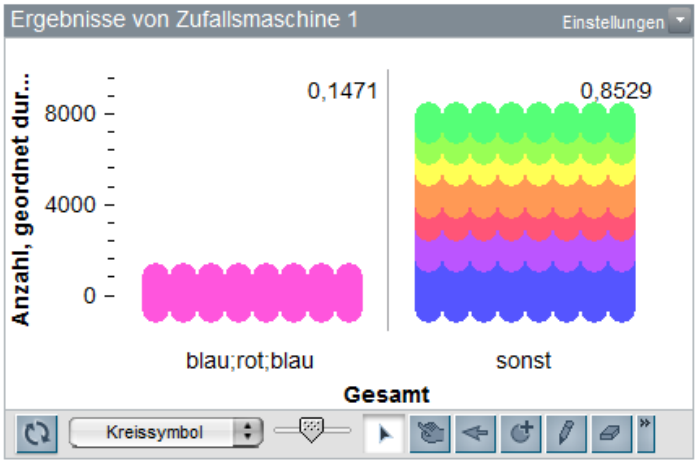


Abbildung 101: Definition des Ereignisses blau-rot-blau im Graphen der simultanen Simulation

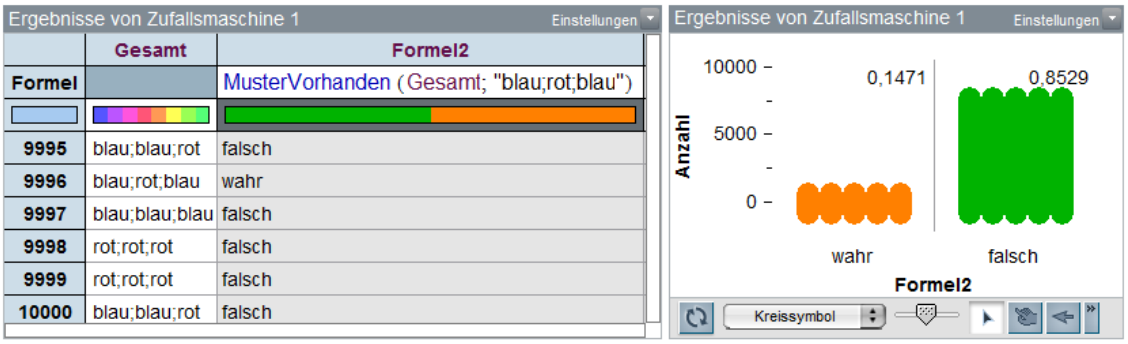


Abbildung 102: Definition des Ereignisses blau-rot-blau in der Tabelle der simultanen Simulation

In Abbildung 102 wurde zur Definition des Ereignisses *Es wird blau-rot-blau gezogen* die Formel *MusterVorhanden(Gesamt; „blau;rot;blau“)* genutzt, die prüft, ob in der Liste Gesamt die Abfolge blau-rot-blau vorliegt. Aus dem Graphen in Abbildung 102 lässt sich erkennen, dass die Abfolge blau-rot-blau aufgetreten ist und ihre Wahrscheinlichkeit lässt sich schätzen auf ca. 14,7%.

Für die Bearbeitung der beiden Fragestellungen bietet der Weg über eine sequentielle Simulation ebenfalls eine mögliche Lösung. Dies erfordert jedoch mehr Aufwand bzw. Wissen bezüglich TinkerPlots. Für die sequentielle Simulation wird das Modellierungsinterface anders genutzt. Angepasst an die Vorstellung der Urne ist die Auswahl des Bauteils Box naheliegend. Hat ein Nutzer die Vorstellung, dass er aus einer Urne zieht und diese Urnenziehung dreimal wiederholt, so kann die gesamte Zufallsmaschine eine Durchführung des Zufallsexperiments abbilden wie in Abbildung 103. Dazu müssen die

Teilexperimente in einzelnen Durchgängen repräsentiert werden, dies entspricht den Einstellungen ‚1' bei Ziehung und ‚3' bei Durchgänge.

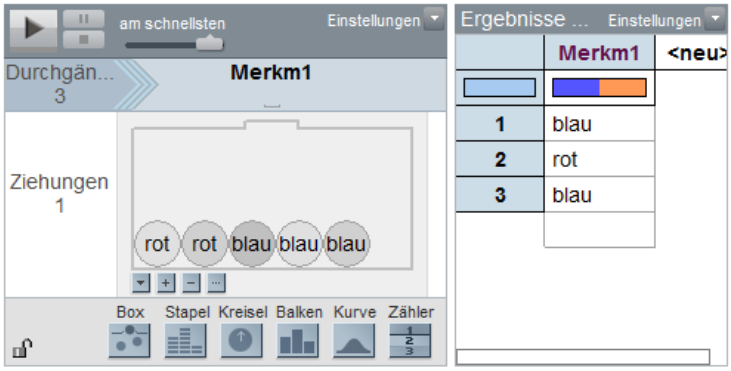


Abbildung 103: Sequentielle Modellierung der Urnenziehung

Die Ergebnisse der Teilexperimente des dreimaligen Ziehens mit Zurücklegen aus einer Urne werden in Abbildung 103 (rechts) in einem Merkmal in den einzelnen Zeilen repräsentiert. Um nun das interessierende Ereignis festzulegen, müssen mehrere Formeln verwendet werden. Hierzu gibt es keine einfache Lösung, so dass dies ein Lösungsweg für einen Experten in TinkerPlots darstellt. Dieser Lösungsweg kann auf gleichartige Situationen übertragen werden. Zuerst kann unter den Ergebnismerkmalen die Runlänge bestimmt werden durch die Auswahl von ‚Runlänge über alle Durchgänge' in Abbildung 104 und anschließend muss geprüft werden, ob die Abfolge zweier blauer Kugeln eine Runlänge 2 verursacht hat.

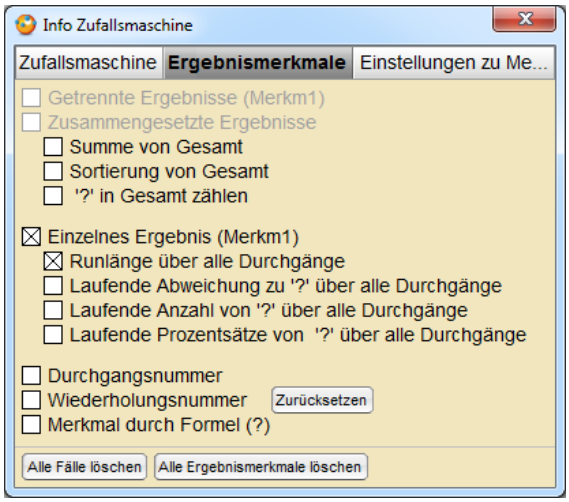


Abbildung 104: Definieren einer Zufallsgröße beim sequentiellen Simulieren der Urnenziehung

Im Hintergrund agiert nun eine Formel[26], die in einer neuen Spalte der Tabelle die Run-

[26] Die genaue Formel lautet: *Runlänge(Merkm1)*, diese muss jedoch vom Nutzer nicht eingegeben werden, wenn er das Ergebnismerkmale-Fenster verwendet.

länge zu Merkmal1 ermittelt. Um zu prüfen, ob zwei blaue Kugeln eine Runlänge von mindestens zwei verursacht haben, also direkt hintereinander gezogen wurden, kann die Darstellung des Ereignisses im Graphen entsprechend Abbildung 105 gewählt werden.

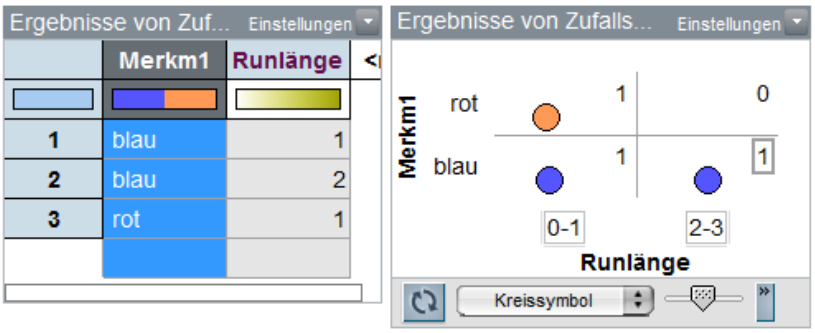


Abbildung 105: Definition des Ereignisses "mindestens zwei blaue Kugeln hintereinander" bei der sequentiellen Simulation

Der in Abbildung 105 grau umrandete Wert ‚1' stellt das Auftreten von mindestens zwei blauen Kugeln hintereinander dar und kann nun als Messgröße gesammelt werden.

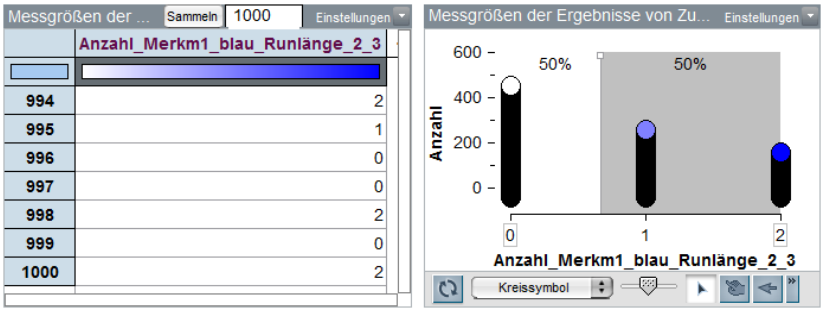


Abbildung 106: Messgrößensammeln und Auswerten bei der sequentiellen Simulation zur Urnenziehung

Wiederum agiert bei dieser Simulationsumsetzung eine Formel im Hintergrund zur Messgrößendefinition. Diese muss vom Nutzer jedoch weder gewusst noch eingegeben werden. Die Auswertung der Ergebnisse geschieht nun wiederum in einem Graph (Abbildung 106, rechts).

Für die zweite Frage, ob die Folge blau-rot-blau gezogen wurde, kann der Weg über eine verschachtelte Formel gewählt werden. Die Modellierung in der Zufallsmaschine bleibt gleich (siehe Abbildung 103). Das interessierende Ereignis kann in der Tabelle mit der Formel

(Merkm1="blau")und(NachfolgeWert(Merkm1)="rot")und(NachfolgeWert(NachfolgeWert(Merkm1))="blau") definiert werden. Hiermit wird überprüft, ob die Folge blau-rot-blau gezogen wurde. Dies geschieht in der Tabelle (Abbildung 107, links) und hier ist die Verstehenshürde für den Nutzer vergleichsweise hoch. Die Ergebnistabelle ist

nicht für die Auswertung von Merkmalen gedacht, somit können sich durchaus Schwierigkeiten beim Verstehen eines Auswertungsmerkmals ergeben. Die korrekten Einträge, sofern die Abfolge blau-rot-blau gezogen wurde, lauten *wahr, falsch, falsch.* Dies liegt daran, dass in der ersten Zeile die Bedingung für die Folgezeilen überprüft wird, und für diese Zeile stimmt hier die Bedingung. In den weiteren Zeilen stimmt die Bedingung nicht mehr, da der Eintrag von Merkmal1 in der zweiten Zeile nicht „blau“ ist und die Nachfolgewerte nicht rot und blau sind. Sobald also ein *wahr* in der ersten Zeile steht, wurde die Abfolge gezogen und von TinkerPlots korrekt erkannt. Dies lässt sich nun in einem Graph darstellen (Abbildung 107, rechts) und die Anzahl von *wahr* als Messgröße sammeln.

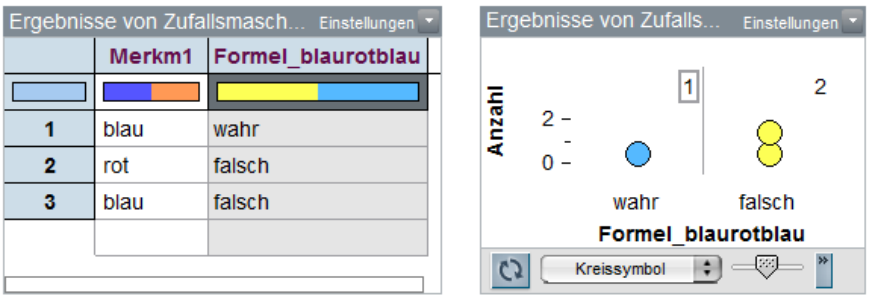


Abbildung 107: Definition des Ereignisses blau-rot-blau in der Tabelle der sequentiellen Simulation

Die gesammelten Messgrößen lassen sich nun wieder über einen Graph auswerten (Abbildung 108). Es lässt sich erkennen, dass in ca. 14% der Fälle (bei 1000 Wiederholungen) die Abfolge blau-rot-blau gezogen wurde.

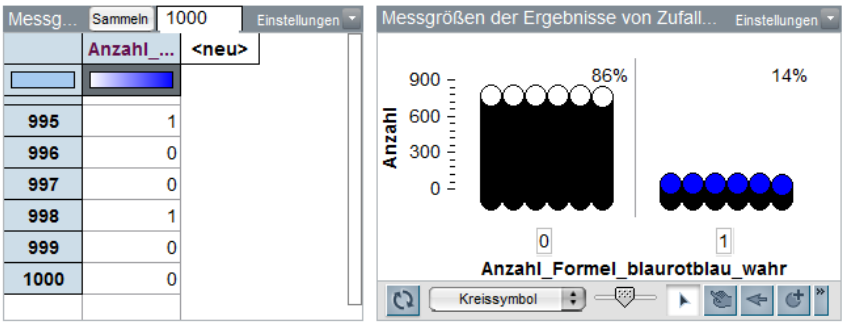


Abbildung 108: Auswertung zur Folge blau-rot-blau in der sequentiellen Simulation

Bewertung Hier wurden zwei mögliche Simulationsalternativen aufgezeigt, um ein mehrstufiges Zufallsexperiment mit unabhängigen gleichartigen Teilexperimenten durchzuführen. Die Simulation dazu kann sowohl simultan als auch sequentiell erfolgen. Wie sind nun die Simulationen zu bewerten und wo liegen Unterschiede in den Simulationsumsetzungen für den Nutzer? Positiv ist zunächst, wie bereits im letzten Kapitel ausführlich dargestellt, dass der Nutzer vorab keine Überlegungen treffen muss, welche Simulationsart er verwenden möchte. Allerdings unterscheiden sich die Simula-

tionen beträchtlich hinsichtlich des Ablaufs. Während die simultane Simulation einfach und direkt umgesetzt werden kann, wird für die sequentielle Simulation viel Hintergrundwissen bezüglich TinkerPlots benötigt und diese ist nicht als intuitiv einzustufen.

Es wurde zunächst für ein stochastisch unabhängiges Zufallsexperiment mit gleichartigen Teilexperimenten die simultan-gebündelte Simulationsart gezeigt. Diese bietet sich an bei gleichartigen Teilexperimenten im Vergleich zur simultan-hintereinandergeschalteten Art, denn die Vorstellungen, dreimal aus einer Urne mit Zurücklegen zu ziehen (entspricht simultan-gebündelt) oder je einmal aus drei gleichartigen Urnen zu ziehen (entspricht simultan-hintereinander) sind bei gleichartigen Teilexperimenten sehr ähnlich. Je mehr Teilexperimente ein Zufallsexperiment hat, umso aufwändiger wird die Simulationsumsetzung der simultan-hintereinandergeschalteten Art, weshalb an dieser Stelle der simultan-gebündelten Art der Vorzug gegeben werden kann. Das Modellierungsinterface lässt sich hier intuitiv gemäß der Vorstellung einer Urne mit fünf Kugeln und dem Ziehen aus dieser bedienen. Für das Interface ‚Ereignisse und Zufallsgrößen' leistet das automatisch erstellte Merkmal ‚Gesamt' einen wichtigen Beitrag, da dadurch viele Umsetzungen von Ereignissen und Zufallsgrößen vereinfacht werden und teilweise sogar durch vordefinierte Ergebnismerkmale nahezu ohne Formeleingabe realisiert werden können (siehe zum Merkmal Gesamt auch ausführlich Kapitel 3.6). Unabhängig davon, ob der Fokus auf dem Gesamtergebnis eines Zufallsexperiments liegt oder auf der Abfolge der zugehörigen Teilexperimente, lässt sich mit Hilfe des Merkmals Gesamt in beiden Fällen gut arbeiten. Sofern ein interessierendes Merkmal in der Tabelle definiert werden soll, muss die entsprechende Formelkenntnis beim Benutzer vorhanden sein. In diesem Fall kann die Auswertung in einem Graph jedoch einfacher sein, da mit nur einem Merkmal im Graph operiert werden muss und häufig nur eine Ausprägung von den anderen separiert werden muss. Die alleinige Darstellung eines interessierenden Ereignisses in einem Graphen kann bei einer großen Anzahl an möglichen Ergebnissen schnell unübersichtlich werden. Man denke an dieser Stelle an eine Urne, in der zum Beispiel acht verschiedenfarbige Kugeln liegen. Somit empfiehlt sich diese Art der Definition in einem Graph eher für Zufallsexperimente mit einer übersichtlichen Anzahl von Ergebnissen. Der Vorteil ist, dass bei diesem Weg das Auswerten gleichzeitig geschieht.

Das Modellierungsinterface wird bei der sequentiellen Simulation ähnlich bedient wie bei der simultanen Simulation. Allerdings wird Bereich 4 (Durchgänge) hier mit zum Modellierungsinterface gezählt im Gegensatz zur simultanen Simulation. Die Betrachtungsweise ist hier eine andere für den Nutzer. Die Ergebnisse einer Durchführung des Zufallsexperiments werden nun durch die Ergebnistabelle abgebildet und die Definition der interessierenden Ereignisse und Zufallsgrößen muss im Graphen geschehen. Dazu kann es notwendig sein, zunächst weitere Auswertungsmerkmale zu erzeugen und diese in die Darstellung im Graphen mit einzubeziehen. Eine gute Formelkenntnis durch den

Nutzer ist hier hilfreich. In der Tat ist die größte Herausforderung bei der sequentiellen Simulationsumsetzung ein interessierendes Ereignis oder eine interessierende Zufallsgröße so im Graph darzustellen, dass von dem entsprechenden Wert Messgrößen zur Wiederholung des Zufallsexperiments gesammelt werden können. Dabei repräsentiert ein Fall im Graph eine gezogene Kugel von drei, also ein Ergebnis eines Teilexperiments. Diese Ergebnisse so anzuordnen, dass sinnvolles Weiterarbeiten, sprich eine Messgrößendefinition, möglich ist, ist für den Nutzer oft eine Schwierigkeit. Es ist etwas einfacher, den Fokus auf das Gesamtergebnis zu legen, denn wenn der Fokus auf einer Abfolge von Teilergebnissen liegt, dann lässt sich dies nur über entsprechende verschachtelte Formeln in der Tabelle realisieren, die zum einen kompliziert einzugeben sind und zum anderen für den Nutzer schwierig zu interpretieren sind. Je mehr Teilexperimente ein Zufallsexperiment besitzt, desto komplizierter wird die weitere Auswertung einer sequentiellen Simulation.

Als Fazit lässt sich festhalten, dass für mehrstufige Zufallsexperimente mit gleichartigen Teilexperimenten die simultane Simulationsart in gebündelter Form bevorzugt benutzt werden sollte. Liegt der Fokus auf dem Gesamtergebnis eines Zufallsexperiments, so können sowohl die simultane als auch die sequentielle Simulationsart gut eingesetzt werden. Liegt der Fokus jedoch auf einer Abfolge der Teilergebnisse eines Zufallsexperiments, so kann dies bei der sequentiellen Simulation schnell zu einer Herausforderung bei dem Finden und der Auswahl einer passenden Formel führen. Hier ist die simultane Simulationsart eindeutig zu empfehlen, da bei dieser Simulationsart weitgehend auf Formeln verzichtet werden kann und sowohl eine kleine Anzahl als auch eine große Anzahl an Teilexperimenten sich hiermit gut realisieren lassen.

Mehrstufige, stochastisch unabhängige Zufallsexperimente mit verschiedenen Teilexperimenten

Beispiel Für diesen Bereich führt Maxara (2009, 143) als Beispiel das Spiel „Die Siedler von Catan“ mit der Erweiterung Städte und Ritter an. In diesem Spiel muss mit drei Würfeln gewürfelt werden, davon sind ein roter und ein weißer Würfel normale sechsseitige Würfel mit Augenzahlen und ein weiterer sechsseitiger Würfel zeigt entweder eine gelbe, eine grüne, eine blaue Festung oder ein schwarzes Seeräuberschiff auf den restlichen drei Seiten. Ein für den Spieler negatives, aber häufig auftretendes Ergebnis ist das folgende: Mit dem roten und dem weißen Würfel wird die Augensumme Sieben gewürfelt und zusätzlich zeigt der dritte Würfel ein Seeräuberschiff. Man interessiert sich für die Wahrscheinlichkeit des Auftretens dieses Ereignisses.

Dieses Beispiel wird als typisch für Würfelexperimente angesehen, in denen mit unterscheidbaren Würfeln experimentiert wird. Mit ihrem Beispiel hat Maxara dies in einen Kontext gesetzt, der durch die Auswahl des „Spiel des Jahres 1995“ und der stetigen Neuauflage dieses Spiels einen schönen Zugang bietet.

Möglichkeiten Beispiele dieser Art lassen sich nur simultan simulieren.

Umsetzung Für die simultane Simulation kann nur die hintereinandergeschaltete Simulationsart verwendet werden, weil nur durch diese die verschiedenen Teilexperimente modelliert werden können. Für die Würfelmodellierung ist in Abbildung 109 jeweils das Bauteil Kreisel verwendet worden, weil bei diesem die Beschriftung des dritten Würfels am besten lesbar ist. Zunächst wurde für die einzelnen Würfel jeweils das Bauteil Box gewählt und die Kugeln entsprechend beschriftet. Durch das Ändern des Bauteils wurden beim dritten Würfel automatisch gleiche Ausprägungen zusammengeführt, so dass für den Nutzer durch den Bauteilwechsel direkt erkennbar ist, dass die Wahrscheinlichkeit für das Würfelergebnis „Seeräuberschiff“ 0,5 beträgt. Die Ziehungsanzahl steht der Situation entsprechend auf Drei, da jeder Würfel pro Durchgang einmal gewürfelt wird.

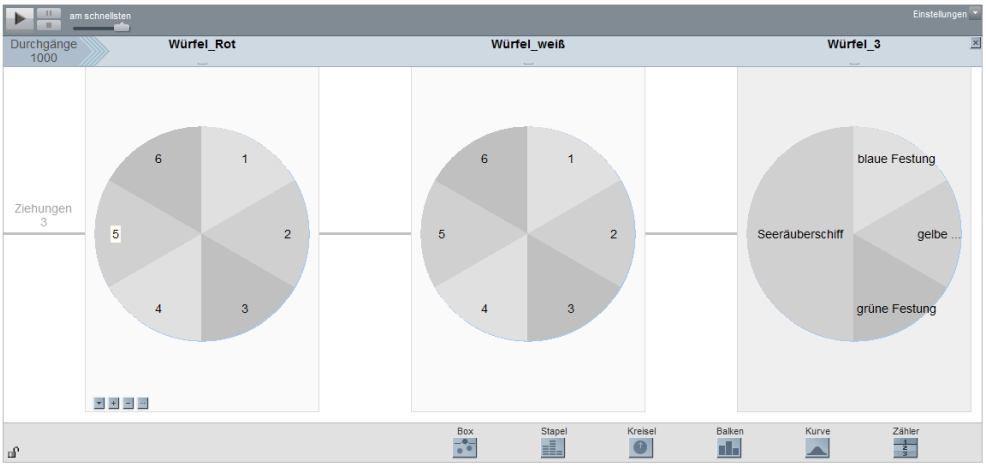


Abbildung 109: Modellierung der Siedler-Würfel in der simultanen Simulation

In der Tabelle lassen sich nun weitere Merkmale definieren, um das Ereignis zu bestimmen.

In Abbildung 110 ist zusätzlich die Augensumme der ersten beiden Würfel als Ergebnismerkmal definiert. Dies kann nicht per vordefiniertem Ergebnismerkmal erfolgen, sondern muss von Hand durch die Formel *Würfel_Rot+Würfel_weiß* geschehen.

Das vordefinierte Merkmal ‚Summe‘ kann ausschließlich auf dem Merkmal Gesamt operieren und durch das Vorliegen von nicht numerischen Werten in der gegebenen Situation lässt sich das Merkmal ‚Summe‘ hier nicht anwenden. Aus diesem Grund ist für die beiden ersten Würfel die Zufallsgröße *Augensumme* per Formel definiert.

Ergebnisse von Zufallsmaschine 1 Einstellungen

	Gesamt	Würfel_Rot	Würfel_weiß	Würfel_3	Augensumme
Formel					Würfel_Rot + Würfel_weiß
994	1;2;gelbe Festung	1	2	gelbe Festung	3
995	2;3;grüne Festung	2	3	grüne Festung	5
996	4;4;grüne Festung	4	4	grüne Festung	8
997	6;6;Seeräuberschiff	6	6	Seeräuberschiff	12
998	1;3;gelbe Festung	1	3	gelbe Festung	4
999	5;5;Seeräuberschiff	5	5	Seeräuberschiff	10
1000	4;6;Seeräuberschiff	4	6	Seeräuberschiff	10

Abbildung 110: Definition der Zufallsgröße Augensumme der ersten beiden Siedler-Würfel

Das Merkmal *Augensumme* kann nun in einem Graphen wie in Abbildung 111 zusammen mit *Würfel_3* dargestellt werden, so dass sich das interessierende Ereignis erkennen lässt (Abbildung 111, graue Markierung).

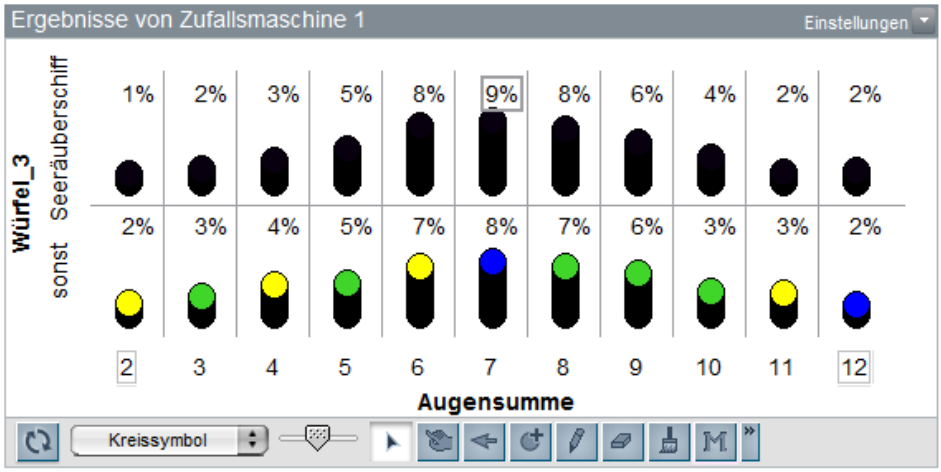


Abbildung 111: Darstellung des interessierenden Ereignisses im Graph zur simultanen Simulation der Siedler-Würfel

Alternativ dazu lässt sich das Ereignis „die ersten beiden Würfel haben als Augensumme 7 und der dritte Würfel zeigt ein Seeräuberschiff“ auch direkt durch die Formel *(Würfel_Rot+Würfel_weiß=7)und(Würfel_3=“Seeräuberschiff“)* umsetzen wie in Abbildung 112 geschehen und in Abbildung 113 ausgewertet.

Durch die Verschiedenheit der Teilexperimente ist in dieser Situation das Merkmal Gesamt keine zusätzliche Hilfe.

Natürlich ließe sich in einem Graphen das Merkmal Gesamt darstellen und es könnten alle interessierenden Ergebnisse von den restlichen getrennt werden wie in Abbildung 114 (also die Ergebnisse ‚1-6-Seeräuberschiff‘; ‚6-1-Seeräuberschiff‘; ‚2-5-

Seeräuberschiff'; usw.), aber das wäre doch ein erheblicher Aufwand für den Nutzer und ist somit auch fehleranfälliger.

Ergebnisse von Zufallsmaschine 1 — Einstellungen

	Gesamt	Würfel_Rot	Würfel_weiß	Würfel_3	Augensumme_7_und_Seeräuberschiff
Formel					(Würfel_Rot + Würfel_weiß = 7) und (Würfel_3 = "Seeräuberschiff")
250	4;3;Seeräuberschiff	4	3	Seeräuberschiff	wahr
251	3;5;gelbe Festung	3	5	gelbe Festung	falsch
252	1;1;Seeräuberschiff	1	1	Seeräuberschiff	falsch
253	3;4;Seeräuberschiff	3	4	Seeräuberschiff	wahr
254	6;5;Seeräuberschiff	6	5	Seeräuberschiff	falsch
255	4;4;blaue Festung	4	4	blaue Festung	falsch
256	4;1;grüne Festung	4	1	grüne Festung	falsch

Abbildung 112: Definieren des Merkmals "Ereignis" in der Tabelle zu den Siedler-Würfeln

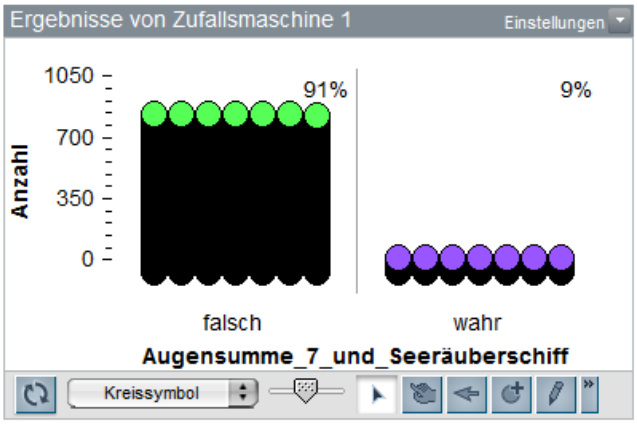


Abbildung 113: Auswerten des Ereignisses zu den Siedler-Würfeln

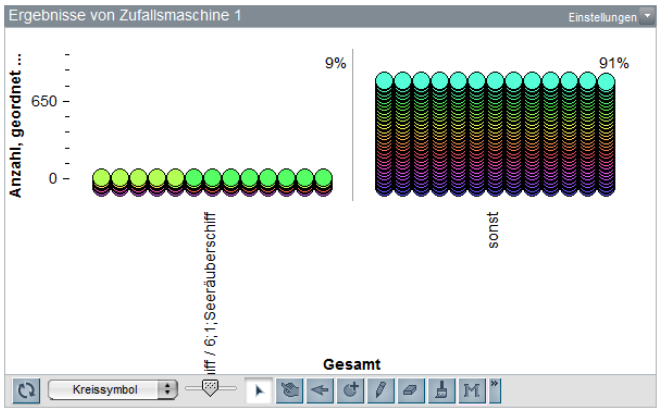


Abbildung 114: Darstellung des Ereignisses nur durch das Merkmal Gesamt der Siedler-Würfel

Eine sequentielle Simulation kann für Zufallsexperimente mit verschiedenen Teilexperimenten nicht durchgeführt werden, weil die zugehörige Versuchsprotokollierung in Form der Tabelle auf einen Fall beschränkt wäre und sich davon keine unterscheidbaren Messgrößen sammeln ließen. Die verschiedenen Teilexperimente lassen sich nur hintereinandergeschaltet realisieren, somit werden die Teilergebnisse immer in verschiedenen

Merkmalen, also verschiedenen Spalten der Tabelle, realisiert, eine sequentielle Simulation ist also unmöglich.

Bewertung Mehrstufige, stochastisch unabhängige Zufallsexperimente mit unterschiedlichen Teilexperimenten lassen sich nur simultan in TinkerPlots realisieren. Das ist für den Nutzer zunächst keine Einschränkung, da dies der intuitiven Umsetzung solcher Zufallsexperimente entspricht. Die verschiedenen Teilexperimente in verschiedenen Bauteilen zu realisieren ist naheliegend bei mehrstufigen Zufallsexperimenten und nicht anders möglich in TinkerPlots. Eine solche Darstellung von mehreren Würfeln über verschiedene Kreisel wird von Konold und Kazak (2008) als positiv für Lernende bezüglich der Einfachheit einer Simulation herausgestellt. Allerdings können bei verschiedenen Teilexperimenten die vordefinierten Auswertungsmerkmale nicht verwendet werden, da sie alle auf sämtlichen Teilergebnissen des Merkmals ‚Gesamt' operieren. Die Vereinfachung durch das Merkmal ‚Gesamt' fällt also für das Interface Ereignisse und Zufallsgrößen weg, außer es wird in einem Graphen dazu verwendet, alle interessierenden Ergebnisse von allen anderen zu trennen. Dies kann je nach Anzahl der möglichen Ergebnisse und je nach Anzahl der Teilexperimente jedoch sehr aufwändig für den Benutzer sein und erfordert viel Eigenarbeit im Bestimmen der interessierenden Ergebnisse.

Einen sprechenden Namen für die verschiedenen Merkmale bzw. Teilexperimente zu wählen, kann die Simulation wesentlich vereinfachen. Da zur Definition der interessierenden Ereignisse und Zufallsgrößen Formeln verwendet werden müssen, ist das Zugreifen auf das benötigte Merkmal bei einem sprechenden Namen leichter. Hier sollte der Benutzer also die Möglichkeit zur Merkmalsumbenennung nutzen, auch wenn die automatisch erzeugten Merkmalsnamen zunächst praktisch erscheinen. Wie oben gezeigt lassen sich auch hier die gesuchten Wahrscheinlichkeiten auf unterschiedlichen Wegen, das heißt mit unterschiedlichen Formeln, bestimmen. Durch die *und*-Verknüpfung bei Formeln lässt sich mitunter wie oben das interessierende Ereignis durch eine lange Formel bestimmen. Für die Auswertung muss der Nutzer dann nur dieses eine Merkmal verwenden. Das kann im Auswertungsinterface mitunter leichter sein, als zwei oder in entsprechenden Situationen sogar noch mehr Auswertungsmerkmale miteinander in einem Graph darzustellen und darüber das gesuchte Ereignis zu bestimmen.

Schließlich bleibt immer noch die Möglichkeit für den Nutzer, alle zum interessierenden Ereignis zugehörigen Ergebnisse selbst zu bestimmen und für die Auswertung allein das Merkmal Gesamt zu verwenden. Dann müssen bei der Simulation keine Formeln verwendet werden, was durchaus zur Vereinfachung auf technischer Ebene führt. Auf der anderen Seite müssen dann jedoch alle benötigten Ergebnisse kombinatorisch durch den Nutzer bestimmt werden, was seine ganz eigenen Schwierigkeiten, aber auch Vorteile

birgt, wenn man zum Beispiel an junge Nutzer denkt, die wenig Erfahrung mit Kombinatorik haben.

4.2.3 Simulation mehrstufiger, stochastisch abhängiger Zufallsexperimente

Zu diesem Typ von Zufallsexperimenten gehören drei unterschiedliche Szenarien, die in Abschnitt 4.1.2 beschrieben wurden. Es lassen sich Zufallsexperimente unterscheiden, bei denen die Teilexperimente durch Urnenziehungen ohne Zurücklegen modelliert werden können, des weiteren gibt es Zufallsexperimente, bei denen die Teilexperimente voneinander abhängig sind, aber nicht durch eine Urne ohne Zurücklegen modelliert werden können.

Mehrfache Urnenziehung ohne Zurücklegen simulieren

Beispiel Zur Analyse wird das von Maxara als repräsentativ gekennzeichnete Beispiel des Ziehens von Buchstaben aus einer Urne verwendet. „Betrachten wir als Beispiel eine Urne, die mit den Buchstaben „E“, „E“, „A“, „A“, „V“ und „V“ gefüllt ist. Aus dieser Urne sollen nun drei Buchstaben ohne Zurücklegen gezogen werden. Wie groß ist die Wahrscheinlichkeit, dass der Name EVA gezogen wird“? (Maxara 2009, 148).

Möglichkeiten Zufallsexperimente, die sich durch mehrfaches Ziehen ohne Zurücklegen aus einer Urne modellieren lassen, können in TinkerPlots sowohl simultan als auch sequentiell umgesetzt werden. Dabei lässt sich für die sequentielle Simulation noch eine kreative Nutzung der Bauteile der Zufallsmaschine unterscheiden, sofern eine bestimmte Abfolge von Teilergebnissen das interessierende Ereignis darstellt. Dies wird ausführlich weiter unten dargelegt.

Umsetzung Durch eine simultane Simulation lässt sich diese Situation leicht modellieren. Entsprechend der intuitiven Nutzervorstellung lässt sich eine Urne bzw. in TinkerPlots eine Box mit den Kugeln entsprechend der Situation bestücken (Abbildung 115). Zusätzlich ist in Abbildung 115 die Ziehungsanzahl drei eingestellt und das Ziehen ohne Zurücklegen angegeben, um die Situation korrekt zu modellieren.

Abbildung 115: Modellierung zur "EVA" als simultane Simulation

Für das Interface Ereignisse und Zufallsgrößen ergeben sich nun wieder zwei Möglichkeiten. Erstens kann das interessierende Ereignis mit Hilfe einer Formel als Auswertungsmerkmal definiert werden. Dies kann durch die Formel *MusterVorhanden(Gesamt; „E;V;A“)* geschehen, die wiederum auf dem Merkmal ‚Gesamt‘ operiert (Abbildung 116, Spalte Formel).

Ergebnisse von Zufallsmaschine 1 Einstellungen

	Gesamt	Formel	Merkm1	Merkm2	Merkm3
Formel		MusterVorhanden (Gesamt; "E;V;A")			
995	V;E;A	falsch	V	E	A
996	V;E;A	falsch	V	E	A
997	V;E;E	falsch	V	E	E
998	V;A;A	falsch	V	A	A
999	E;V;V	falsch	E	V	V
1000	V;A;V	falsch	V	A	V

Abbildung 116: Definieren des Auswertungsmerkmals für das Beispiel EVA der simultanen Simulation

Hiermit wird überprüft, ob die Buchstabenfolge E; V; A mit dem Standardtrennzeichen Semikolon gezogen wurde. Diese Formel existiert nicht als vordefinierte Formel, gehört jedoch zu den Listenbefehlen, die die Definition von Ereignissen per Formel wesentlich vereinfachen.

Die Auswertung ist dann ebenfalls nicht schwer, in einer geeigneten Darstellung wie zum Beispiel in Abbildung 117 kann geschätzt werden, dass die Wahrscheinlichkeit, dass beim dreimaligen Ziehen ohne Zurücklegen aus der Urne die Buchstabenfolge EVA gezogen wird, bei sieben Prozent liegt.

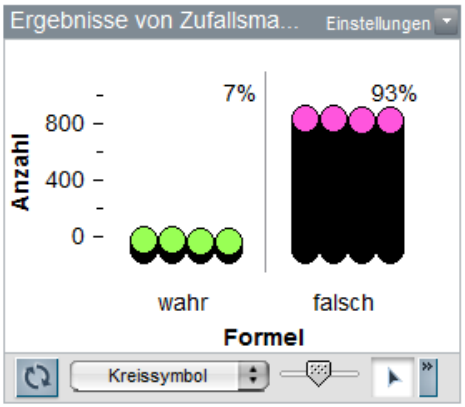


Abbildung 117: Auswertung des interessierenden Ereignisses zu EVA per Formel in der simultanen Simulation

Eine alternative Definition des interessierenden Ereignisses ist über das Merkmal ‚Gesamt‘ in einem Graph wie in Abbildung 118 möglich.

In Abbildung 118 ist das interessierende Ereignis „E;V;A“ als gezogene Buchstabenfolge von den anderen Ergebnissen getrennt und somit ohne Formeleingabe definiert. Die Auswertung geschieht dann in der gleichen Darstellung durch Einblendung der Prozente.

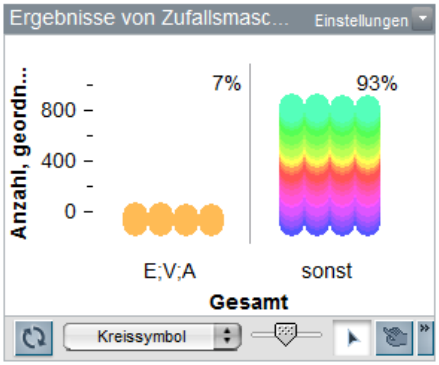


Abbildung 118: Definieren des interessierenden Ereignisses im Graph für EVA der simultanen Simulation

Dieses Beispiel lässt sich auch sequentiell realisieren. Dazu sieht die Modellierung in der Zufallsmaschine ähnlich aus wie für die simultane Simulation (Abbildung 115), der Unterschied besteht in der Wahl der Ziehungs- und Durchgangsanzahl. Die Ziehungsanzahl muss ‚1‘ betragen und für die Anzahl der Durchgänge muss ‚3‘ gewählt werden, um eine sequentielle Simulation durchzuführen wie in Abbildung 119.

Auch bei dieser Modellierung muss darauf geachtet werden, dass ohne Zurücklegen eingestellt wird, um die Situation korrekt abzubilden. Hierzu kann die Einstellung ‚ohne Zurücklegen‘ des Pfeilmenüs unterhalb der Box benutzt werden. Nun werden die Teilziehungen in Zeilen repräsentiert.

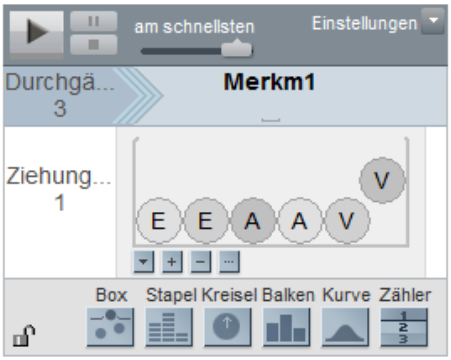


Abbildung 119: Sequentielle Modellierung für das Beispiel EVA

Um das interessierende Ereignis (die Zugfolge EVA) zu definieren, wird es in der Tabelle über eine zusammengesetzte Formel definiert, die die Abfolge der gezogenen Buchstaben prüft (Abbildung 120).

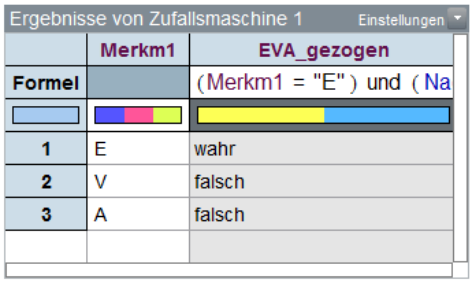


Abbildung 120: Definition des interessierenden Ereignisses der sequentiellen Simulation in der Tabelle

Mit der Formel

(Merkm1="E") und (Nachfolgewert(Merkm1)="V") und (Nachfolgewert(Nachfolgewert(Merkm1)="A"),

die in Abbildung 120 unter dem Merkmal ‚EVA_gezogen' steht, wird überprüft, ob EVA in der angegebenen Reihenfolge gezogen wurde. Dabei wird ausgehend von jeder Zeile die Formel überprüft, so dass der Wert in der ersten Zeile ausschlaggebend ist. Steht hier ein „wahr", so wurde EVA gezogen, steht hier ein „falsch", so wurde eine andere Buchstabenfolge gezogen. Für die zweite und dritte Zeile wird dies erneut überprüft, da hier aber zuerst ein „V" bzw. „A" gezogen wurde, wird „falsch" ausgegeben. Mithilfe des Messgrößensammelns kann das Merkmal Eva_gezogen ausgewertet werden. Dazu wird die Anzahl von „wahr" gesammelt (Abbildung 121, graue Markierung) und anschließend ausgewertet (Abbildung 122).

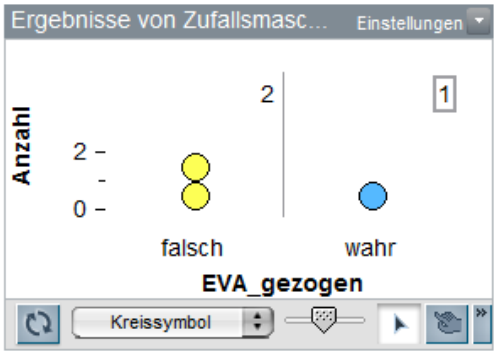


Abbildung 121: Darstellung des interessierenden Ereignisses mit Messgrößendefinition der sequentiellen Simulation

Die Formel, die für diese Situation benötigt wird, ist durch die Verschachtelung in TinkerPlots eher unüblich, aber eine gute Möglichkeit, die Simulation fortzuführen, wenn sie sequentiell begonnen wurde. Die Auswertung über Messgrößen ist dann wieder einfach und durch wenig Arbeit zu erledigen.

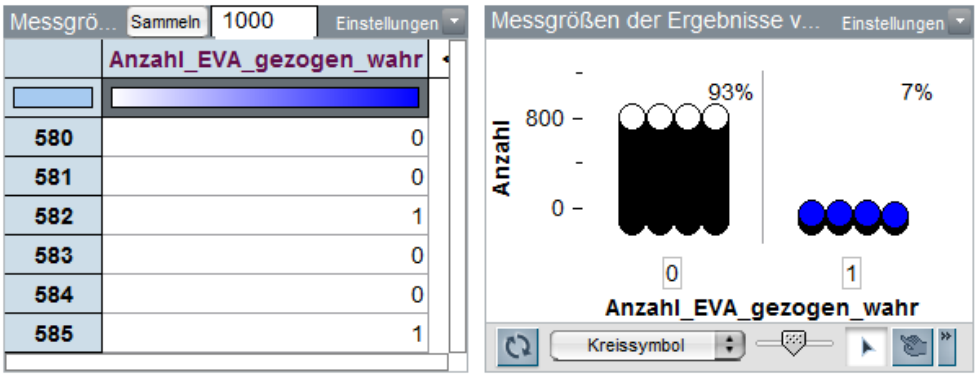


Abbildung 122: Gesammelte Messgrößen zu EVA und Auswertung

Eine weitere Möglichkeit diese Situation sequentiell zu modellieren ist eine kreative Verwendung von zwei Bauteilen der Zufallsmaschine wie in Abbildung 123.

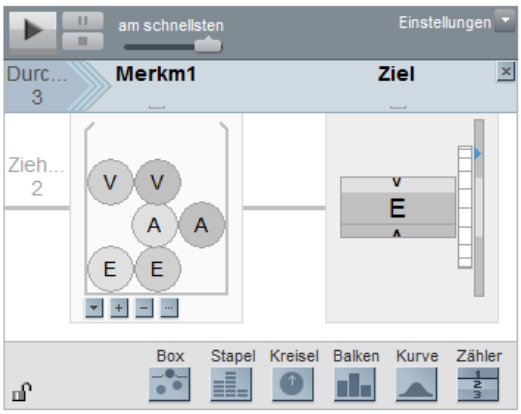


Abbildung 123: Sequentielle Modellierung zu EVA durch zwei Bauteile

In Abbildung 123 ist im ersten Bauteil Box die Urne abgebildet, die die sechs Buchstaben des Beispiels enthält und für das zusätzlich ohne Zurücklegen eingestellt wurde. Im zweiten Bauteil wird das Bauteil Zähler verwendet, das die drei Einträge E, V und A in genau dieser Reihenfolge hat und eine Zielliste geordneter Zeichen festlegt, mit dem (in einem weiteren Merkmal der Tabelle, s. u.) die gezogenen Buchstaben aus Merkmal1 überprüft werden können. Da es sich beim Bauteil Zähler nicht um ein Zufallsgerät handelt, sondern um eine geordnete Liste von Zeichen, ist dieses Bauteil geeignet, um eine bestimmte (Ziel-) Zeichenfolge festzulegen: In der Modellierung von Abbildung 123 wird für das Merkmal Ziel in der ersten Ziehung immer ein E, in der zweiten Ziehung immer ein V und in der dritten Ziehung immer ein A ausgegeben.

In der Tabelle kann in einem dritten Merkmal ‚Eva_getroffen' überprüft werden, ob die zufällig gezogenen Buchstaben aus Merkmal1 mit der Zielvorgabe des Merkmals ‚Ziel' übereinstimmen (Abbildung 124).

Ergebnisse von Zufallsmaschine 1 Einstellungen

	Gesamt	Merkm1	Ziel	Eva_getroffen
Formel				Merkm1 = Ziel
1	A;E	A	E	falsch
2	V;V	V	V	wahr
3	A;A	A	A	wahr

Abbildung 124: Definition des interessierenden Ereignisses in der Tabelle der zweiten sequentiellen Simulation zu EVA

Hierbei ist das Merkmal Gesamt nicht von Interesse. Erst wenn im Merkmal Eva_getroffen dreimal ein „wahr“ auftaucht, wurde die richtige Buchstabenfolge in ‚Merkmal1‘ zufällig gezogen. Dies lässt sich in einem Graphen darstellen (Abbildung 125) und die Anzahl „wahr“ als Messgrößen sammeln.

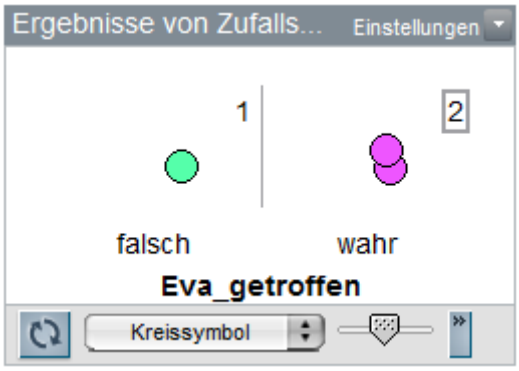


Abbildung 125: Definition der zu sammelnden Messgröße zu ‚Eva_getroffen‘

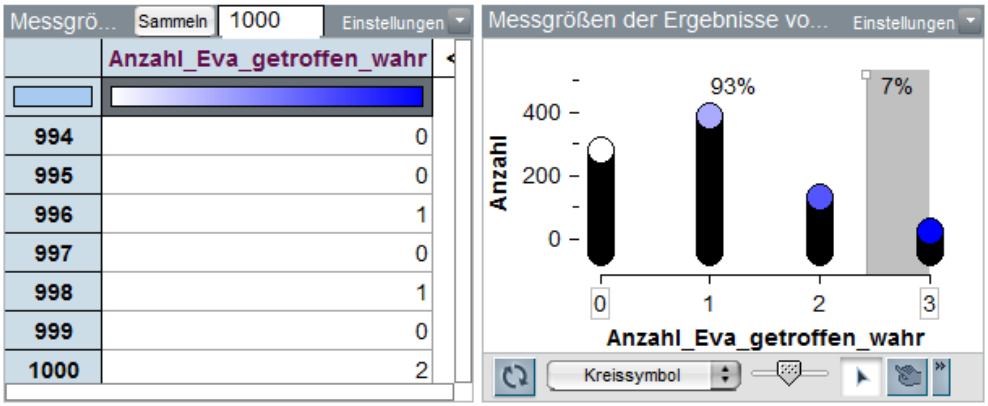


Abbildung 126: Messgrößentabelle und -auswertung zur zweiten sequentiellen Simulation von EVA

In Abbildung 126 sind die gesammelten Messgrößen in der Tabelle (links) zu sehen und rechts befindet sich die Auswertung. Die grau hinterlegten Fälle sind die, bei denen alle

drei zufällig gezogenen Buchstaben des Merkmals1 mit denen des Merkmals ‚Ziel' übereinstimmten, so dass die gesuchte Wahrscheinlichkeit wieder auf sieben Prozent geschätzt werden kann.

Bewertung Das Modellierungsinterface lässt sich sowohl für die simultane als auch für die sequentielle Simulation leicht der Situation und der Nutzervorstellung anpassen. Durch das Bauteil Box, das eine Repräsentation der Urne im Objekt der Zufallsmaschine darstellt, ist jede Modellierung einer Urnenziehung eng an der Vorstellung des Nutzers zur jeweiligen Situation gegeben. Für den Nutzer ist jedoch wichtig, darauf zu achten, dass ohne Zurücklegen gezogen wird. Dies muss in einem Untermenü der Zufallsmaschine eingestellt oder überprüft werden. Hierin liegt ein Hindernis in der Benutzung, da es leicht zu einer falschen Einstellung, die nicht erkannt wird, kommen kann.

Interessant ist bei der Modellierung die Wahl der Ziehungs- und der Durchgangsanzahl, da diese die Simulationsart und somit das Interface zu Ereignissen und Zufallsgrößen bestimmen. Bei der simultanen Simulationsart lassen sich interessierende Ereignisse entweder mit Hilfe von Formeln in der Tabelle bestimmen oder durch Separieren der zugehörigen Ergebnisse des Merkmals ‚Gesamt' im Graphen. Die Formeleingabe wird durch das Merkmal ‚Gesamt' vereinfacht, für das in TinkerPlots eigene Formeln existieren, die auf der durch ‚Gesamt' erzeugten Liste operieren können. Dadurch sind vielfältige Definitionen weiterer Auswertungsmerkmale möglich (vgl. Kapitel 3.6). Dabei ist es unerheblich, aus wie vielen Teilexperimenten das Zufallsexperiment besteht, da über die Listenbefehle gleichzeitig mit einer Formel auf alle Ergebnisse von Teilexperimenten zugegriffen werden kann. Die Darstellung des Merkmals ‚Gesamt' in einem Graphen ist ein Weg zur Definition von Ereignissen, der prinzipiell immer möglich ist. Durch die Operation Trennen können zu dem interessierenden Ereignis gehörige Ergebnisse separiert werden. Sind durch das Zufallsexperiment jedoch viele Ausgänge möglich, so kann es dem Benutzer leicht passieren, dass er den Überblick verliert. Gegebenenfalls sollte der Benutzer über kombinatorisches Wissen verfügen, um alle zugehörigen Ergebnisse identifizieren und separieren zu können. Beide Wege zur Definition des interessierenden Ereignisses sind recht universell einsetzbar. Die Formel *MusterVorhanden* kann für eine nahezu beliebige zu überprüfende Zeichenfolge eingesetzt werden und somit eine Vielzahl von stochastischen Fragestellungen abdecken. Die Definition über das Merkmal Gesamt im Graphen ist auch nahezu immer möglich, aber bei einer Vielzahl von möglichen Ergebnissen bei unterschiedlich großen Wahrscheinlichkeitsräumen und abhängig von der Anzahl der Teilexperimente eventuell sehr aufwändig.

Bei der sequentiellen Simulation ist die Definition von Ereignissen und Zufallsgrößen bei mehrstufigen, abhängigen Zufallsexperimenten oft etwas komplizierter als bei der simultanen Simulation. Wird herkömmlich sequentiell simuliert (damit ist gemeint, dass ein Bauteil verwendet wird, die Ziehungsanzahl Eins eingestellt wird und die Durch-

gangsanzahl der Anzahl der Teilexperimente entspricht), so kann das Bestimmen eines interessierenden Ereignisses durch eine Formel komplexer werden, je mehr Teilexperimente das Zufallsexperiment umfasst. Wird jedoch sequentiell simuliert, indem als weiteres Bauteil der Zähler verwendet wird, um ein Ziel zu definieren, so ist der anschließende Abgleich per Formel in der Tabelle recht einfach. Diese Art zu simulieren funktioniert jedoch nur bei ausgewählten Beispielen, bei denen eine Art Zielmerkmal über den Zähler definiert werden kann, also bei Zufallsexperimenten, bei denen nach einer Abfolge von Zeichen in einer bestimmten Reihenfolge gefragt ist.

Das Auswertungsinterface über den Graphen ist bei allen Simulationsumsetzungen ähnlich, besondere Hindernisse sind hier nicht auszumachen.

Zufallsexperimente mit verschiedenen, abhängigen Teilexperimenten simulieren

Zu dieser Gruppe von Zufallsexperimenten zählen diejenigen, die sich nicht durch ein einziges Urnenmodell beschreiben lassen, sondern die aus verschiedenartigen, abhängigen Teilexperimenten bestehen. Maxara (2009, S. 158ff) unterscheidet drei typische Arten dieser Zufallsexperimente, zunächst Aufgabenstellungen mit bedingten Wahrscheinlichkeiten, weiterhin einige Glücksspielsituationen und zuletzt die Simulation von Markowketten.

Diese Art von Zufallsexperimenten lässt sich nur simultan simulieren, Markowketten können mit der Zufallsmaschine von TinkerPlots nicht simuliert werden.

Simulieren zu Aufgabenstellungen mit bedingten Wahrscheinlichkeiten

Beispiel Als Beispiel für eine Aufgabenstellung mit bedingten Wahrscheinlichkeiten führt Maxara (2009, 159) das Beispiel der Brustkrebsuntersuchung an, zitiert nach Wassner (2004, 48) „Bei Frauen ab 40 Jahren werden Routineuntersuchungen auf Brustkrebs durchgeführt. Das Untersuchungsverfahren ist die sogenannte „Mammografie“. Aus der Literatur ist folgendes bekannt: Die Wahrscheinlichkeit, dass eine Frau der Altersgruppe zwischen 40 und 50 Jahren Brustkrebs (B) hat, beträgt 0,8%. Die Wahrscheinlichkeit, dass diese Krankheit mit einer Mammografie erkannt wird (M+), wenn sie vorliegt, beträgt 91,5%. Die Wahrscheinlichkeit, dass eine Mammografie fälschlicherweise auf Brustkrebs hinweist, obwohl die Krankheit gar nicht vorliegt, beträgt 2,6%. Wie groß ist die Wahrscheinlichkeit, dass eine Frau dieser Altersgruppe tatsächlich Brustkrebs hat, wenn sie einen positiven Mammografiebefund erhalten hat?“

Eine solche Aufgabe steht prototypisch für eine ganze Klasse von Aufgaben zu bedingten Wahrscheinlichkeiten.

Möglichkeiten Aufgabenstellungen, denen bedingte Wahrscheinlichkeiten innewohnen, lassen sich in TinkerPlots nur simultan simulieren.

Umsetzung Eine Lösestrategie hierzu liegt in der Erstellung eines Baumdiagramms, das sich auch mit der Zufallsmaschine von TinkerPlots realisieren lässt. In Abbildung 127 geschieht die Simulation simultan hintereinandergeschaltet. Um die einzelnen Wahrscheinlichkeiten zu repräsentieren, wurde in Abbildung 127 als Bauteil jeweils der Kreisel gewählt und die Anzeige auf Anteile gestellt. Durch die zwei Bauteilspalten steht die Ziehungsanzahl automatisch auf zwei, die Merkmalsnamen wurden der Situation entsprechend umbenannt.

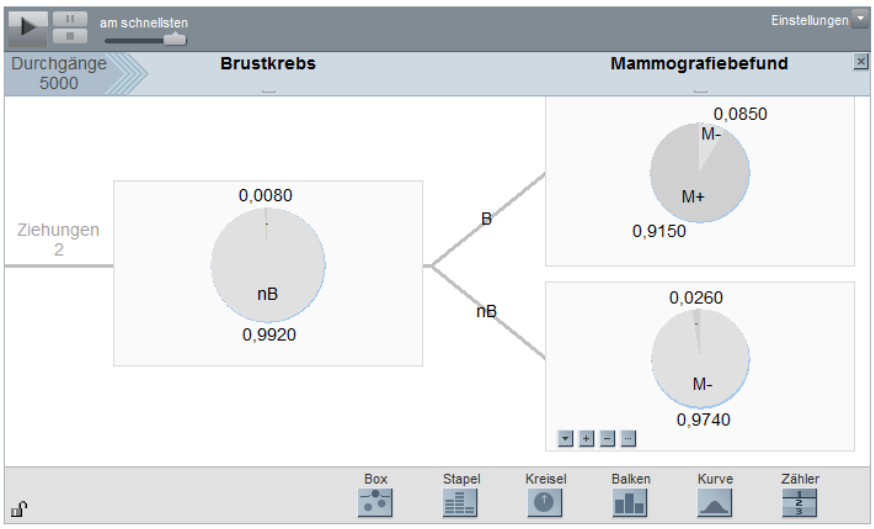


Abbildung 127: Simultane Simulation der Brustkrebs-Aufgabe

In Abbildung 127 findet sich eine intuitive Umsetzung der Brustkrebs-Aufgabe. Das Modellierungsinterface lässt sich hierbei entsprechend der Angaben bedienen. Als Unterschied zwischen der Simulation und einem Baumdiagramm lässt sich festhalten, dass die Pfadwahrscheinlichkeiten in den Bauteilen stehen und die Pfade mit den „Bedingungen“ beschriftet sind im Vergleich zu einem Baumdiagramm. Dies erfordert vom Nutzer ein kurzes Nachdenken, lässt sich jedoch intuitiv bedienen.

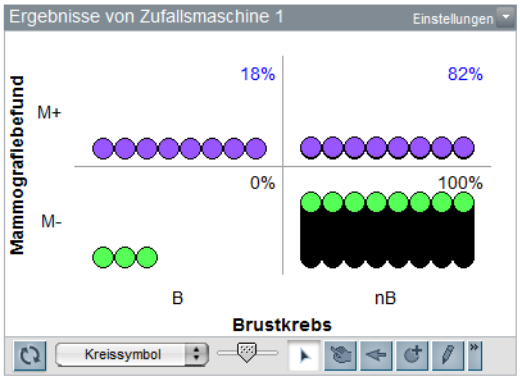


Abbildung 128: Auswertung der Brustkrebsaufgabe

Zur Auswertung können die automatisch erzeugten Merkmale der Tabelle in einer Graphik als Vierfeldertafel dargestellt werden und mit Einblendung der entsprechenden Prozente (Zeilenprozente in Abbildung 128) kann das gesuchte Ereignis identifiziert werden. Die Wahrscheinlichkeit, dass eine Frau zwischen 40 und 50 Jahren tatsächlich Brustkrebs hat, wenn sie einen positiven Mammografiebefund erhalten hat, kann somit auf 18% geschätzt werden.

Alternativ lässt sich im Sinne von Wassner, Biehler, Schweynoch und Martignon (2007) eine solche Aufgabenstellung auch mit natürlichen Häufigkeiten bearbeiten, was eine didaktische Vereinfachung der Situation darstellen kann. Hierzu müsste die Anzeige in Abbildung 128 auf Anzahlen statt Prozente umgestellt werden.

Sequentiell lässt sich eine solche Aufgabenstellung nicht simulieren.

Bewertung Sofern der Nutzer ein Baumdiagramm vor Augen hat, lässt sich eine Situation mit bedingten Wahrscheinlichkeiten intuitiv in die Zufallsmaschine von TinkerPlots übertragen. Dabei ist lediglich darauf zu achten, dass die Bauteile, die auf Abzweigungen folgen, zu dem jeweiligen Pfad passen, weil TinkerPlots die Beschriftungen von Abzweigungen automatisch vornimmt (vgl. Kapitel 3.5.1). In der Bauteilwahl ist der Nutzer insoweit frei, wie es zu der zu simulierenden Situation passt. Sehr nutzerfreundlich kann bei solchen Fragestellungen auf Formeln völlig verzichtet werden, die Definition und gleichzeitige Auswertung der interessierenden Ereignisse geschieht direkt mit den simulierten Merkmalen in einem Graphen. Das zusammengesetzte Merkmal ‚Gesamt' wird hierbei nicht gebraucht.

Anders als auf eine simultane Art kann eine Situation mit bedingten Wahrscheinlichkeiten nicht simuliert werden. Dies ist jedoch kein Hindernis in der Bedienung, da der Weg über die simultane Simulation für den Nutzer naheliegend ist.

Simulieren zu Aufgaben aus Glücksspielkontexten mit abhängigen, verschiedenartigen Teilexperimenten

Beispiel Als nächstes wird das Beispiel eines Glücksspiels betrachtet, das für eine weitere Art von mehrstufigen Zufallsexperimenten mit verschiedenen Teilexperimenten steht. „Betrachten wir hier folgendes fiktives Spiel: Es wird zunächst eine faire Münze geworfen. Fällt Wappen, so wird ein fairer Würfel geworfen, fällt Zahl, so wird die Münze erneut geworfen. Bei einem Einsatz von einem Euro tritt folgende Gewinnverteilung in Kraft:

$$X: \begin{cases} ZW, ZZ \rightarrow -5€ \\ W1, W2, W3 \rightarrow 2€ \\ W4, W5 \rightarrow 3€ \\ W6 \rightarrow 6€ \end{cases}$$

Ist das Spiel fair?" (Maxara 2009, 162).

Möglichkeiten Glücksspielsituationen, hinter denen Zufallsexperimente mit abhängigen, verschiedenartigen Teilexperimenten stehen, lassen sich in TinkerPlots nur simultan simulieren.

Umsetzung Auch dieses Beispiel lässt sich mit Hilfe eines Baumdiagramms darstellen, das sich in TinkerPlots in der Zufallsmaschine umsetzen lässt. Auf der ersten Stufe, das heißt im ersten Merkmal, wird die Münze modelliert, auf der zweiten Stufe wird je nach Ausgang des ersten Münzwurfs entweder eine Münze oder ein Würfel modelliert und auf der dritten Stufe, das heißt im dritten Merkmal, wird der Gewinn modelliert. Dies ist in Abbildung 129 zu sehen.

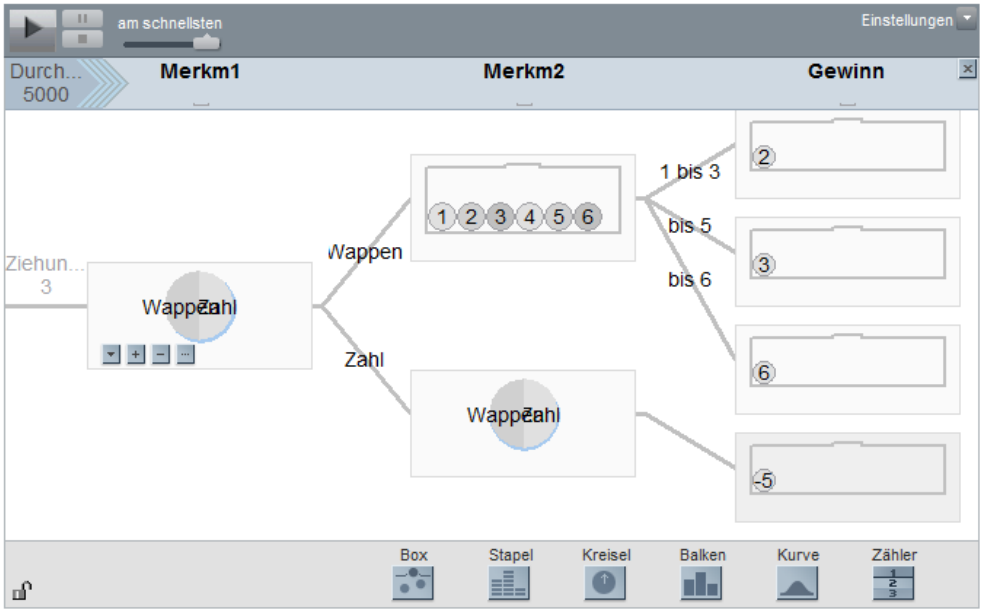


Abbildung 129: Simultane Simulation eines Glücksspiels

Die Boxen, die zu dem Merkmal ‚Gewinn' gehören, beinhalten jeweils nur ein Element, da streng genommen an dieser Stelle keine Zufallsauswahl mehr herrscht, sondern der Gewinn ausschließlich von der zuvor in Merkmal2 gewürfelten Zahl oder geworfenen Münze abhängt. Die untere Box des Merkmals2 ließe sich auch durch die unterste Box des Merkmals ‚Gewinn' ersetzen, denn wenn es vielleicht noch nicht in der Aufgabenstellung ersichtlich wird, dann doch spätestens in der Modellierung mit Hilfe der Zufallsmaschine wie in Abbildung 129. Hier lässt sich gut erkennen, dass ohne weitere Abzweigung der Pfad zu „-5" führt, sollte beim ersten Münzwurf (Merkm1) Zahl gefallen sein. Der Vollständigkeit halber ist der zweite Münzwurf in Merkm2 jedoch aufgeführt.

Die interessierende Zufallsgröße ist der Nettogewinn, sprich der Gewinn abzüglich des Einsatzes. Dieser kann in der Tabelle in einem weiteren Merkmal definiert werden durch die Formel *-1+Gewinn* (Abbildung 130, rechte Spalte).

Ergebnisse von Zufallsmaschine 1 Einstellungen

	Gesamt	Merkm1	Merkm2	Gewinn	Netto_Gewinn
Formel					– 1 + Gewinn
4994	Zahl;Zahl;-5	Zahl	Zahl	-5	-6
4995	Zahl;Wappen;-5	Zahl	Wappen	-5	-6
4996	Zahl;Wappen;-5	Zahl	Wappen	-5	-6
4997	Zahl;Wappen;-5	Zahl	Wappen	-5	-6
4998	Wappen;3;2	Wappen	3	2	1
4999	Wappen;1;2	Wappen	1	2	1
5000	Wappen;4;3	Wappen	4	3	2

Abbildung 130: Definition der Zufallsgröße „Netto_Gewinn“ des Glücksspiels

Alternativ dazu könnte der Einsatz auch als Merkmal in der Simulation realisiert werden in einer eigenen Merkmalsspalte, zum Beispiel vor der ersten oder hinter der letzten in Abbildung 129. Nähme man dafür die Box mit einer Kugel, beschriftet mit -1, so ließe sich in der Tabelle die gesuchte Zufallsgröße auch mithilfe der beiden Merkmale *Einsatz+Gewinn* realisieren. Dies ist Geschmackssache des Nutzers, sofern ihm die Vorstellung eher behagt, zum Beispiel die Simulation mit dem Einsatz zu starten, so ließe sich auch dies leicht realisieren.

Die Auswertung erfolgt wiederum in einem Graphen. In Abbildung 131 ist zu sehen, dass das arithmetische Mittel der Zufallsgröße ‚Nettogewinn‘ bei cirka -2 liegt, dass also im Durchschnitt auf lange Sicht mit einem Verlust von zwei Euro zu rechnen ist. Somit ist das Spiel nicht fair.

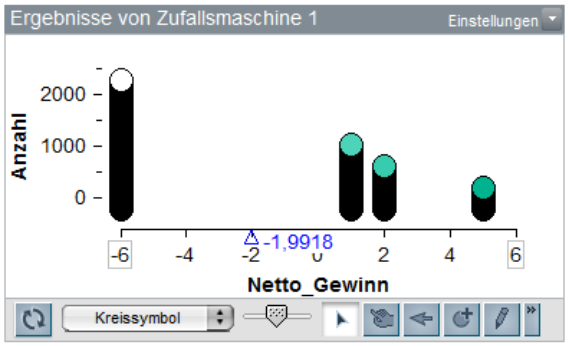


Abbildung 131: Auswertung der Zufallsgröße „Netto_Gewinn“ des Glücksspiels

Bewertung Auch ein Zufallsexperiment dieser Art lässt sich in TinkerPlots recht intuitiv simulieren. Wie zuvor kann die Modellierung der Situation in der Zufallsmaschine eng an der Vorstellung des Nutzers geschehen. Einzig die Beschriftung der Abzweigungen kann sich als etwas heikel herausstellen (siehe Ausführungen in Kapitel 3.5.1), aber mit der Hinweisfunktion von TinkerPlots kann auch diesen Schwierigkeiten begegnet werden, so dass mit hoher Sicherheit bei den Abzweigungen kein logischer Fehler in die Simulation durch den Nutzer eingebaut werden kann. Der Nettogewinn, der bei Glücks-

spielbeispielen dieser Art häufig als Zufallsgröße gefragt ist, lässt sich leicht per Formel in der Tabelle definieren. Eine große Formelkenntnis ist an dieser Stelle nicht nötig. Gleiches gilt für die Auswertung, diese geschieht wie immer in einem Graphen und die benötigte Darstellung ist leicht zu erzeugen.

Ein Vorteil der Modellierung einer solchen Situation mit direkten Abhängigkeiten ist, dass sich Zusammenhänge erkennen lassen, die durch die Darstellung offen gelegt werden können. Je vielschichtiger ein solches Modell jedoch wird, damit ist gemeint je mehr Teilexperimente es aufweist oder je mehr Bedingungen je Stufe unterschieden werden, desto mühseliger wird die Modellierung in TinkerPlots, dessen graphische Oberfläche durch den Computerbildschirm limitiert ist und deren Anzeigegröße der einzelnen Objekte nicht angepasst werden kann.

Wie bei dem Beispiel zuvor gilt, dass eine andere Simulationsart nicht in Frage kommt, um solche Beispiele mit TinkerPlots zu simulieren. Aber auch hier besteht für den Nutzer keine Notwendigkeit, einen anderen Weg wählen zu wollen, da die oben beschriebene Simulation nah an der Nutzervorstellung zu der Situation geschehen kann.

Simulieren von Markowketten

Markowketten lassen sich mit Hilfe der Zufallsmaschine von TinkerPlots nicht modellieren, weil eine rekursive Modellierung damit nicht möglich ist. Die einzige Möglichkeit besteht darin, den von Maxara beschriebenen Zugang über Formeln in der Tabelle zu wählen (Maxara 2009, 166f), allerdings widerspricht dies dem Modellierungsgedanken von TinkerPlots über die Zufallsmaschine. Aus diesem Grund werden zugehörige Beispiele hier nicht weiter ausgeführt, sondern auf die Ausführungen von Maxara verwiesen.

Zusammenfassung

Für mehrstufige Zufallsexperimente mit abhängigen Teilexperimenten zeigt sich, dass die simultane Simulation die einzige Möglichkeit zum Modellieren darstellt. Der Aufbau der Zufallsmaschine unterstützt den Nutzer, seine intuitiven Vorstellungen zu Urnenmodellierungen in TinkerPlots direkt umzusetzen und somit automatisch eine simultane Simulation durchzuführen. Besteht ein Zufallsexperiment aus Teilexperimenten, die sich als Ziehen ohne Zurücklegen aus einer Urne modellieren lassen, so unterstützt TinkerPlots dessen Modellierung direkt durch die Einstellungsmöglichkeiten an der Zufallsmaschine. Die Modellierung hierzu ist technisch einfacher zu erstellen als die zu vielen verschiedenartigen Teilexperimenten, weil ein Bauteil die Situation komplett abbildet. Sind die Teilexperimente verschiedenartig, abhängig oder unabhängig, so lässt sich mit der Zufallsmaschine die Vorstellung eines Baumdiagramms modellieren, das sehr universal einsetzbar ist. An dieser Stelle kann für den Nutzer die Abbildung der

Zusammenhänge der Teilexperimente in der Zufallsmaschine eine weitere Verstehenshilfe sein.

Zufallsgrößen und interessierende Ereignisse lassen sich je nach Fragestellung sehr direkt über das Merkmal ‚Gesamt' in einem Graphen darstellen und auswerten. Ansonsten lassen sie sich über Formeln in der Tabelle realisieren, diese sind häufig einfach wie in den obigen Beispielen, so dass der Nutzer selten komplizierte Formeln benötigt.

Die sequentielle Simulationsart ist nicht geeignet, um mehrstufige Zufallsexperimente mit abhängigen Teilexperimenten zu modellieren, allerdings ist dies kein Nachteil für den Nutzer, weil die simultane Simulation an dieser Stelle der Weg ist, der intuitiv gewählt wird.

Für Markowketten bietet die Zufallsmaschine von TinkerPlots keine Unterstützung an, da solche Problemstellungen nicht modellierbar sind.

4.2.4 Simulation von Wartezeitproblemen

Maxara (2009, 170f) klassifiziert Wartezeitprobleme in solche, bei denen es um die mittlere Wartezeit auf den ersten Treffer, die mittlere Wartezeit auf den k-ten Treffer oder um das Problem der vollständigen Serie geht. Diese drei Arten von Wartezeitproblemen werden in den nächsten drei Abschnitten betrachtet. Maxara beschränkt sich dabei auf Zufallsexperimente mit stochastisch unabhängigen Teilexperimenten, diese Einschränkung wird hier übernommen.

TinkerPlots bietet dabei hohes Unterstützungspotential an, weil bereits in der Zufallsmaschine eine Abbruchbedingung für die Simulation definiert werden kann.

Simulieren eines Wartezeitproblems bis zum ersten Treffer

Beispiel Als einfaches, aber typisches Beispiel wählt Maxara die mittlere Wartezeit bis zur ersten Sechs beim Werfen eines Würfels. Wie oft muss im Durchschnitt geworfen werden, bis bei einem Würfel die erste Sechs fällt? Die Simulation hierzu wurde bereits im vorigen Kapitel ausgeführt (Kapitel 3.7). Es folgt an dieser Stelle eine knappe Zusammenfassung der Simulation.

Möglichkeiten Eine Problemstellung mit Warten auf den ersten Treffer lässt sich ausschließlich sequentiell simulieren.

Umsetzung Für die Modellierung eines Würfels stehen die Bauteile Box, Stapel, Kreisel und Balken zur Verfügung. Da es um das Werfen eines Würfels geht, muss die Ziehungsanzahl zunächst auf ‚1' gesetzt werden, wie in Abbildung 132.

Eine Durchgangsanzahl ist nicht angegeben in dieser Modellierung, weil diese gerade durch die Abbruchbedingung definiert wird und später die interessierende Zufallsgröße darstellt. Im Fenster „Info Zufallsmaschine" kann eine Abbruchbedingung definiert

werden. Dazu stehen für dieses Beispiel zwei Möglichkeiten zur Verfügung. Entweder fasst man das Auftreten einer Sechs als Muster auf und trägt in das Feld *Wiederholen bis folgendes Muster auftritt* eine ‚6' ein (Abbildung 133, Markierung) oder man benutzt den Formeleditor zur Definition einer Abbruchbedingung in dem Feld *Wiederholen bis zur Bedingung* (Abbildung 133) und trägt dort *Merkm1 = 6* ein. In beiden Fällen wird die Zufallsmaschine veranlasst, solange Kugeln (mit Zurücklegen) aus der Box zu ziehen, bis eine Sechs gezogen wurde.

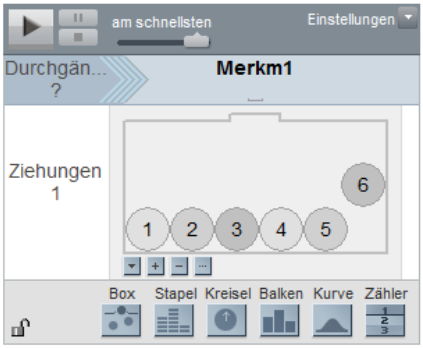


Abbildung 132: Modellierung des Würfelwurfs mit Abbruchbedingung

Es gibt in TinkerPlots keine weitere Möglichkeit als diese beiden, eine Abbruchbedingung zu formulieren. Solange sich die gesuchte Abbruchbedingung als Muster interpretieren lässt, ist die Unterstützung für den Nutzer durch TinkerPlots sehr hoch, weil keine Formeln verwendet werden müssen. Andernfalls steht der komplette Formelapparat zur Verfügung, wodurch eine Vielzahl von Abbruchbedingungen definierbar wird.

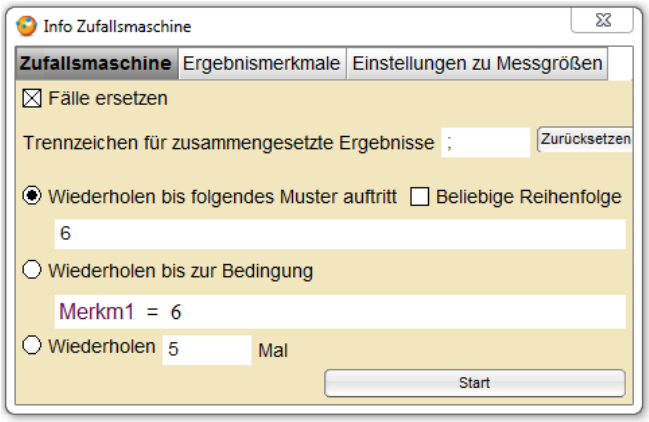


Abbildung 133: Definierte Abbruchbedingung für das Auftreten der ersten Sechs

Nach der Definition der Abbruchbedingung muss die Anzahl der benötigten Durchgänge in einem Graphen dargestellt und als Messgröße gesammelt werden (für detaillierte Ausführungen siehe Kapitel 3.7). In Abbildung 134 ist dieser Prozess zu sehen, einschließlich der Auswertung von 1000 gesammelten Messgrößen.

Der Auswertung in Abbildung 134 zufolge lässt sich schätzen, dass im Mittel sechsmal geworfen werden muss, bis die erste Sechs auftritt.

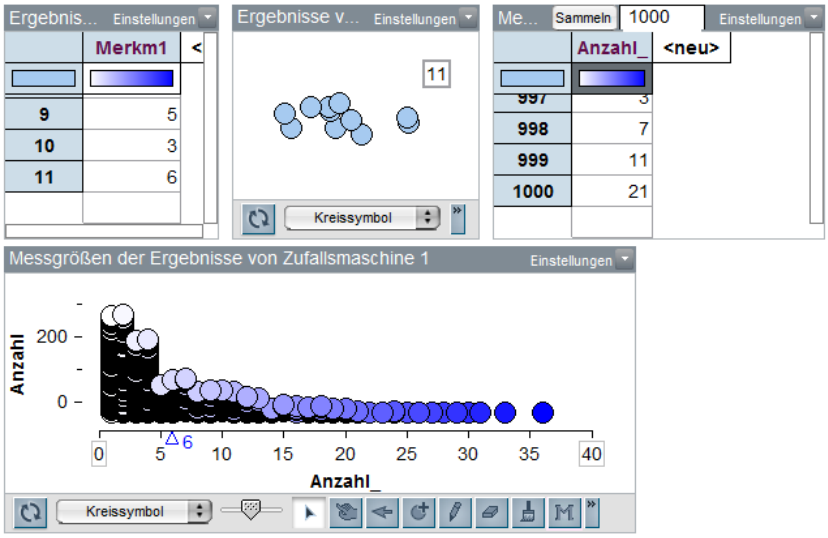


Abbildung 134: Festlegen, Sammeln und Auswerten der Messgröße "Anzahl Würfe bis zur ersten Sechs"

Simulieren eines Wartezeitproblems bis zum k-ten Treffer

Beispiel Ändert man das vorige Beispiel dahingehend ab, dass nun etwa die mittlere Wartezeit beim Würfeln bis zum dritten Treffer, also bis zum dritten Mal eine Sechs fällt, so muss zwingend mit dem Formeleditor gearbeitet werden.

Möglichkeiten Ein Wartezeitproblem mit Warten auf den k-ten Treffer lässt sich wie ein Wartezeitproblem mit Warten auf den ersten Treffer nur sequentiell simulieren.

Umsetzung Dies lässt sich nicht mehr als Muster für TinkerPlots interpretieren, so dass allgemein mit der Formel *Anzahl(Merkm1="Ausprägung")=k* als Abbruchbedingung gearbeitet werden muss. Im Beispiel des Wartens auf die dritte Sechs ist dies *Anzahl(Merkm1=6)=3*, wie in Abbildung 135 zu sehen.

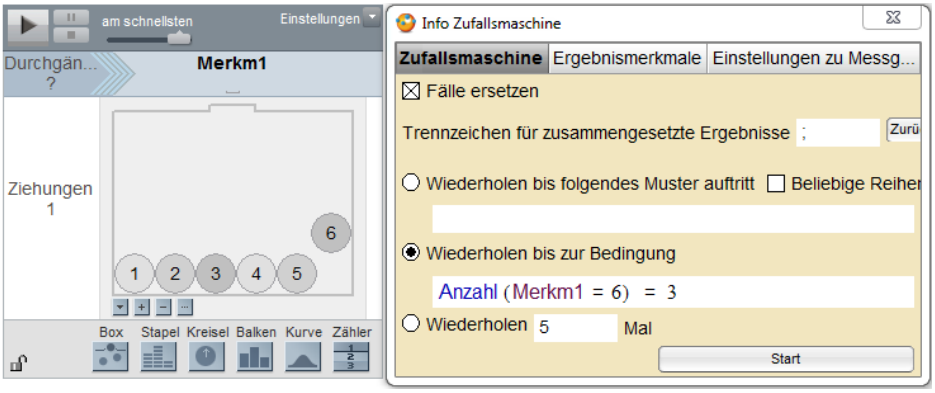


Abbildung 135: Definition der Abbruchbedingung "Warten bis zur dritten Sechs"

Wie beim vorherigen Beispiel des Wartens auf die erste Sechs stellt nun die gesamte Wurfanzahl die gesuchte Zufallsgröße dar und somit kann die Anzahl der Durchgänge als Messgröße gesammelt und ausgewertet werden (Abbildung 136).

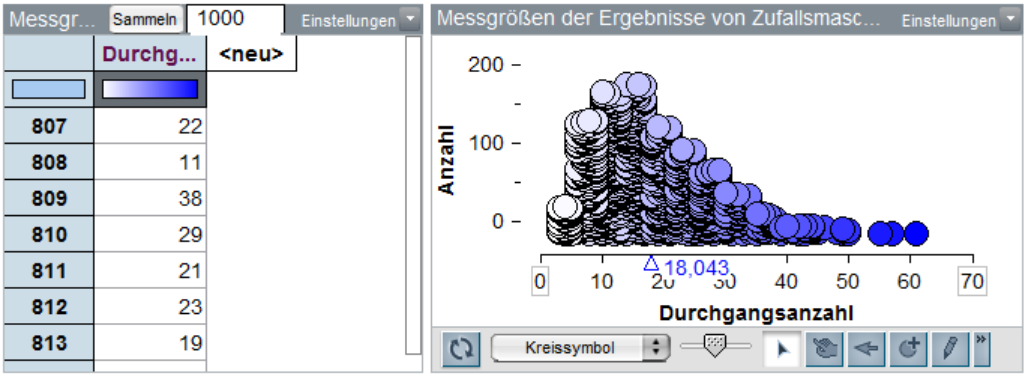


Abbildung 136: Sammeln und Auswerten der Messgröße "Warten bis zur dritten Sechs"

Bewertung Wartezeitprobleme, bei denen die mittlere Wartezeit bis zum k-ten Treffer (dies schließt den ersten Treffer mit ein) simuliert werden soll, lassen sich ebenfalls nur auf diese Weise realisieren. Die Prozessvorstellung eines Nutzers zu einer solchen Problemstellung wird durch TinkerPlots gut abgebildet. Abgesehen davon, dass die entsprechende Messgröße nur per Formel definiert werden kann, läuft die Simulation analog zu einem Wartezeitproblem, bei dem auf den ersten Treffer gewartet wird. Natürlich ließen sich auf diese Art und Weise auch mehrstufige Zufallsexperimente realisieren, indem die Anzahl der Ziehungen entsprechend erhöht würde oder weitere Bauteilspalten ergänzt würden. Auch eine Modellierung eines Zufallsexperiments mit abhängigen Teilexperimenten als Wartezeitproblem ließe sich auf diese Weise simulieren.

Simulieren eines Wartezeitproblems bis zur vollständigen Serie

Beispiel Das folgende Beispiel illustriert eine Situation des Wartens auf eine vollständige Serie. „In jedem siebten Überraschungsei ist eine Figur einer Serie mit fünf verschiedenen Figuren versteckt. Angenommen jede Figur kann mit gleicher Wahrscheinlichkeit gezogen werden. Wie viele Eier muss man im Mittel kaufen, um alle fünf Figuren mindestens einmal zu besitzen?“ (Maxara, 2009, 175).

Möglichkeiten Diese Situation lässt sich nur sequentiell simulieren.

Umsetzung Interpretiert man das Auftreten einer Figur in jedem siebten Ei, so gehören 30 weitere Eier dazu, die nicht mit einer der gesuchten Figuren bestückt sind. Anders ausgedrückt liegt für jede der fünf verschiedenen Figuren die Wahrscheinlichkeit bei 1/35 für deren Auftreten und die Chance, dass keine Figur gezogen wird, bei 30/35. Diese Situation lässt sich in TinkerPlots beispielsweise direkt mit dem Bauteil Stapel modellieren, so wie in Abbildung 137.

In Abbildung 137 sind sechs Stapel gebildet, von denen Figur 1 bis Figur 5 jeweils ein Element besitzt und der Stapel ohne Beschriftung 30 Elemente umfasst. Diese Modellierung spiegelt die Situation der Figurenverteilung nahe an der Nutzervorstellung wider. Alternativ ließe sich die Situation auch mit anderen Bauteilen modellieren.

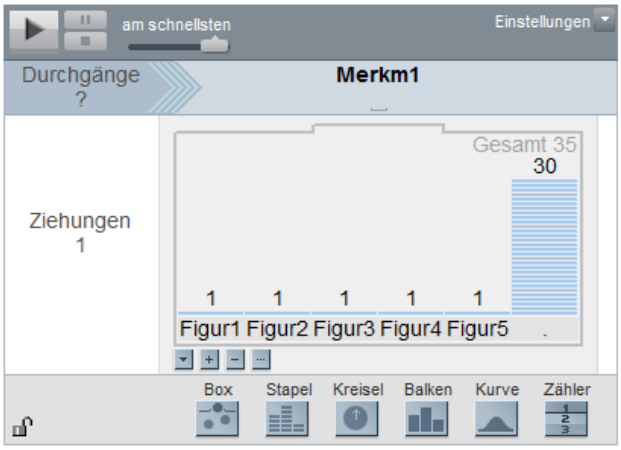


Abbildung 137: Sequentielle Modellierung der Überraschungseier-Situation

Nun interessiert die Durchgangsanzahl, bis jede der Figuren 1 bis 5 mindestens einmal aufgetreten ist (Abbildung 139).

Die Formel *AnzVerschiedeneWerte(Merkm1)=5* (Abbildung 138) lässt die Zufallsmaschine solange laufen, bis fünf verschiedene Werte gezogen wurden. Da die 30 Eier, die nicht mit Figuren bestückt sind, keine Beschriftung haben, werden diese auch nicht als Wert gezählt.

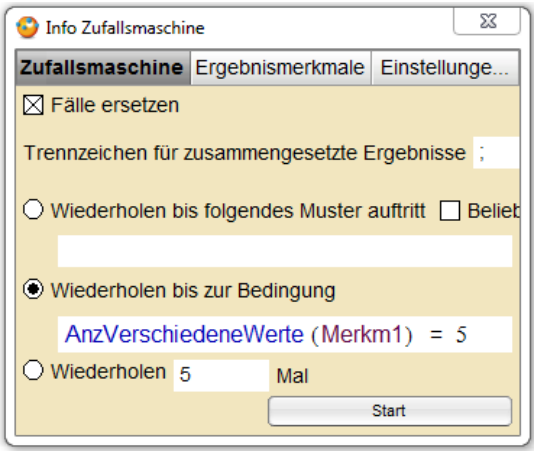


Abbildung 138: Definieren der Abbruchbedingung für die Überraschungseier-Aufgabe

Die Ergebnisliste sieht dann zum Beispiel so aus wie in Abbildung 139 links.

In Abbildung 139 wurden 21 Durchgänge benötigt, bis alle fünf Figuren gezogen wurden. Die Zufallsgröße *X: Anzahl der gezogenen Eier, bis alle fünf Figuren mindestens*

einmal gezogen wurden lässt sich als Messgröße der Anzahl im Graph definieren (Abbildung 139, rechts).

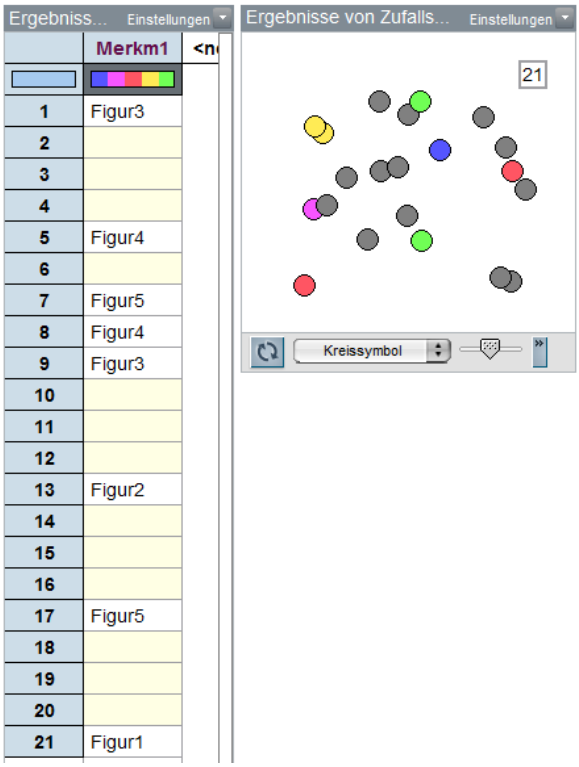


Abbildung 139: Beispielergebnis einer Simulation zur Überraschungseier-Aufgabe

Nun muss noch die Anzahl der Durchgänge als Messgröße gesammelt und ausgewertet werden (Abbildung 140).

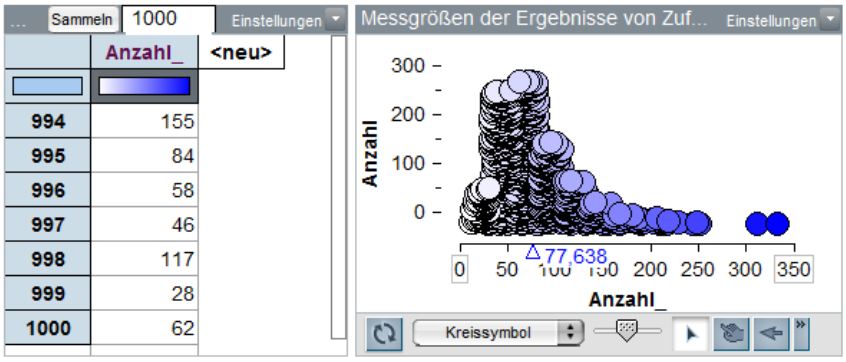


Abbildung 140: Auswerten der Ergebnisse zur Überraschungseier-Aufgabe

Es lässt sich erkennen, dass bei 1000 Simulationswiederholungen im Mittel 77,638 Eier gezogen werden mussten, um alle fünf Figuren zu bekommen.

Bewertung Auch ein Problem der vollständigen Serie lässt sich ausschließlich sequentiell in TinkerPlots simulieren. Wiederum gilt, dass die Modellierung dabei nah an der

Vorstellung des Nutzers geschehen kann, mit anderen Bauteilen sind auch abstraktere Modellierungen möglich. Eine Veränderung des Anteils jeden Treffers (im obigen Beispiel jeder Figur) ist ebenfalls leicht möglich, zum Beispiel durch Hinzufügen weiterer Elemente zu einem Stapel. Die Abbruchbedingung muss mittels Formel erfolgen, dies lässt sich für eine vollständige Serie jedoch gut mit der Formel *AnzVerschiedeneWerte(Merkm1)* definieren. Das Festlegen der Zufallsgröße ist einfach gestaltet durch das Auswählen der Durchgangsanzahl oder der Anzahl im Graph als Messgröße. Wie sonst auch ist die Auswertung am Ende der Simulation über den Graphen leicht.

Zusammenfassung

Die Simulation von Wartezeitproblemen lässt sich ausschließlich per sequentieller Simulation realisieren. Dabei ist die Modellierung wie sonst auch in vielen Fällen nah an der Nutzervorstellung möglich durch die verschiedenen Bauteile und den prinzipiellen Aufbau der Zufallsmaschine. Der Benutzer muss sich im Vorfeld keine Gedanken über die auszuwählende Simulationsart machen, da sich diese automatisch ergibt. Die größte technische Fehlerquelle liegt in der falschen Wahl für die Ziehungsanzahl, die in vielen Situationen „1“ betragen muss. Das Festlegen der Abbruchbedingung geschieht entweder über die Eingabe eines Ereignisses über die Musterdefinition in der Zufallsmaschine oder, etwas universeller, über die Eingabe einer Formel. Dabei gibt die folgende Tabelle 9 einen kurzen Überblick über häufig verwendete Abbruchbedingungen in Bezug auf das Werfen eines Würfels.

Tabelle 9: Häufige Abbruchbedingung und deren Umsetzung für einen Würfelwurf

Abbruchbedingung	Formel in TinkerPlots
Warten bis zur ersten Sechs	*Merkm1 = 6*
Warten bis zur r-ten Sechs	*Anzahl(Merkm1=6) = r*
Warten bis r-mal hintereinander eine gleiche Zahl gefallen ist	*Siehe oben*
Warten, bis jede Zahl mindestens einmal gefallen ist	*AnzVerschiedeneWerte(Merkm1) = 6*

Die Musterdefinition als Abbruchbedingung eignet sich vor allem bei mehrstufigen Zufallsexperimenten, wenn beispielsweise fünf Würfel geworfen werden und man sich für das Ereignis „Große Straße“, also die Zahlenfolge 1, 2, 3, 4, 5 interessiert. In diesem Fall ist die Mustereingabe sehr komfortabel, da sich das interessierende Ereignis direkt eingeben lässt, so wie es auftreten würde. Eine besondere Hilfe sind dabei die Variablen V1, V2, … und der Asterix „*“, der für einen beliebigen Wert steht.

4.3 Ergänzungen

Zusätzlich zu den bisherigen Ausführungen werden in den folgenden zwei Unterabschnitten noch zwei Ergänzungen gegeben, die sich nicht in die obige Struktur einbinden lassen. Zunächst wird eine Lernumgebung zum Gesetz der großen Zahlen als

stochastische Ergänzung vorgestellt, da dies den Hintergrund für Simulationen liefert. Diese lässt sich beispielsweise als vorbereitete Lernumgebung nutzen, um Erfahrungen mit dem Gesetz der großen Zahlen zu machen. Als weitere Ergänzung aus Softwaresicht wird eine Funktion der Zufallsmaschine vorgestellt, mit der sich die Inhalte der Zufallsmaschine verbergen lassen, um Schlüsse aus den Ergebnissen einer Simulation auf eine zugrundeliegende Wahrscheinlichkeit ziehen zu können.

4.3.1 Ergänzung: Gesetz der großen Zahlen

Es folgt nun noch eine Beispielsimulation zur Veranschaulichung des Gesetzes der großen Zahlen als Ergänzung zu den vorherigen Beispielen. Dies lässt sich beispielsweise als Demonstration einsetzen und zeigt die Grundlagen, auf denen eine Simulation aufbaut. Dies wird über den Münzwurf realisiert, weil dies der Standardkontext ist, in dem das Gesetz der großen Zahlen thematisiert und/oder eingeführt wird. Dabei ist die Entwicklung der relativen Häufigkeiten bei immer größer werdenden Wurfanzahlen von Interesse.

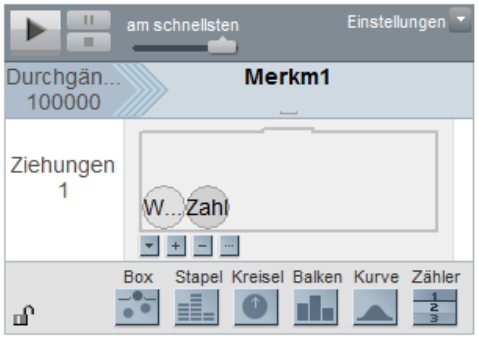


Abbildung 141: Modellierung einer Münze

Es erfolgt zunächst eine simultane Simulation zu einer Münzwurfserie zur Veranschaulichung der Entwicklung der relativen Häufigkeiten von Wappen. Die Modellierung in Abbildung 141 erfolgt entsprechend der Standardvorstellung über die zwei Seiten einer Münze als beschriftete Kugeln in einer Urne. Es werden 100 000 Münzwürfe durchgeführt, das Maximum dessen, was bei einem Simulationsdurchlauf in TinkerPlots möglich ist.

Ergebnisse von Zufallsmaschine 1 — Einstellungen

	Merkm1	laufendeAnzahl_Wappen	Relative_Häufigkeit	Anzahl_Würfe
Formel		VorgängerWert (laufende	laufendeAnzahl_Wappen / index	index
99998	Wappen	50209	0,5021	99998
99999	Zahl	50209	0,502095	99999
100000	Wappen	50210	0,5021	100000

Abbildung 142: Benötigte Merkmale für die Visualisierung zum Gesetz der großen Zahlen

In einer Tabelle (Abbildung 142) wird nun mit Hilfe von Formeln die Entwicklung der relativen Häufigkeit von Wappen berechnet. Dazu werden drei weitere Merkmale benötigt. In Merkm1 sind zunächst die Simulationsergebnisse zu sehen, welche Seite der Münze gefallen ist. Zunächst wird die absolute Anzahl an gefallenen Wappen bestimmt über die vordefinierte Formel *Laufende Anzahl von ‚Wappen' über alle Durchgänge* in der Spalte „laufendeAnzahl_Wappen". Die zugrunde liegende Formel wählt TinkerPlots automatisch. In der nächsten Spalte „Relative_Häufigkeit wird daraus die relative Häufigkeit bestimmt, mit der bis zu der jeweiligen Zeile ‚Wappen' gefallen ist, indem die laufende Anzahl von Wappen durch den Index geteilt wird. Für die Darstellung im Graphen wird zusätzlich noch das Merkmal „Anzahl_Würfe" über die Index-Funktion bestimmt. In einer vorbereiteten Lernumgebung müssen die benötigten Formeln von den Lernenden nicht unbedingt verstanden werden. Es kann genügen zu erklären, in welchem Merkmal welche Information dargestellt wird. Somit ist die Verstehenshürde für die graphische Auswertung niedriger, als wenn erwartet wird, dass die kompletten Formeln verstanden und benutzt werden sollen. Es kann zusätzlich die kognitive Last verringert werden, indem die Formelzeile komplett ausgeblendet wird.

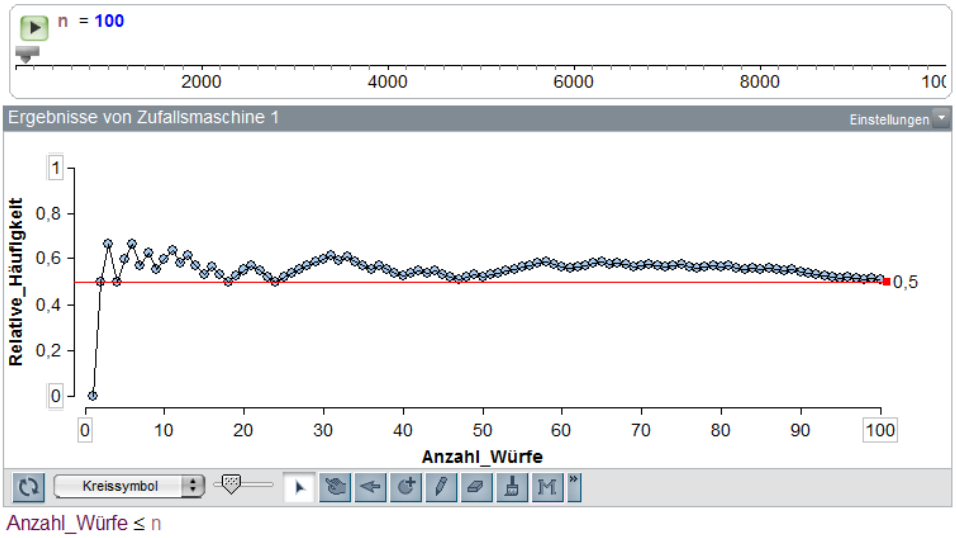


Abbildung 143: Relative Häufigkeit von Wappen bei n = 100 Münzwürfen

Mit Hilfe eines Reglers (Abbildung 143, oben) und eines Filters (Abbildung 143, unterhalb des Graphen) lassen sich nun verschiedene Umfänge einer Münzwurfserie darstellen. Abbildung 143 zeigt die Entwicklung der relativen Häufigkeit von Wappen bei 100 Münzwürfen, Abbildung 144 zeigt 1000 Münzwürfe und Abbildung 145 zeigt 10 000 Münzwürfe und die jeweilige Entwicklung der relativen Häufigkeit von Wappen.

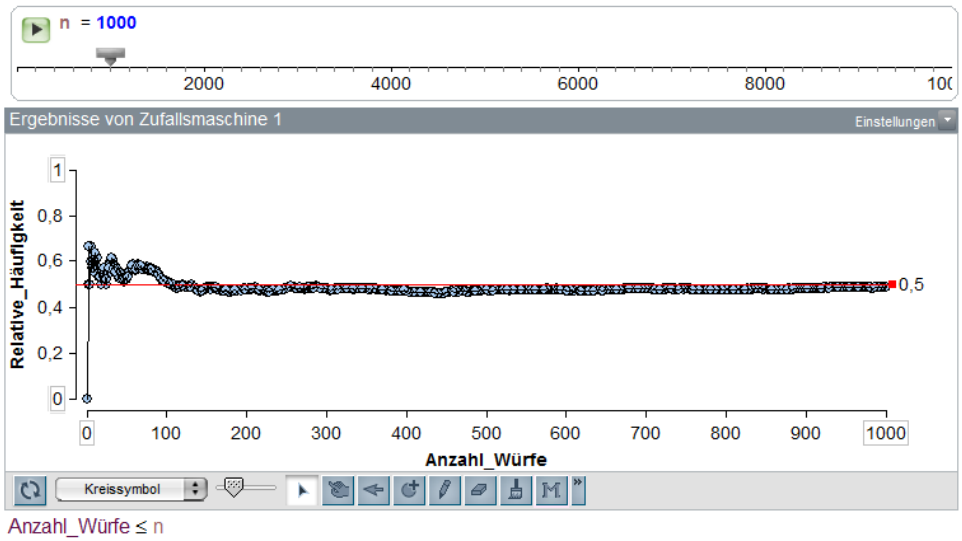


Abbildung 144: Relative Häufigkeit von Wappen bei n = 1000 Münzwürfen

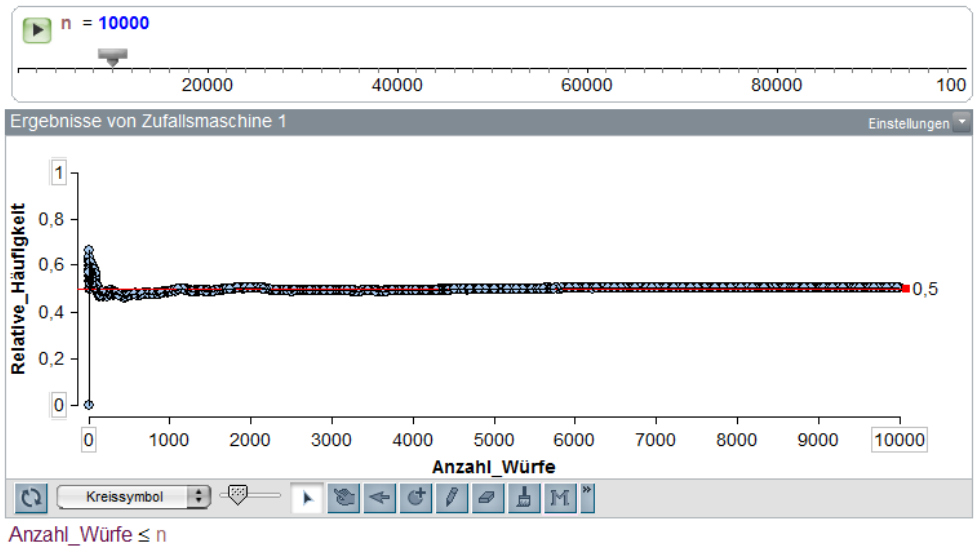


Abbildung 145: Relative Häufigkeit von Wappen bei n = 10 000 Münzwürfen

Anmerkung: Die Darstellung von 100 000 Fällen in einem Graphen erfordert eine sehr hohe Rechenleistung. Dadurch ist es möglich, das Programm zum Absturz zu bringen, auf jeden Fall bringt es mehrere Minuten Wartezeit mit sich. Ein Hantieren im Graph ist mit so vielen Fällen kaum noch möglich, insofern wird hier auf die Darstellung der relativen Häufigkeit von Wappen bei n = 100 000 Würfen verzichtet und dem Leser empfohlen, dies nur durchzuführen, wenn er einen leistungsstarken Rechner zum Einsatz bringt.

In TinkerPlots kann in einem Graph immer nur eine Trajektorie von Münzwürfen dargestellt werden. Durch Betätigen des Startbuttons in der Zufallsmaschine kann jedoch

jederzeit eine neue Münzwurfserie erzeugt werden, was bei Anzahlen bis zu 10 000 Fällen auch binnen weniger Sekunden oder Sekundenbruchteilen erfolgt.

Als Erweiterung können nun noch mehrere Münzwurfserien verschiedener Umfänge und deren Streuung dargestellt werden. Dazu muss die Simulation mit verschiedenen Umfängen wiederholt und die jeweilige Durchgangszahl als Messgröße (Merkmal n in Abbildung 146) gesammelt werden.

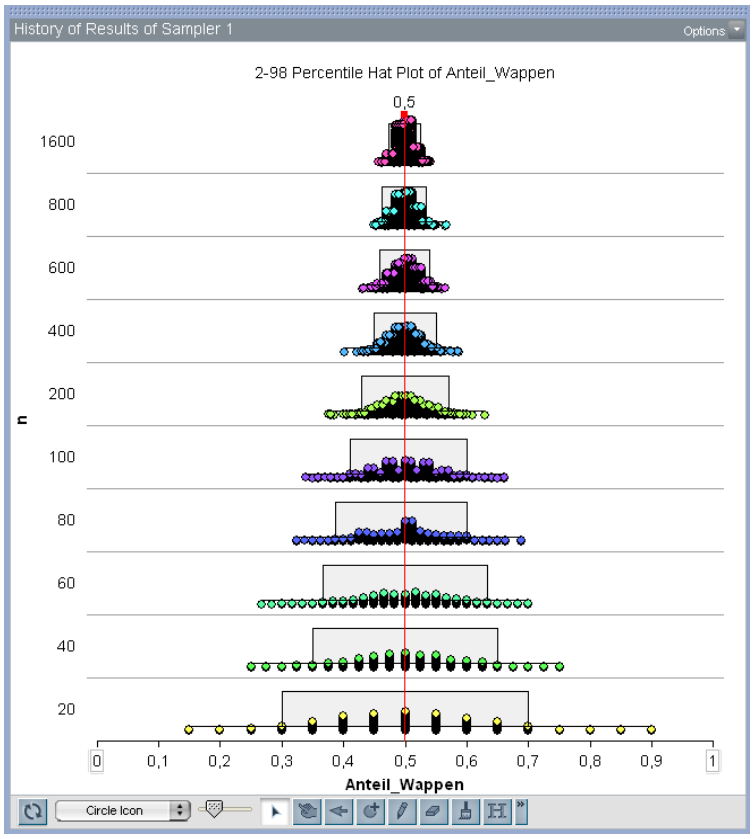


Abbildung 146: Darstellung der relativen Häufigkeit von Wappen in Abhängigkeit des Stichprobenumfangs bei eingeblendetem mittlerem 96%-Hut und Referenzlinie bei 0,5

In Abbildung 146 sind Stichprobenumfänge von n = 20, 40, 60, 80, 100; 200; 400; 600; 800 und 1600 dargestellt. Es lässt sich gut erkennen, wie die Schwankungsbreite mit zunehmendem Stichprobenumfang abnimmt.

4.3.2 Ergänzung: Die versteckte Zufallsmaschine

Aus didaktischer Sicht ist die Funktion *Verbergen* der Zufallsmaschine interessant. So kann eine Aufgabe gestellt werden, bei der die Inhalte der Zufallsmaschine verborgen werden, um anhand der Stichprobe zu entscheiden, was sich in der Zufallsmaschine verbirgt. Daran lassen sich beispielsweise das Thema ‚Genauigkeit in Abhängigkeit von der Stichprobengröße‘ oder auch das Gesetz der großen Zahlen mit einem anderen Zu-

gang erfahrbar machen. Eine solche „Rate-Aufgabe“ kann beispielsweise wie folgt lauten: *In der Zufallsmaschine verbirgt sich ein gezinkter zwölfseitiger Würfel. Finde heraus, welche Zahl wie oft auf diesem Würfel vorkommt. Versuch dabei, mit so wenig Durchgängen wie möglich auszukommen, um dennoch ein verlässliches Ergebnis zu erhalten.* Die zur Aufgabe gehörige Zufallsmaschine sieht dann wie in Abbildung 147 aus.

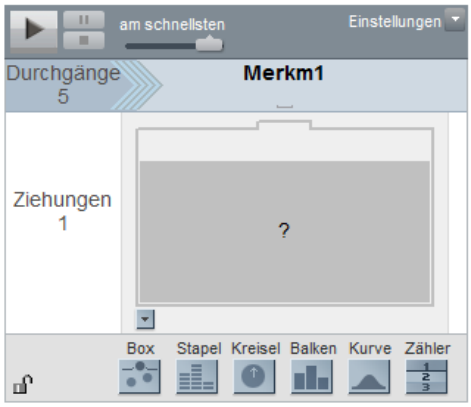


Abbildung 147: Zufallsmaschine mit verborgenem Inhalt

Anhand einer Verteilung wie in Abbildung 148 kann nun versucht werden, auf die Zahlen des zwölfseitigen Würfels zu schließen oder durch eine Erhöhung des Stichprobenumfangs, das heißt der Durchgangsanzahl, ein verlässlicheres Ergebnis zu erhalten.

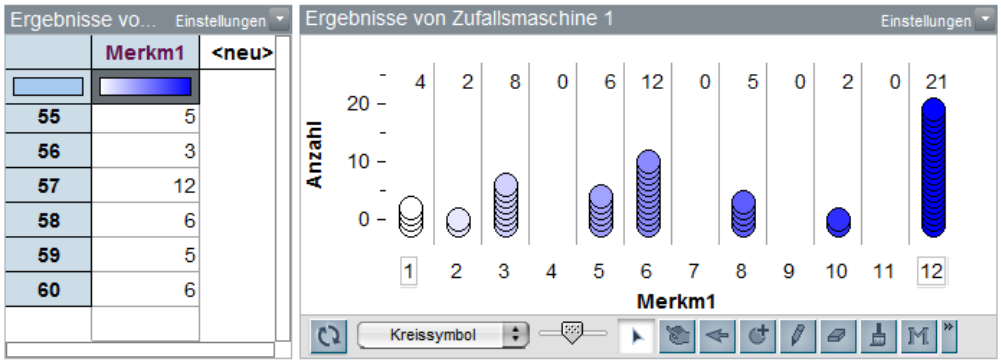


Abbildung 148: Beispielverteilung für n = 60 Würfe des versteckten Dodekaeders

Die Zufallsmaschine in Abbildung 147 lässt sich normal bedienen, es kann die Anzahl der Ziehungen und der Durchgänge variiert werden und über das Pfeilmenü lassen sich auch die Inhalte wieder anzeigen. Nun kann durch eine sukzessive Erhöhung der Durchgänge und der Darstellung der Ergebnisse in einem Graphen versucht werden, die zwölf Zahlen zu raten, die sich in der Zufallsmaschine befinden. Soll ein Lerner die Zufallsmaschine nicht manipulieren können, insbesondere also auch nicht die Inhalte wieder anzeigen können, so lässt sich mit Hilfe des Schlosses links unten in Abbildung 147 zusätzlich die Zufallsmaschine sperren, wenn gewünscht auch mit einem Passwort ver-

sehen, so dass ausschließlich das Steuerungsfeld mit dem Start-, Stop- und Pausebutton bedient werden kann. Die Ansicht der Zufallsmaschine sieht dann aus wie in Abbildung 149.

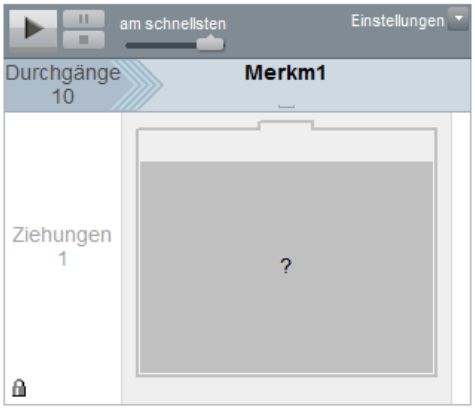


Abbildung 149: Gesperrte Zufallsmaschine mit verborgenem Inhalt

Die Anzahl der Durchgänge lässt sich für die gesperrte Zufallsmaschine in Abbildung 149 nur noch über das Einstellungsmenü rechts oben variieren, indem im Fenster „Info Zufallsmaschine“ eine andere Wiederholungszahl eingegeben wird (Abbildung 150).

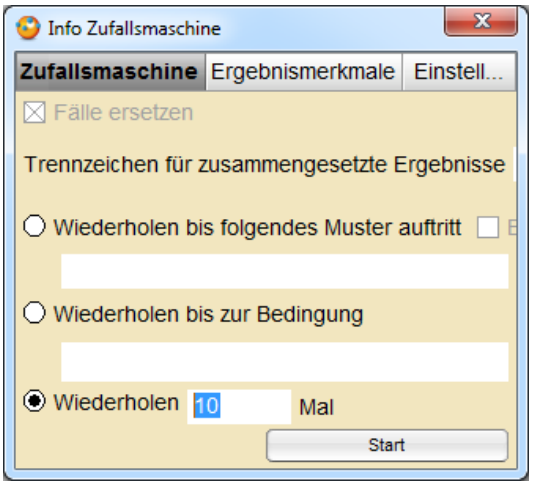


Abbildung 150: Veränderungsmöglichkeit der Durchgangsanzahl bei einer gesperrten Zufallsmaschine

Somit kann der Lerner die gestellte Aufgabe bearbeiten, ohne eine Möglichkeit zu haben, die Inhalte der Zufallsmaschine unerlaubt sichtbar zu machen.

Mit dieser Funktion lassen sich spielerisch erste Erfahrungen machen bezüglich des Schließens von einer Stichprobe auf die zugrunde liegende Wahrscheinlichkeitsverteilung. Dieser spielerische Aspekt trägt zu einer hohen Motivation bei solchen Aufgaben bei. Aus Lehrersicht ist eine entsprechende Zufallsmaschine leicht einzurichten und kann mit Hilfe des Passwortschutzes eine vorbereitete Lernumgebung darstellen.

4.4 Zusammenfassung

In Kapitel 3 wurde ausführlich die Unterstützung der Software TinkerPlots zur stochastischen Simulation analysiert. Dabei wurde ausgegangen von den technischen Gegebenheiten der Software. In Kapitel 4 bis hierher wurde anhand von stochastischen Problemstellungen das didaktische Potential der Software untersucht. Für den Einsatz in der Lehre wurde ein speziell auf TinkerPlots angepasster Simulationsplan entwickelt (Kapitel 2.4), auf dessen Grundlage verschiedene Schnittstellen zwischen Stochastik und Software nach der Klassifizierung von Maxara (2009) für TinkerPlots unterschieden wurden. Eine möglichst einfache und intuitive Bedienung der Schnittstellen diente als Kriterium für die Bewertung des Unterstützungspotentials von TinkerPlots zur stochastischen Simulation. Eine solche Analyse ist aus Sicht der instrumentellen Genese (Kapitel 2.2) nötig gewesen, um durch die Kenntnis der Grenzen und Möglichkeiten eines Werkzeugs dieses zu einem Instrument für den Nutzer werden zu lassen.

Der Vorteil der in Kapitel 3.5.2 identifizierten Simulationsarten ist, dass Simulationen in TinkerPlots immer in ähnlichen Schritten abfolgen und deshalb die Benutzung der Software wesentlich vereinfacht ist. Dieser intuitiv strukturierte Aufbau von Simulationen, nur wenige, überschaubare Schritte und eine leichte Anpassung der Simulationen sind wesentliche Merkmale der Software TinkerPlots.

Das **Modellierungsinterface** ist gegeben durch die Zufallsmaschine. In dieser lassen sich fünf verschiedene Bauteile zur Modellierung gegebener Situationen verwenden, von denen das Bauteil Box am universellsten die Grundvorstellung einer stochastischen Urne (Sedlmeier 1999) unterstützt. Aber auch die anderen Bauteile bieten mit ihren verschiedenen Eigenschaften und vor allem der Wechselmöglichkeit der Darstellungen ineinander ein hohes intuitves Unterstützungspotential der Visualisierung mentaler Modelle, ohne dass Formeln oder Programmiersprachen nötig sind. Gerade das Bauteil Kreisel unterstützt die Modellierung nicht gleichverteilter Ergebnisräume, was in anderen Programmen durchaus schwierig sein kann (McNamara 2015). Das Definieren von mehreren Bauteilspalten für mehrstufige Zufallsexperimente (sowohl für abhängige als auch für unabhängige Teilexperimente) hat ebenfalls einen hohen didaktischen Wert, da hierdurch Abhängigkeiten und Strukturen von Teilexperimenten direkt visualisiert werden können. Jedoch kann dies bei mehrstufigen Zufallsexperimenten abhängig von der Monitorgröße auch schnell an seine Grenzen stoßen, wenn die Bedienbarkeit der Software aufgrund der starren Anzeigegröße behindert wird.

Es wurden im Wesentlichen die vier Simulationsarten simultane Simulation, sequentielle Simulation, Simulation durch Stichprobenziehung und Simulation durch Randomisierung unterschieden. Dabei ist die Simulation durch Stichprobenziehung eine spezielle Art der simultanen Simulation. Die Simulation durch Randomisierung spielt zwar nur für eine geringe Anzahl an Zufallsexperimenten eine Rolle, allerdings ist gerade hierfür

die didaktische Unterstützung extrem groß (vgl. Kapitel 3.5.2.4). Ein großer Vorteil der Software TinkerPlots ist, dass sich viele Zufallsexperimente sowohl durch eine simultane Simulation als auch durch eine sequentielle Simulation realisieren lassen, ein Nutzer muss also nicht vorher überlegen, auf welche Art eine Simulation erstellt werden muss. Der Vorteil der simultanen Simulation ist das hohe Maß an Visualisierung eines mentalen Modells, allerdings werden hier zur Auswertung häufig Formeln benötigt. Der Vorteil einer sequenziellen Simulation ist, dass diese häufig ohne Formeln auskommt, allerdings werden hierzu mehr Objekte benötigt, wodurch die Arbeitsfläche und die Zuordnung der einzelnen Objekte unübersichtlich werden kann.

Das **Interface zu Ereignissen und Zufallsgrößen** kann über Merkmale oder über Messgrößen realisiert werden. Für eine simultane Simulation werden (meistens) Formeln benötigt. Hierfür ist durch die vordefinierten Auswertungsmerkmale das Unterstützungspotential von TinkerPlots hoch, da hierdurch keine Formelkenntnis benötigt wird. Das Messgrößenkonzept verzichtet komplett auf Formeln, wodurch die Fehleranfälligkeit hierbei äußerst gering und für den Nutzer sehr leicht zugänglich ist.

Das **Interface Auswerten** wird ausschließlich durch Graphen realisiert. Dieses Interface ist benutzerfreundlich gestaltet und verzichtet auf vorgefertigte Diagrammtypen. Allerdings lassen sich beispielsweise keine Funktionen in Graphen implementieren, so dass die Nutzung dieses Interfaces etwas eingeschränkt ist.

Mögliche Fehlerquellen

In Kapitel 3.4 wurden 9 Bereiche für funktionale Einheiten zur Simulation in TinkerPlots definiert. In den einzelnen Bereichen können sich mögliche Fehlerquellen ergeben.

Bereich 1a und 1b: Bauteile

Für die korrekte Modellierung einer Problemsituation kann die Wahl des richtigen Bauteils entscheidend sein. Da beispielsweise das Bauteil Kreisel die Einstellung ‚ohne Zurücklegen' nicht erlaubt, ist es für entsprechende Zufallsexperimente nicht geeignet. Allerdings kann trotzdem eine funktionierende Simulation mit diesem Bauteil erstellt werden, die aber möglicherweise nicht zur gegebenen Situation passt. Eine weitere Fehlerquelle ist das Bauteil Kurve, das den Anschein eines „stetigen" Bauteils hat, jedoch ebenso wie die anderen nur diskrete Werte erzeugt. Auch hierin kann eine Fehlerquelle liegen.

Bereich 2: Ziehungen (Merkmale) und *Bereich 3: Durchgänge (Wiederholung)*

Die eingestellte Ziehungsanzahl bestimmt in gewisser Weise die Simulationsart. Es ist jedoch möglich, dass je nach gegebener Situation ein semantischer Konflikt auftritt, wie

die Ziehungs- bzw. Durchgangsanzahl zu interpretieren ist. Vor allem bei sequentiellen Simulationen und bei Wartezeitproblemen birgt die Ziehungsanzahl Fehlerquellen.

Bereich 4: Steuerungselemente

Für diesen Bereich sind keine Fehlerquellen identifiziert worden. Im Gegenteil kann das Einstellen der Geschwindigkeit zum Verstehensprozess einer Simulation beitragen.

Bereich 5a und 5b: Ergebnistabelle

Je nach Simulationsart und nach Anzahl der Teilexperimente müssen hier verschiedene Merkmale identifiziert werden. Wenn sich auf die automatische Namensvergabe für Merkmale verlassen wird, so ist eine Zuordnung einzelner Merkmale möglicherweise schwierig, weil diese nur fortlaufend nummeriert werden. Ggf. ist ein Umbenennung der Merkmale hier sinnvoll. Eine weitere Fehlerquelle in diesem Bereich sind die zu definierenden Ereignisse und Zufallsgrößen, sofern sie über Formeln definiert werden. Das Benutzen der verschiedenen Formeln mit ihrer Syntax kann eine Fehlerquelle darstellen.

Bereich 6: Graph (Auswerten)

Das Darstellen des benötigten Merkmals kann eine Fehlerquelle sein, wenn im Bereich zuvor (oder in Bereich 2) die Merkmale nicht umbenannt wurden und hier das für eine Situation falsche Merkmal ausgewählt und dargestellt wird. Eine weitere potentielle Fehlerquelle besteht hier in der Philosophie dieses Objekts, das das Erzeugen von Diagrammen mittels der drei Operationen Trennen, Stapeln und Ordnen erlaubt, dem Nutzer allerdings keinen Hinweis gibt, wann eine „fertige“ Graphik erstellt wurde. Mitunter können so „unsinnige“, d. h. nicht interpretierbare Graphiken erstellt werden.

Bereich 7: Messgrößensammeln

Beim Messgrößensammeln sind wenig Fehlerquellen vorhanden, allenfalls kann ein falscher Wert als zu sammelnde Messgröße definiert werden.

Bereich 8: Anzahl Messgrößen

Dieser Bereich birgt eine Fehlerquelle, wenn eine zu kleine Anzahl an Messgrößen gesammelt wird, um eine reliable Aussage über die zu schätzende Größe zu tätigen. Desweiteren muss beachtet werden, dass TinkerPlots die in einem Vorgang zu sammelnden Messgrößen automatisch auf 5000 begrenzt, sofern eine größere Zahl eingegeben wird.

Bereich 9: Messgrößen auswerten

Für diesen Bereich gelten die gleichen Fehlerquellen wie für Bereich 6.

Generell stellt die Verknüpfung der stochastischen Inhalte mit den Objekten in TinkerPlots und den simulierten Daten eine mögliche Fehlerquelle dar. Dies kann sich auf al-

len Ebenen wiederfinden, bei der Modellierung eines Zufallsexperiments, bei der Realisierung von Ereignissen und Zufallsgrößen bis hin zur Interpretation der Ergebnisse.

4.5 Fazit

Es hat sich gezeigt, dass TinkerPlots recht umfassende Möglichkeiten für die Simulation von Zufallsexperimenten bietet. Auf vielfältige Weisen können einstufige, mehrstufige abhängige und mehrstufige unabhängige Zufallsexperimente simuliert werden, wobei die Simulationen je nach Art mehr oder weniger intuitiv bezüglich der verschiedenen Simulationsarten gestaltet werden können. Zur Unterstützung der Simulation kann der Simulationsplan aus Kapitel 2.4 genutzt werden, in welchem die einzelnen Bereiche mit Zahlen aufgenommen sind, und somit alle wichtigen Schritte einer Simulation abgebildet sind, um die kognitive Last des Nutzers zu verringern.

Zusammenfassen lässt sich das Unterstützungspotential in drei wesentlichen Punkten:

1. TinkerPlots ist eine Software zur Modellierung und Repräsentation von stochastischen Zufallsexperimenten. Die Zufallsmaschine erlaubt in vielen Fällen eine Visualisierung kognitiver Modelle, ohne dass Formeln verwendet werden müssen.
2. TinkerPlots ist ein Werkzeug zur Simulation. Hiermit können Wahrscheinlichkeiten geschätzt werden umd Wahrscheinlichkeitsverteilungen können durch empirische Verteilungen angenähert werden.
3. Mithilfe eines Modells kann TinkerPlots zum Experimentieren genutzt werden. Das visuelle Abbild eines Modells der Realität oder eines mathematischen Modells lässt vielfältige Explorationen zu.

Grenzen des Werkzeugs ergeben sich bei der Simulation mehrstufiger Zufallsexpimente mit vielen abhängigen Teilexperimenten, bei der Simulation von Markowketten und bei Simulationen, die verschachtelte Formeln zur Definition von Ereignissen benötigen.

Durch das eigenständige Erstellen von Simulationen mit TinkerPlots sollen aus didaktischer Sicht das Verständnis für das Modellieren von stochastischen Situationen gefördert werden, tiefere Einsichten bezüglich Ereignissen und Zufallsgrößen erlangt werden und simulierte Ergebnisse besser interpretiert werden. Hierzu ist jedoch eine Beherrschung des Werkzeugs als Instrument durch den Nutzer nötig. Die Fallstudie der nächsten Kapitel beleuchtet unter anderem die Anwendung der Software TinkerPlots durch Studierende zur stochastischen Simulation bei Randomisierungstests.

Teil II: Statistisches Schließen mit Randomisierungstests: Lerneinheit und Explorative Fallstudie zur Durchführung von Randomisierungstests mit TinkerPlots

5 Grundlagen zu Teil II: Inferenzstatistik und Randomisierungstests

Das Analysieren von Daten, um Schlussfolgerungen zu ziehen und Entscheidungen zu treffen, liegt im Kern jeder statistischen Untersuchung. Solche Methoden gehören zur Inferenzstatistik, die auch beurteilende bzw. schließende Statistik genannt wird. Es können (statistische) Hypothesen überprüft werden und bestimmte Fragen können datenbasiert beantwortet werden. Solche Fragen, die sich auf eine bestimmte Art von Unsicherheit beziehen, nämlich verursacht durch Daten von Zufallsstichproben und nicht von ganzen Populationen, Prozessen oder Verteilungen. Es geht also darum, über vorliegende Daten hinaus Schlussfolgerungen zu ziehen, immer unter Bedacht, dass die so gezogenen Schlüsse mit Unsicherheit behaftet sind, wie das folgende Zitat verdeutlicht.

> „Statistical inference moves beyond the data in hand to draw conclusions about some wider universe, taking into account that variation is everywhere and the conclusions are uncertain.“ (Moore 2007, xxviii)

Biehler und Engel (2015, 221f) formulieren in der Einleitung ihres Kapitels im Handbuch der Mathematikdidaktik, dass

> „… Schülerinnen und Schüler darauf vorbereitet werden [sollen], als verantwortungsbewusste Staatsbürger risikobehaftete Situationen rational zu analysieren und an Entscheidungsprozessen unter Unsicherheit teilzunehmen.“

Es ist jedoch nicht nur für Schülerinnen und Schüler wichtig, Entscheidungen unter Unsicherheit verstehen und treffen zu können, sondern allgemein für „verantwortungsbewusste Staatsbürger“, die beinahe täglich mit datenbasierten Studien und Entscheidungen konfrontiert werden. In diesem Sinne als verantwortungsbewussten Staatsbürger sollen unter anderem auch Lehramtsstudierende der Primarstufe befähigt werden, die die Zielgruppe der Studie in Kapitel 9 darstellen.[27] Inferenzstatistik lässt sich in der wissenschaftlichen Diskussion unterteilen in „klassische“ Inferenz und in Bayes-Inferenz. Die klassische Inferenz beruht auf einem objektivistischen Wahrscheinlichkeitsbegriff, wohingegen die Bayes-Inferenz auf einem subjektivistischen Wahrscheinlichkeitsbegriff beruht (Wickmann 1990). In der vorliegenden Arbeit wird die klassische Inferenz betrachtet, die vor allem von Sir Ronald Aymer Fisher (z. B. Fisher 1935) und von der durch Fishers Arbeiten angeregten Diskussion und Weiterentwicklung durch Neyman

---

[27] Genau diese Kompetenz im Bereich der Zivilstatistik (Engel 2017, Ridgway 2016) zu fördern hat das EU-Projekt ProCivicStat[27] zum Ziel erklärt, in welchem die Autorin dieser Arbeit mitarbeitet und welches die vorliegende Arbeit mit inspiriert hat. ProCivicStat steht für „Promoting Civic Engagement via Exploration of Evidence: Challenges for Statistics Education“ und ist eine durch die Europäische Kommission geförderte strategische Partnerschaft der Universitäten Durham, Haifa, Ludwigsburg, Paderborn, Porto und Szeged. Weitere Informationen finden sich unter http://www.procivicstat.org (abgerufen am 29.06.2018).

und Pearson (z. B. Neyman & Pearson 1933) geprägt wurde (für eine Diskussion der beiden Ansätze siehe z. B. Gigerenzer et al. 1999, ab S. 113). Mit Hilfe von Verfahren der Inferenzstatistik wird versucht, Rückschlüsse auf Wahrscheinlichkeitsverteilungen oder datengenerierende Prozesse (beispielsweise das Würfelwerfen, vgl. z. B. Konold & Kazak 2008) zu ziehen. Garfield und Ben-Zvi (2008a) unterscheiden Fragen der Inferenzstatistik in zwei Teilbereiche. Erstens Fragen bezüglich Verallgemeinerungen (aus Stichproben) von in Daten gefundenen Beobachtungen und zweitens Fragen bezüglich einer Ursache (bei randomisierten Experimenten). Letzteres fordern zum Beispiel auch Wild und Pfannkuch (1999), nämlich dass im Statistikunterricht nicht nur nach Verallgemeinerungen in Bezug auf Populationen sondern auch in Bezug auf Ursachen gefragt werden sollte.

> „Statistics education should really be telling students something every scientist knows, 'The quest for causes is the most important game in town.' It should be saying: 'Here is how statistics helps you in that quest'." (Wild & Pfannkuch 1999, 238).

In diesem grundlegenden Artikel beschreiben Wild & Pfannkuch Inferenzstatistik als einen wichtigen Teil des komplexen Bildes, das sie unter dem Schlagwort „Statistical Thinking" entwickeln.

5.1 Statistical Literacy, Statistical Reasoning und Statistical Thinking

Inferenzstatistik ist eine Komponente von Statistical Literacy, ein Zusammenschluss von Kompetenzen, die "verantwortungsbewusste Staatsbürger" (Biehler & Engel 2015) anstreben sollten, um mit statistischen Daten, ihren Darstellungen und Interpretationen kompetent umgehen zu können. Das dahinter liegende Konzept ist umfassend und in der Statistikdidaktik breit diskutiert (z. B. Gal 2002, Schield 1999, Sproesser 2015, Wallman 1993). Bereits 1993 argumentiert Wallman: „statistical literacy is the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions." (Wallman 1993, 1). Gal verbindet Statistical Literacy mit Inferenzstatistik und nennt dies „statistical reasoning: Knowing how statistical conclusions or inferences are reached" (Gal 2002, 10) und beschreibt dies als eine der Wissensgrundlagen für Statistical Literacy. Als Abgrenzung von Statistical Literacy, Statistical Reasoning und Statistical Thinking findet sich folgende Beschreibung

> „Statistical literacy includes basic and important skills that may be used in understanding statistical information or research results. (…) Statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information. (…) Statistical thinking involves an understanding of the nature of sampling, how we make inferences from samples to populations, and why designed experiments are needed in order to establish causation." (Ben-Zvi & Garfield 2004, 7)

Für diese Autoren stellt Statistical Thinking das höchste zu erreichende Konzept dar. In diesem Sinn setzen sich Wild & Pfannkuch mit Statistical Thinking auseinander und

gehen detailliert der Frage nach, was dieses „statistische Denken“ ausmacht. Sie erarbeiten ein umfassendes und viel beachtetes Modell zu Statistical Thinking, bei dem Lernende lernen sollen, statistische Untersuchungen nach dem Vorbild der Arbeitsweise professioneller Statistiker durchzuführen. Sie formulieren, dass „applied statistics is part of the information gathering and learning process which, in an ideal world, is undertaken to inform decisions and actions” (Wild & Pfannkuch 1999, 225). Die Autoren stellen ein vierdimensionales Modell zum statistischen Denken vor, das seitdem vielbeachtet in der Statistikausbildung ist. Die erste Dimension dieses Modells ist „The Investigative Cycle“, der unter dem Namen PPDAC-Cycle bekannt geworden ist. Die zweite Dimension des Frameworks umfasst „Types of Thinking“, die dritte Dimension ist „The Interogative Cycle“ und die vierte Dimension sind „Dispositions“. Ein „Thinker“ im statistischen Denken befindet sich in allen diesen Dimensionen zugleich (Wild & Pfannkuch 1999, 225). Für die vorliegende Arbeit ist vor allem Dimension Eins, der berühmt gewordene „Investigative Cycle“ von Interesse, der den Ablauf einer statistischen Untersuchung abbildet. Das statistische Denken in einer statistischen Untersuchung bezieht sich auf das Lernen und Treffen von Entscheidungen. Der PPDAC-Cycle besteht aus den fünf Komponenten *Problem, Plan, Data, Analysis,* und *Conclusions* (vgl. Abbildung 151), die zyklisch durchlaufen werden.

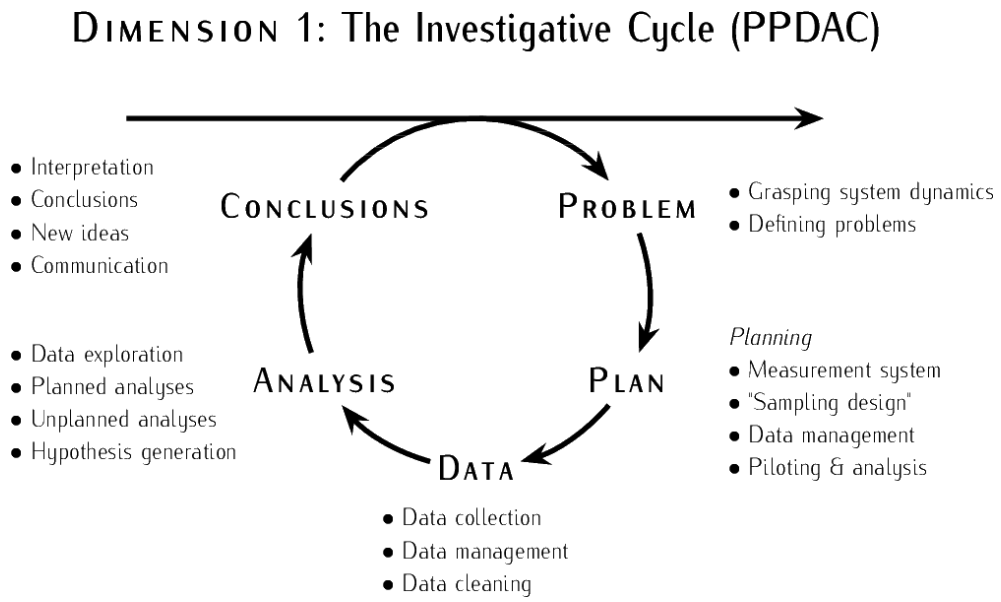


Abbildung 151: The Investigative Cycle (PPDAC) von Wild und Pfannkuch (1999), nachgebildet nach Wild und Pfannkuch (1999, 226)

In der ersten Phase *Problem* werden statistische Fragen und Hypothesen generiert und das zu untersuchende Problem wird hier definiert. In der zweiten Phase *Plan* wird die Datenerhebung geplant und das zugehörige (Versuchs-)Design festgelegt. Zur dritten Phase *Data* gehört das Erheben der Daten mit sich anschließendem Aufbewahren und Bereinigen der Daten. In der vierten Phase *Analysis* werden die Daten sowohl nach ei-

nem vorher festgelegten Plan geplant als auch ungeplant untersucht und in der fünften Phase *Conclusions* schließlich werden die Daten interpretiert und Schlussfolgerungen gezogen. Hierzu gehört auch das Kommunizieren der Ergebnisse und ggf. neuer Ideen. Alle fünf Phasen gehören gleichermaßen zu einer statistischen Untersuchung und sollten von Anwendern verstanden werden. Ähnliche Beschreibungen eines Untersuchungszyklusses finden sich sowohl in den deutschen Bildungsstandards (Kultusministerkonferenz 2003, 12, Kultusministerkonferenz 2012, 21) als auch in den Empfehlungen des Arbeitskreises Stochastik der Gesellschaft für Didaktik der Mathematik (Arbeitskreis Stochastik 2003, 22). Vor allem betonen die Autoren (Wild & Pfannkuch 1999, 227f) die Bedeutung von fünf fundamentalen (Denk-)Konzepten für Statistical Thinking. *Recognition for the need for data* (als Grundlage, um Entscheidungen treffen zu können), *transnumeration* (als Fähigkeit, die Darstellung von Daten zu wechseln bzw. zu transformieren für ein besseres Verständnis), *consideration of variation*, *reasoning with statistical models* und *integrating the statistical and contextual* (als Einbezug des Kontexts, aus dem die Daten stammen). Sie stellen fest, dass bei Anwendern, also bei statistisch Denkenden, unabhängig von Vorwissen die Suche nach Ursachen für beobachtete Phänomene häufig im Vordergrund steht. Das Nutzen von geeigneten Modellen, um Ursachen nachzuspüren, steht vor allem in der zweiten Dimension des Frameworks im Vordergrund bei *reasoning with models*.

> „From these models, we make inferences. We assume that the data has been randomly generated according to the model and use probability as the link between population/process and data. This is the very heart of the statistics we teach." (Wild & Pfannkuch 1999, 241)

Die so entwickelte Inferenzstatistik als ein Teil des Untersuchungsprozesses hat im Kern die Frage, wie aus vorliegenden Daten auf zugrundliegende Wahrscheinlichkeitsmodelle oder -prozesse oder deren Annahmen geschlossen werden kann (vgl. Burrill & Biehler 2011). Modelle in diesem Sinne von Statistical Literacy bzw. Statistical Thinking werden nach Shaughnessy (2007) sowohl von Schülern als auch von Erwachsenen benötigt, um ein Mindestmaß an Wissen und Fähigkeiten im Bereich Statistik zu erlangen, um als informierte Bürger datenbasierte Entscheidungen treffen zu können. Eine gute Zusammenfassung des „statistischen Denkens" bietet die folgende: „Das statistische Denken bezieht sich auf das Lernen und Treffen von Entscheidungen auf der Basis von variierenden Daten, die Unsicherheit implizieren." (Biehler & Engel 2015, 224).

5.2 Informelle Inferenzstatistik

Inferenzstatistik besteht üblicherweise aus Verfahren von miteinander verwobenen Konzepten, zu denen Wahrscheinlichkeit, (Zufalls-)Stichprobe, Teststatistik, Referenzverteilung, Null- und Alternativhypothese, P-Werte, Signifikanzniveau und die Logik des Schlussfolgerns gehören (Harradine et al. 2011, Liu & Thompson 2009). „Inference is a foundational area in statistics, and learning and teaching about inference is a key

concern of statistics education“ (Pratt & Ainley 2008, 3). Allerdings sind viele Schwierigkeiten in Bezug auf das Lernen und Verwenden von inferenzstatistischen Methoden bzw. den zugehörigen Konzepten bekannt (z. B.; Batanero 2000, Castro Sotos, Vanhoof, Van den Noortgate & Onghena 2007, Cumming, Williams & Fidler 2004, Harradine et al. 2011, Vallecillos 1999). Haller und Krauss (2002) beispielsweise berichten darüber, dass auch Experten Schwierigkeiten haben, Verfahren der Inferenzstatistik korrekt zu verstehen oder anzuwenden. Deshalb wird seit einigen Jahren als Vorstufe zu formaler Inferenzstatistik das informelle Schlussfolgern (informal inferential reasoning, siehe z. B. Makar & Rubin 2009, Pratt & Ainley 2008, Rossman 2008), als eine Herangehensweise genutzt, um leichter in Inferenzstatistik einzuführen. Dabei kann ‚informell‘ auf verschiedene Ebenen abzielen, wie folgende Beschreibung zeigt. „What is informal could depend on the nature of the inferential tasks being studied, on the complexity of the statistical or probabilistic concepts involved, on the educational stage, and on other factors.” (Pratt & Ainley 2008, 3).

Ein Ansatz, der drei Komponenten des informellen statistischen Schlussfolgerns zeigt, ist in Abbildung 152 zu sehen. Die erste Komponente dieses Modells beinhaltet das Erkennen der Unsicherheit, die einer statistischen Schlussfolgerung innewohnt. Makar & Rubin formulieren hierzu, dass das verallgemeinernde Schlussfolgern aus Daten das eigentliche Ziel ist, wobei sie im Sinne von Friel, Curcio und Bright (2001) die Stufe „beyond the data“, also Erkenntnisse, die über die vorliegenden Daten hinausgehen, anstreben. Dies ist die zweite Komponente ihres Modells. Die dritte Komponente entspricht der Forderung nach der expliziten Quantifizierung einer gefundenen Wahrscheinlichkeit zum Beispiel mit Hilfe des P-Werts, wie sie z. B. auch Biehler und Engel (2015) vorschlagen und wie sie bereits von Fisher (1935) eingeführt wurde.

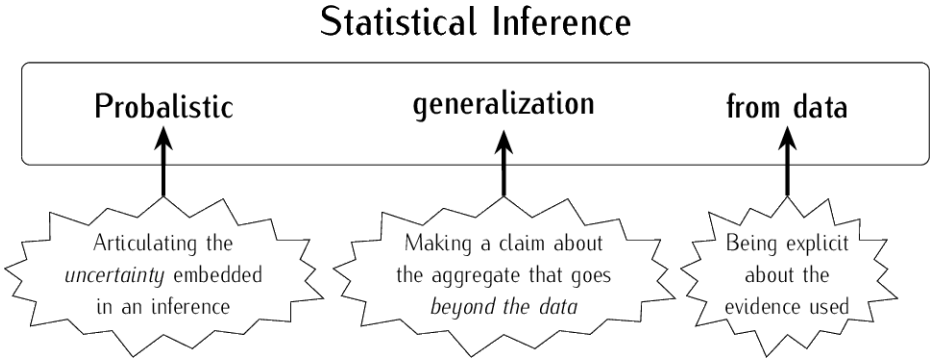


Abbildung 152: Modell für informelles Schlussfolgern von Makar und Rubin (2009), nachgebildet nach Makar und Rubin (2009, 85)

Im Kontext des „informellen Hypothesentestens“ als einem Verfahren der (informellen) beurteilenden Statistik formulieren Biehler & Engel folgendes Ziel: „Es geht beim informellen Schließen weder um eindeutige Ja-nein-Schlüsse noch um eine exakte Berechnung von Fehlerwahrscheinlichkeiten innerhalb eines probabilistischen Modells,

sondern um die Heranbildung von Intuitionen und Heuristiken, eine beobachtete Abweichung im Kontext vergangener Abweichungen zu beurteilen.“ (Biehler & Engel 2015, 241). Gigerenzer definiert dabei Heuristik als „eine Regel, die den Prozess – nicht nur das Ergebnis – einer Problemlösung beschreibt“ (Gigerenzer & Gaissmaier 2006, 332), wobei auf bereits erlernte Fähigkeiten zurückgegriffen wird und im Sinne des Kontexts bestimmte Sachverhältnisse (Umweltstrukturen) vereinfacht werden. Hierauf baut informelles Schlussfolgern auf, so dass ein Ziel ist, dass nicht alle zugehörigen Konzepte der Inferenzstatistik (wie oben aufgezählt) auf einmal verstanden werden, sondern Ideen von einzelnen Konzepten entwickelt werden (vgl. Rubin, Hammerman & Konold 2006, Zieffler, Garfield, DelMas & Reading 2008) und erst später formal durchdrungen werden. Es ergibt sich ein Kontinuum zwischen formeller und informeller Inferenzstatistik, je nachdem, welche und wie Konzepte des Schlussfolgerns formal verwendet werden (Makar & Rubin 2017). Es geht auch weniger um die Frage, ob oder ob nicht ein bestimmtes Verfahren als informell definiert wird, sondern darum, zu untersuchen, „what range of approaches are most beneficial for helping students appreciate and master the power of statistical inference.“ (Makar & Rubin 2017, 275). Letztlich kann als Ziel aller hier genannten Autoren interpretiert werden, dass der Fokus beim Ansatz des informellen Schlussfolgerns auf dem Verstehen des Prozesses der beurteilenden Statistik liegt und weniger auf dem rezeptartigen Durchführen einer Methode.

Ein diskutierter Ansatz des informellen Schließens ist es, über P-Werte in das Hypothesentesten einzusteigen. Dies ist in den angelsächsischen Ländern verbreitet, in Deutschland gibt es einige wenige Ansätze dazu (Lergenmüller, Schmidt, Krüger, Biehler & Vehling 2012, Meyfarth 2006, Oesterhaus & Biehler 2013, Podworny & Biehler 2014). Ein typischer Ansatz des informellen Schließens ist es, formale Berechnungen von Wahrscheinlichkeiten durch aus Simulationen geschätzte Wahrscheinlichkeiten zu ersetzen (z. B. Garfield, delMas & Zieffler 2012, Lock et al. 2013). Harradine et al. (2011) sehen in Randomisierungstests den Übergang von informellem Schlussfolgern zu formellem Schlussfolgern, und nach einer ausführlichen Analyse von Schwierigkeiten in der Inferenzstatistik fordert Ridgway für die schulische Ausbildung sogar, dass „hypothesis testing should be replaced with randomisation tests“ (Ridgway 2016, 546).

5.3 Resamplingverfahren als Einstieg in Inferenzstatistik

Ein junger Ansatz, um in die Inferenzstatistik einzusteigen, ist die Nutzung von Resamplingverfahren als nichtparametrische Verfahren. Dazu gehören Randomisierungstests (Permutationstests) und Bootstrapping, die als computergestützte Methoden auf der Verfügbarkeit immer leistungsfähigerer Computer beruhen. Diese Methoden finden immer mehr Verwendung in allen Feldern, in denen Statistik angewendet wird (Engel 2007, Hesterberg, Moore, Monaghan, Clipson & Epstein 2009, Moore 2007).

In der didaktischen Literatur gibt es verschiedene Gruppen von Forschern, die große Vorteile darin sehen, über Randomisierungstests oder Bootstrapping in Verbindung mit Simulationen die Logik des Schlussfolgerns einzuführen. Hierzu einige Stimmen:

> „I suggest that simulation of randomization tests provides an informal and effective way to introduce students to the logic of statistical inference." (Rossman 2008, 17)

> „Er [der Bootstrap] ist ein Instrument, um grundlegende Ideen der schlussfolgernden Statistik zu erkunden, zu visualisieren und konkreter fassbar zu machen." (Engel & Grübel 2008, 3)

> „The randomization method offer simplicity and direct access to the logic of inferential reasoning. Furthermore, they open up an opportunity to give all students access to understanding inference." (Pfannkuch et al. 2011, 903)

> „The new learning emphasis should be on the reasoning and logic underpinning inference or hypothesis testing and that the mathematical procedures and manipulation of symbols of normal-based hypothesis testing should be replaced with the randomization method, which seems to be more conceptually accessible both visually and verbally (Budgett et al. 2012, 3)

Gerade in Verbindung mit dem P-Wert können Randomisierungstests von Vorteil sein:

> „We believe that this simulation of a randomization test can lead students to a deeper understanding of the concepts of statistical significance and p-value than, say, conducting t-tests." (Chance & Rossman 2006, 5)

Eine ganze Reihe von Vorteilen aus didaktischer Sicht zur Einführung in die beurteilende Statistik über Randomisierungstests finden sich bei Holcomb, Chance, Rossman, Tietjen und Cobb (2010). Davon sei ein Argument herausgegriffen:

> „These activities [simulation based randomization tests] require virtually no prerequisite knowledge, so students can engage in them from the very beginning of the course. With such an important and difficult concept as statistical significance, we argue that an early experience, with frequent follow-up, allows a similar reasoning process to be reinforced over and over in new settings." (Holcomb, Chance, Rossman, Tietjen, et al. 2010, 3)

Watson und Moritz (1999) sowie Makar und Confrey (2002) schlagen vor, nach dem Vergleich von Verteilungen Randomisierungstests zu thematisieren, um Lernenden eine Möglichkeit zu geben, ihre in der jeweiligen Stichprobe gefundenen Unterschiede zu verallgemeinern. Dies findet sich auch bei Madden (2011), die einen Vorschlag macht, inhaltlich an Gruppenvergleiche anzuknüpfen und über Randomisierungstests in *informal inferential reasoning* einzuführen, um kausale Schlussfolgerungen ziehen zu können, was auch von Pfannkuch et al. (2015) als Hauptargument für die Verwendung von Randomisierungstest gesehen wird.

Durch die verschiedenen Anregungen erfolgt im nächsten Abschnitt ein genauerer Blick auf das Verfahren des Randomisierungstests.

5.4 Der Randomisierungstest als ein Resamplingverfahren

Unter Resampling versteht man (heutzutage) computerbasierte Verfahren zur statistischen Datenanalyse, die vor allem für die Untersuchung von Datensätzen verwendet werden, deren Verteilungseigenschaften nicht bekannt sind. Es handelt sich also um so-

genannte verteilungsfreie Testverfahren (z. B. Götz 1993). Unbekannte Verteilungseigenschaften können beispielsweise bei kleinen Stichproben, abhängigen Daten oder bei komplexen statistischen Problemen vorliegen (Reimer 2009).

Randomisierungstests gehen zurück auf Fisher (1925) und Pitman (1937). Beim Randomisierungstesten soll die exakte Verteilung einer beliebigen Teststatistik, die unter einer bestimmten Nullhypothese spezifiziert wird, erhalten werden, indem die Originaldaten sehr häufig neu gemischt werden. Fishers exakter Test ist ein Beispiel eines Randomisierungstests, dieser beruht auf der Analyse von Daten in Kontingenztabellen und erzeugt alle möglichen Ergebnisse (Fisher 1935, 12). Pitman hat gezeigt, dass der Randomisierungstest über solche Daten hinaus auf beliebige Daten erweitert werden kann und folgert „but the essential point of the method [test for significance of difference of means] is that we do not have to worry about the populations which we do not know, but only about the sample values which we do know.” (Pitman 1937, 129).

Die Grundidee des Randomisierungstests nach Fisher bzw. Pitman ist, dass alle denkbaren Möglichkeiten durchgespielt werden und die kumulierte Wahrscheinlichkeit bestimmt wird, mit der ein Ergebnis, das ebenso extrem wie oder noch extremer als ein Beobachtetes ist, auftritt. Der Vorteil dieses Verfahrens ist, dass es unerheblich ist, wie die Versuchseinheiten gewonnen wurden (Edgington 1995, Ludbrock & Dudley 1998). Beim Durchführen des tatsächlichen Experiments werden die Versuchseinheiten der verschiedenen Gruppen dann unterschiedlichen Bedingungen oder Behandlungen (treatments) ausgesetzt und bestimmte Kennzahlen gemessen. Die Gruppengrößen müssen dabei nicht gleich sein (ebd.).

Hierzu ein Beispiel (angelehnt an Lock et al. 2013, 240, eigene Übersetzung), das das Verfahren des Randomisierungstests aufzeigen soll, so wie es der Studie in Kapitel 9 zugrunde liegt.

“Finger-Klopfen und Koffein": Für ein Experiment wählten Forscher 20 männliche College-Studenten aus, die ihre Finger mit einer hohen Geschwindigkeit auftippen sollten. Diese 20 Personen wurden zufällig auf zwei Gruppen von jeweils 10 Personen aufgeteilt. Jeder Teilnehmer trank zwei Tassen Kaffee. Für die Teilnehmer der ersten Gruppe enthielt der Kaffee etwa 200 mg Koffein, die Teilnehmer der zweiten Gruppe erhielten entkoffeinierten Kaffee. Nach einer Zeit von zwei Stunden wurde von jedem Teilnehmer erneut die Fingertipprate gemessen (in Tipps pro Minute). Die Teilnehmer wussten nicht, ob ihre Getränke Koffein enthielten oder nicht, ebenso wenig wie der Versuchsleiter. Diese Informationen hatte nur der das Experiment begleitende Statistiker. Das Ziel des Experiments war es herauszufinden, ob Koffein eine Erhöhung der durchschnittlichen Tipprate bewirkt. Die in diesem Experiment gemessenen Fingertipprraten sind in Tabelle 1 zusammengefasst.

Tabelle 1:Fingertippraten von jeweils 10 unterschiedlichen Versuchspersonen[28]

Koffein	*246*	*248*	*250*	*252*	*248*	*250*	*246*	*248*	*245*	*250*
Kein Koffein	*242*	*245*	*244*	*248*	*247*	*248*	*242*	*244*	*246*	*242*

Die Versuchseinheiten (im Beispiel: männliche College-Studenten) wurden auf zwei Gruppen N_1 und N_2 mit n_1 bzw. n_2 Versuchseinheiten aufgeteilt, wobei im Beispiel $n_1 = n_2 = 10$ gilt. Dieser Vorgang ist schematisch in Abbildung 153 dargestellt.

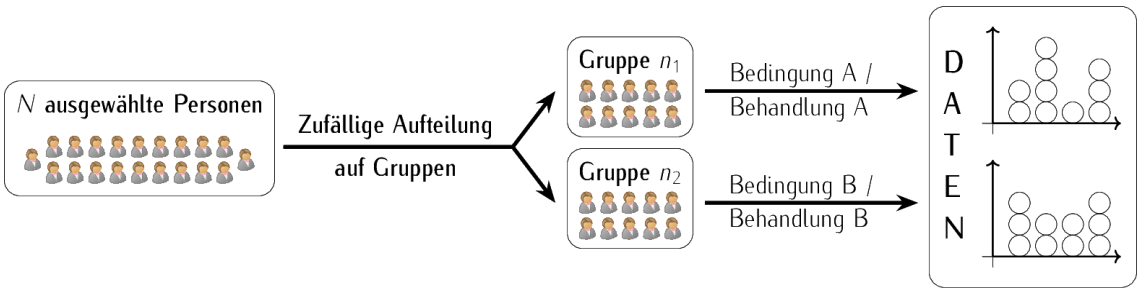


Abbildung 153: Schematische Darstellung eines Experiments mit zwei randomisierten Gruppen, angelehnt an Ramsey und Shafer (2013, 10)

Die Daten wurden in dem Experiment hinsichtlich ihres Durchschnitts ausgewertet. In der Gruppe N_1: *Koffein* wurde ein Mittelwert von 248,3 Fingertipps pro Minute erreicht, in der Gruppe N_2: *Kein Koffein* wurde ein Mittelwert von 244,8 Fingertipps pro Minute erlangt.

Um die Daten eines Experiments zu untersuchen, wird eine Teststatistik benötigt. Eine häufig verwendete Teststatistik T ist zum Beispiel die Differenz der Mittelwerte der beiden Gruppen N_1 und N_2:

$$T = \overline{N_1} - \overline{N_2}$$

Das wäre im obigen Beispiel

$$T_{beob} = \overline{N_1} - \overline{N_2} = 248{,}3 - 244{,}8 = 3{,}5,$$

also ein beobachteter Unterschied zwischen den Mittelwerten der beiden Gruppen von 3,5 Tipps pro Minute.

Für die Teststatistik eines Randomisierungstests können jedoch nicht nur die Unterschiede zwischen arithmetischen Mitteln wie z. B. bei t-Tests verwendet werden, sondern auch Unterschiede zwischen Medianen, geometrischen Mittelwerten, Anteilen,

[28] Die Daten gehen zurück auf Draper und Smith (1981, 425).

mittleren Bereichen, Rängen, Standardabweichungen oder ähnliche Statistiken (Edgington 1995, Ludbrock & Dudley 1998, Reimer 2009).

Es stellt sich nun die Frage, wie die beobachtete Differenz eines Experiments zu beurteilen ist. Hiermit einher geht die Frage, was die Erklärung für die beobachteten Unterschiede sein kann, vor allem, ob es einen kausalen Zusammenhang zwischen einer Behandlung (im Beispiel die Einnahme von Koffein) und dem Ausgang des Experiments gibt. Prinzipiell lassen sich drei mögliche Erklärungen für einen beobachteten Unterschied differenzieren.

1. Die beobachteten Unterschiede liegen an der durchgeführten Behandlung.
2. Die beobachteten Unterschiede liegen an unbekannten Variablen (konfundierenden Variablen), die einen systematischen Fehler verursacht haben.
3. Die beobachteten Unterschiede sind zufällig aufgetreten, liegen also an der zufälligen Einteilung der Versuchspersonen auf die Gruppen.

Die erste Ursache (Behandlung) als Erklärung für die beobachteten Unterschiede ist üblicherweise das (gewünschte) Ziel eines Experiments. Dies lässt sich jedoch mit einem statistischen Test nicht direkt untersuchen. Die zweite Erklärung (Störvariable) beinhaltet Merkmale der Versuchseinheiten, die nicht untersucht wurden, aber die Ergebnisse beeinflussen. So kann in einem bestimmten Experiment beispielsweise das Alter der Versuchsteilnehmer eine Rolle in Bezug auf die Ergebnisse spielen.[29] Wenn zuvor alle älteren Personen einer Gruppe zugewiesen und alle jüngeren der zweiten zugewiesen wurden, so wird das Ergebnis des Experiments dadurch verzerrt. Oder die Gruppen werden nach männlichen und weiblichen Teilnehmern aufgeteilt, oder nach noch anderen Prinzipien, die Einfluss auf das Ergebnis des Experiments haben, in der Untersuchungsfrage jedoch nicht berücksichtigt werden. Die Erklärung dieser konfundierenden Variablen wird jedoch durch die zufällige Aufteilung der Versuchseinheiten auf die Gruppen kontrolliert und kann somit weitgehend[30] ausgeschlossen werden. Aus diesem Grund ist die Randomisierung beim Versuchsaufbau das wichtigste Prinzip (vgl. z. B. auch Fisher 1935 oder Edgington 1995). Bleibt noch die dritte Erklärung, dass die beo-

[29] Dies ist die Beschreibung für eine Situation, in der alle Störfaktoren unbekannt sind. Es gibt auch Situationen, in denen bestimmte Einflussfaktoren bekannt sind. Hierfür bietet sich ein anderes Experimentdesign an (z. B. das Bilden von stratifizierten Stichproben im Hinblick auf bekannte Einflussfaktoren, z. B. Männer vs. Frauen (Saint-Mont 2011, 231)).

[30] Weitgehend und nicht komplett deshalb, weil je kleiner die Stichprobe ist, desto höher die Wahrscheinlichkeit ist, dass „ungünstige“ Gruppenaufteilungen nach einem nicht untersuchten Merkmal bei zufälliger Gruppeneinteilung auftreten. Bei bekannten Störvariablen kann zum Beispiel ein „matched pair design“ Abhilfe schaffen, um systematischen Fehlern vorzubeugen (Good 2005, 100). Der entscheidende Hinweis bezüglich unbekannter Einflussfaktoren lautet: „to keep the errors exchangeable we need to randomly assign experimental units to treatment so that the innumerable factors that can neither be controlled nor observed directly are as likely to influence the outcome of one treatment as another” (Good 2005, 103). Randomisierung ist also ein wichtiges Prinzip für das Design eines Experiments.

bachteten Unterschiede an zufällig auftretenden Schwankungen liegen, die durch die zufällige Gruppeneinteilung zum Vorschein kommen. Dies ist gerade die Verneinung der ersten Ursache, dass die beobachteten Unterschiede abhängig von der Behandlung sind.

Diese Annahme, dass die durchgeführte Behandlung keinen Unterschied bezüglich einer beobachteten Differenz der Teststatistik ausmacht, wird in der Nullhypothese formuliert. Edgington schreibt hierzu

> „The null hypothesis for a traditional randomization test is that the measurement for each person or other unit that is randomly assigned will be the same under one assignment to treatments as under any alternative assignment that could have resulted from the random assignment procedure. Thus, the nullhypothesis (…) is the hypothesis of no differential treatment effect (…).” (Edgington 1995, 2)

Wie hier und anderswo (z. B. Ernst 2004) zu sehen ist, lautet eine mögliche, allgemeine Formulierung der Nullhypothese

H_0: Es gibt keinen Unterschied zwischen den verschiedenen Behandlungen.

Oder anders ausgedrückt:

H_0: Die beobachteten Ergebnisse sind unabhängig von der Gruppenzugehörigkeit und somit von der Behandlung.

Im Beispiel des Koffein-Experiments lautet eine mögliche Formulierung der Nullhypothese

H_0: *Die Ergebnisse der Versuchspersonen sind unabhängig davon, ob zuvor Koffein eingenommen wurde oder nicht.*

Es wird also von der Annahme ausgegangen, dass die zufällige Gruppeneinteilung die Erklärung für die beobachteten Unterschiede ist. Für diese Annahme lässt sich die Wahrscheinlichkeit berechnen, dass ein Ergebnis wie das beobachtete oder sogar ein noch extremeres erreicht wird.

Für den eigentlichen Test werden dann alle möglichen Permutationen gebildet, wie die N Versuchseinheiten auf Gruppen der Größe n_1 und n_2 aufgeteilt werden können. Die Gruppengrößen des ursprünglichen Experiments werden dabei beibehalten, so dass es bei zwei Gruppen der Größen n_1 und $n_2 = N - n_1$ genau $K := \binom{N}{n_1}$ mögliche Kombinationen gibt.[31] Unter der Nullhypothese sind alle diese Kombinationen gleichwahrscheinlich und hätten genauso gut wie die beobachtete Kombination auftreten können, da die Versuchseinheiten zufällig auf die Gruppen aufgeteilt wurden.

[31] Es gilt $\binom{N}{n_1} = \binom{N}{n_2}$ wegen der Symmetrie des Binomialkoeffizienten.

Im obigen Beispiel des Fingerklopfens gibt es bei zwei Gruppen der Größe $n_1 = n_2 = 10$ genau $\binom{20}{10}$, also 184756 mögliche Kombinationen, wie die Versuchspersonen auf zwei Gruppen aufgeteilt werden können.

„Randomisierungstests ‚erzeugen' ihre Prüfverteilungen jeweils auf der Basis der konkret erhobenen Daten" (Bortz & Lienert 2008, 227). Für jede der möglichen Kombinationen wird die Teststatistik berechnet und in der Referenzverteilung zusammengetragen. (Andere Begriffe hierfür sind Permutationsverteilung (Reimer 2009), Randomisierungsverteilung (Ernst 2004) oder Prüfverteilung (Bortz & Lienert 2008)).

Nun lässt sich bestimmen, wie extrem der in den Originaldaten beobachtete Unterschied ist, unter Voraussetzung, dass die Nullhypothese gilt. Dies ist der P-Wert. Und zwar „ist der P-Wert des Permutationstests die Wahrscheinlichkeit für die Permutationen, die einen Wert ergeben, der mindestens so stark wie der beobachtete Wert gegen die Nullhypothese spricht" (Neuhäuser 2010, 6). Der P-Wert des Randomisierungstests in Bezug auf die Nullhypothese H_0 lässt sich berechnen als Laplace-Wahrscheinlichkeit mit

$$p = P(T \geq T_{beob})|H_0) = \frac{\sum_{i=1}^{K} I(t_i \geq T_{beob})}{K}.$$

Dabei wird für jede der möglichen, unter H_0 gleichwahrscheinlichen K $(= \binom{N}{n_1})$ Kombinationen betrachtet, ob die Differenz der Merkmalsmittelwerte (t_i) gleich oder größer dem beobachteten Wert ist. Die Indikatorfunktion I liefert 1, wenn dem so ist, ansonsten 0. Durch Summation der Indikatorwerte wird die Gesamtzahl der günstigen Fälle erhalten und durch Division der Anzahl aller möglichen Fälle lässt sich der P-Wert ermitteln.

Es ist klar, dass die Randomisierungsverteilung diskret ist und der P-Wert ein Vielfaches von $1/\binom{N}{n_1}$ sein muss. Jedoch ist nicht jedes Vielfache davon möglich (je nach der Beschaffenheit der Originaldaten). Möchte man nun mit Hilfe des P-Werts eine Entscheidung treffen (Good 2005), ob die Nullhypothese abgelehnt werden kann oder nicht, so folgt man einem Hybridansatz (vgl. hierzu auch Rossman 2008) aus dem Gedankengut von Fisher (der lediglich von „Evidenz" gegen die Nullhypothese spricht) und Neyman & Pearson (die bei einem genügend kleinen P-Wert, z. B. $p < \alpha = 0{,}05$, die Nullhypothese zugunsten einer Alternativhypothese verwerfen). Der Hypbridansatz zeigt sich zum Beispiel darin, eine Entscheidung aufgrund eines vorher festgelegten Niveaus α zu treffen, was aus dem Ansatz von Neyman und Pearson stammt. Wenn allerdings keine Schlussfolgerung gezogen wird, weil der erhaltene P-Wert nicht signifikant ist, so ist dies ein Ansatz von Fisher, der von Neyman und Pearson kritisiert wurde. Spricht man dann noch davon, dass eine Nullhypothese beispielsweise auf dem 0,05 Prozent Niveau verworfen werde, so lässt sich dies weder bei Fisher noch bei Neyman und Pearson wiederfinden. Dieses Zusammenbringen der verschiedenen Theorien wird von Gigerenzer et al. (1999, 130) „Theorientopf" genannt, der von Zeitschriftenheraus-

gebern und Universitäten „institutionalisiert“ worden sei. Auch in der vorliegenden Arbeit soll ein Hybridansatz verwendet werden zur Vereinfachung (vgl. dazu Kapitel 8.2), nachdem ab einem gewissen Niveau α die Nullhypothese abgelehnt werden soll.

Bei einem solchen Entscheidungsprozess können Fehler passieren. Es kann sein, dass man sich dafür entscheidet, die Nullhypothese abzulehnen, obwohl sie in Wirklichkeit wahr ist. Oder man entscheidet sich dafür, die Nullhypothese nicht abzulehnen, obwohl sie in Wirklichkeit falsch ist. Entscheidet man sich dafür, die Nullhypothese abzulehnen, obwohl sie in Wirklichkeit wahr ist, so begeht man den Fehler erster Art. Dieser Fehler lässt sich mit Hilfe eines vorab defininierten Signifikanzniveaus α kontrollieren. Ist der erhaltene P-Wert kleiner als das Signifikanzniveau α, so wird die Nullhypothese abgelehnt und man trifft höchstens mit der Wahrscheinlichkeit α eine falsche Entscheidung.

Lehnt man aufgrund eines kleinen P-Werts die Nullhypothese, also die Annahme, dass die beobachteten Unterschiede an der zufälligen Einteilung der Versuchseinheiten zu den Gruppen liegt, ab, so kann man aufgrund der vor dem Experiment erfolgten Randomisierung auf einen kausalen Zusammenhang zwischen der Behandlung und dem beobachteten Unterschied schließen.

Die Anzahl aller Kombinationen für die Neuzuordnung der Versuchseinheiten auf die Gruppen ist bereits bei kleinen Stichproben relativ groß. Schon für $N = 30$ Versuchseinheiten, die auf zwei gleich große Gruppen aufgeteilt werden, gibt es über 155 Millionen Kombinationen. „The amount of computation required for randomization tests made them impractical in the days when computation had to be done by hand or by mechanical desk calculator” schreibt Edgington (1995, 44), bezugnehmend u. a. auf Fisher und Pitman. Deshalb schlägt er vor, statt systematisch alle Kombinationen zu betrachten, zufällig vom Computer Kombinationen simulieren zu lassen und stellt dazu ein von ihm selbst erstelltes Programm in FORTRAN IV vor (Edgington 1995, 45ff). Auch an anderen Stellen (z. B. Ernst 2004, Good 2005 oder Zieffler et al. 2011) wird vorgeschlagen, statt den exakten P-Wert zu berechnen, ihn über Simulationen zu schätzen. Über die Anzahl der dazu nötigen Simulationsdurchgänge findet sich lediglich bei Ernst eine Angabe, nämlich dass „a few thousand test statistics from the permutation distribution usually are sufficient to get an accurate estimate of the exact *p* value” (Ernst 2004, 682).

Im Beispiel des Fingerklopfens führt eine Simulation[32] im Umfang von n = 5000 zu folgender Referenzverteilung (Abbildung 154).

[32] Ein bezüglich der Simulation ausführlich dargestelltes Beispiel findet sich in Kapitel 9.2.

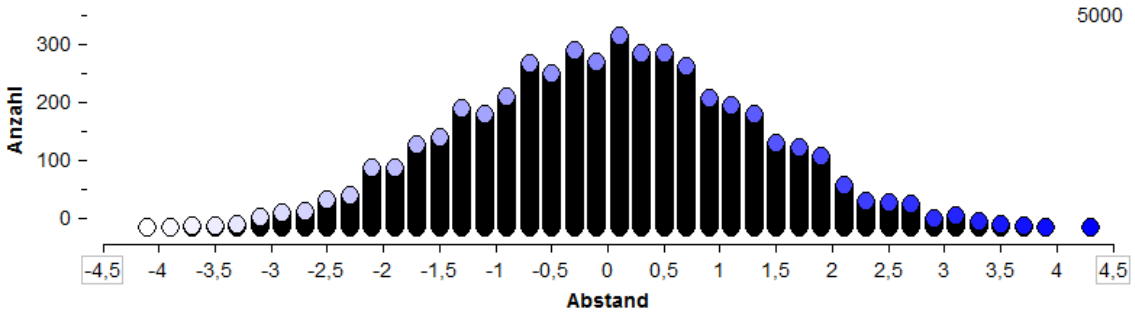


Abbildung 154: Referenzverteilung zum Fingerklopfexperiment

Wie in Abbildung 155 lässt sich die beobachtete Teststatistik einzeichnen und daraus der P-Wert schätzen.

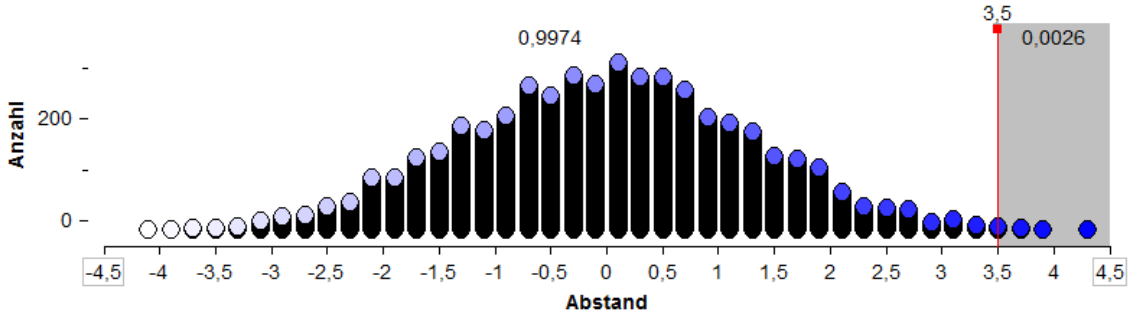


Abbildung 155: Referenzverteilung mit eingezeichneter beobachteter Teststatistik und geschätztem P-Wert

Der geschätzte P-Wert beträgt also bei 5000 Wiederholungen für das Fingerklopf-Experiment 0,0026, ist also kleiner als 1%. Somit liegt ein signifikantes Ergebnis vor und die Nullhypothese kann abgelehnt werden. Durch die zuvor erfolgte Randomisierung der Versuchsteilnehmer auf die beiden Gruppen lässt sich schließen, dass es einen kausalen Zusammenhang zwischen der Einnahme von Koffein und der Fingertipprate gibt.

Es stehen sich also die angenäherte Berechnung des P-Werts durch Simulation mit der exakten Berechnung beim Permutationstest gegenüber. Die Einführung in Randomisierungstests in der konkreten Lerneinheit (Kapitel 8.2), die im Rahmen der vorliegenden Arbeit entwickelt wird, soll über Simulationen und daraus geschätzten P-Werten geschehen, so wie es von Mathematikdidaktikern (u. a., Holcomb, Chance, Rossman, Tietjen, et al. 2010 Rossman 2008 und Wild, Pfannkuch, Regan & Horton 2011) empfohlen wird.

5.4.1 Experimental Research

Ein wichtiger Schritt im PPDAC-Zyklus von Wild und Pfannkuch (1999) ist das zweite „P“, das Planen des Designs der Untersuchung. Sofern ein Experiment aktiv geplant und durchgeführt wird, ist das Design sorgfältig zu wählen. Wenn mit bereits erhobenen Daten (wie es im Lehrkontext häufig der Fall ist) gearbeitet wird, so muss das Design des zugrundeliegenden Experiments genau geprüft werden.

Allgemein findet sich dazu, dass „A *randomization test* is a permutation test based on randomization (random assignment) to test a null hypothesis about treatment effects in a randomized experiment“ (Edgington 1995, 1). Weitere Anforderungen an ein Experiment stellt Edgington nicht, ein Experiment muss also randomisiert sein, das heißt es genügt, wenn die Versuchseinheiten zufällig auf die verschiedenen Gruppen aufgeteilt sind.

In der Praxis lässt sich unterscheiden zwischen Experimenten, für die eine Zufallsstichprobe aus einer Population gezogen wird (dies wird häufig als Populationsmodell bezeichnet) und Experimenten, in denen verfügbare Versuchseinheiten verwendet werden, ohne dass diese eine Zufallsstichprobe einer bestimmten Population darstellen; dies sind häufig Freiwillige, die an einem Experiment oder einer Untersuchung teilnehmen möchten. Für die letzteren Experimente findet sich als gängige Bezeichnung „experimental research“ (z. B. Zieffler et al. 2011). Im Hinblick auf eine mögliche zufällige Einteilung der Versuchseinheiten zu den Gruppen ergeben sich vier Szenarien, in denen Forschung betrieben wird. Sowohl bei Zieffler et al. (2011) als auch bei Ramsey und Shafer (2013) finden sich diese vier Szenarien übersichtlich dargestellt mit den möglichen Schlussfolgerungen, die aus dem jeweiligen Design gezogen werden können, wie in Abbildung 156 dargestellt.

Hat man eine echte Zufallsstichprobe vorliegen, deren Versuchseinheiten randomisiert, also zufällig auf verschiedene Gruppen aufgeteilt wurden, so lassen sich Schlussfolgerungen bezüglich der Population und bezüglich der Wirkung einer Behandlung/Maßnahme ziehen. Wenn eine der beiden Bedingungen (Zufallsstichprobe oder Randomisierung) verletzt ist, so lassen sich im Fall einer erfolgten Randomisierung Schlussfolgerungen bezüglich einer Ursache, aber nicht bezüglich einer Population ziehen und im Fall einer gezogenen Zufallsstichprobe genau anders herum. Hat man weder eine Zufallsstichprobe noch eine Randomisierung, so mag es sich um eine sogenannte „observational study“ handeln, aus der kaum Schlussfolgerungen möglich sind.

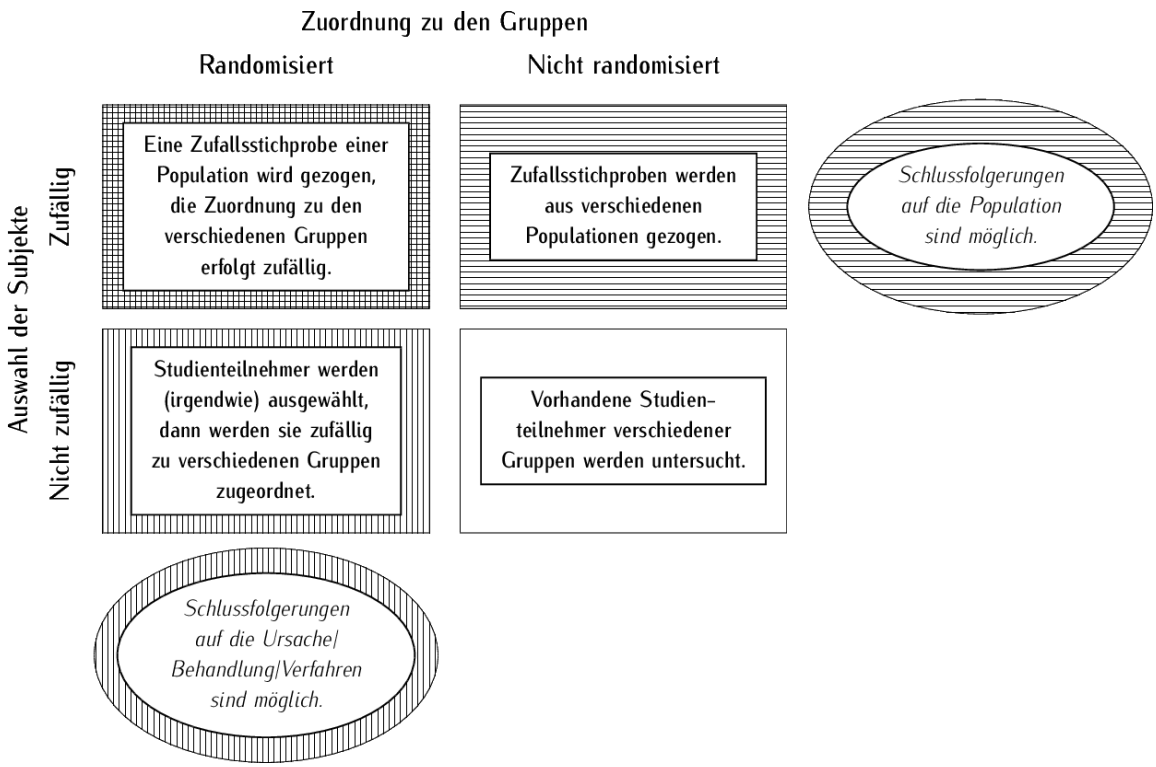


Abbildung 156: Design von Experimenten und mögliche Schlussfolgerungen, angelehnt an Ramsey und Shafer (2013, 9)

Echte Zufallsstichproben sind jedoch ein seltenes Design für Experimente, bei denen die Versuchseinheiten Menschen sind. So sehen zum Beispiel Ludbrock und Dudley (1998) eine Diskrepanz (vor allem in der Biomedizin) zwischen dem, was Statistiker denken, wie Experimente aufgebaut sind und dem, wie sie tatsächlich angelegt sind:

> „Statisticians appear to believe that biomedical researchers do most experiments by taking random samples, and therefore recommend statistical procedures that are valid under the population model of inference. In fact, randomization of a nonrandom sample, not random sampling, is more usual." (Ludbrock & Dudley 1998, 127)

Auch Edgington (1995) betont, dass in Experimenten in Biologie, Erziehung, Medizin oder Psychologie in den seltensten Fällen mit Zufallsstichproben gearbeitet würde, wesentlich verbreiteter sei der Gebrauch von nichtzufälligen Stichproben. Es findet sich also häufig ein Design, in welchem sogenannte experimentelle Forschung betrieben wird, also nichtzufällige Stichproben verwendet werden, deren Versuchseinheiten zufällig auf verschiedene Gruppen aufgeteilt werden. Die zu verwendende Testmethode sollte das Design der Studie widerspiegeln, in diesem Fall ist also der Randomisierungstest angemessen (Ernst 2004, Neuhäuser 2010, Rossman 2008).

In Bezug auf die Auswahl der Versuchseinheiten, die die Stichprobe bilden, wird dies konkretisiert:

„A randomization test is valid for any kind of sample, regardless of how the sample is selected. This is an extremely important property because the use of nonrandom samples is common in experimentation." (Edgington 1995, 7)

5.4.2 Mit welchen Stichprobengrößen kann im Randomisierungstest gearbeitet werden?

Um den Randomisierungstest vorzustellen, wird fast in jeder durchgesehenen Literatur an einer frühen Stelle ein Beispiel verwendet. Interessanterweise basieren fast alle diese Beispiele auf kleinen bis sehr kleinen Stichprobengrößen. Das extremste Beispiel verwenden Zieffler et al. (2011) mit $n_1 = 2$ und $n_2 = 3$, welches allerdings auf einer tatsächlich durchgeführten Studie basiert[33]. Es findet sich jedoch bei diesen Autoren kein weiterer Hinweis auf eine zu verwendende Stichprobengröße für Randomsierungstests, sie weisen lediglich darauf hin, dass der P-Wert von der Stichprobengröße beeinflusst wird. An anderer Stelle findet sich, dass „in the case of randomization, do not rely on sample size assumptions" (Pfannkuch et al. 2011, 904), was sich als Argument auch bei Hesterberg et al. (2009) in ähnlicher Form findet.

Ludbrock und Dudley (1998) haben 252 Randomisierungstest-Studien aus dem biomedizinischen Forschungsfeld analysiert und festgestellt, dass „the group sizes used in the randomized studies were small. Overall, the median group size varied from 6 to 12" (Ludbrock & Dudley 1998, 127). Auch hier zeigt sich, dass in experimenteller Forschung in Vergleichsstudien durchaus mit kleinen Stichproben gearbeitet wird.

„Bei kleineren Stichproben gibt es für die statistische Hypothesenprüfung meistens keine Alternative zu den verteilungsfreien Tests" schreiben Bortz und Lienert (2008, 60) als Argument für die Verwendung des Randomisierungstests in klinischen Studien.

Es hat sich an keiner Stelle ein expliziter Hinweis auf eine minimale Stichprobengröße für die Anwendbarkeit eines Randomisierungstests gefunden.[34] Allerdings ist die Anzahl der möglichen Permutationen wichtig. Wenn es weniger als 20 mögliche Permutationen gibt, dann kann ein P-Wert kleiner als 0,05 nicht erreicht werden (da der P-Wert ein Vielfaches von $1/\binom{N}{n_1}$ ist, siehe oben). Bei gleich großen Gruppen sollte aus diesem Grund mindestens $n_1 = n_2 \geq 3$ sein und für zwei ungleich große Gruppen sollte $n_1 + n_2 \geq 5$ sein.

---

[33] Die Originalstudie kann unter Lochman, J. E., Lampron, L. B., Gemmer, T. C., Harris, S. E. & Wyckoff, G. M (1989). Teacher consultation and cognitive-behavioral interventions with aggressive boys. *Psychology in the Schools*, *26*(2), 179-188 nachgelesen werden.

[34] Vgl. Abschnitt 5.4.4 zur möglichen Problematik in Bezug auf Schlussfolgerungen bei kleinen Stichprobengrößen.

5.4.3 Was lässt sich aus den Ergebnissen eines Randomisierungstests schließen?

Wie bereits oben erwähnt, lassen sich mit Hilfe eines ordentlich designten Randomisierungstests kausale Schlussfolgerungen ziehen, bei Zufallsstichproben lassen diese sich auf die Population verallgemeinern (Ramsey & Shafer 2013). Rossman formuliert hierzu, dass „random *assignment* of units to treatment groups allows for cause-and-effect conclusions to be drawn about the relationship of the explanatory and response variables“ (Rossman 2008, 6, Hervorhebung im Original). Wie eine Antwort hierauf (obwohl bereits früher verfasst) formuliert Ernst „The consequence of this is that any inferences in the randomization model are limited to the subjects in the study” (Ernst 2004, 677). Ludbrock & Dudley machen ebenfalls sehr deutlich, dass über die vorliegende Stichprobe hinaus nicht verallgemeinert werden darf. Grundsätzlich gelten die Schlussfolgerungen für die Versuchseinheiten, mit denen das Experiment durchgeführt wurde.

> „If P = .05 results from randomization and the application of a permutation test, there is no statistical promise for the future. The statistical inference refers only to the actual experiment that has been performed and the P value indicates the probability that the way in which this experiment turned out was merely a matter of chance. However, this need not deter experimenters from inferring that their results are applicable to similar patients, animals, tissues, or cells, though their arguments must be verbal rather than statistical.” (Ludbrock & Dudley 1998, 129)

Ähnlich argumentiert auch Edgington, wie Ludbrock & Dudley mit dem Hinweis, dass neben statistischen Schlussfolgerungen auch „nichtstatistische” Schlussfolgerungen möglich sind, die über die Versuchseinheiten hinaus verallgemeinert werden können.

> „Inferences about treatment effects for other subjects must be nonstatistical inferences – inferences without a basis in probability. Nonstatistical generalization is a standard scientific procedure. We generalize from our experimental subjects to individuals who are quite similar in those characteristics that we consider relevant.” (Edgington 1995, 8)

5.4.4 Kritik am Randomisieren

Die folgende Darstellung ist angelehnt an Saint-Mont (2011).

Es stellt sich die Frage, ob das Verfahren der Randomisierung leistet, was es soll. Die vor der Durchführung eines Experiments vorgenommene zufällige Aufteilung der Versuchseinheiten soll zu Vergleichbarkeit der Gruppen führen und so induktive Schlüsse erlauben. Positiv formuliert sollen durch die Randomisierung potentiell unendlich viele Störvariablen kontrolliert werden. Jedoch ist dazu nötig, dass eine in der Realität durchgeführte Aufteilung der Einheiten auf die Gruppen, beispielsweise durch einen Münzwurf, zu tatsächlich vergleichbaren Gruppen führt. Allerdings muss dies nicht bedeuten, dass alle potentiellen Störvariablen gleich auf die beiden Gruppen aufgeteilt sind. Da es sich a priori immer um eine reale zufällige Zuordnung handelt, kann es wegen der potentiell unendlichen Anzahl von Störvariablen immer ein Ungleichgewicht bezüglich einer möglichen Störvariablen zwischen den Gruppen geben. „Bei den üblichen (kleinen) Stichprobengrößen und zahlreichen Störfaktoren ist die Wahrscheinlich-

keit deshalb auch unter Randomisierung sehr groß, dass zumindest einer der Faktoren extrem ungleich zwischen den beiden Gruppen aufgeteilt ist.“ (Saint-Mont 2011, 228). Wenn man eine solche Störvariable jedoch nicht beobachtet, so findet man keinen Anhaltspunkt dafür in den Daten und schreibt die im Experiment beobachteten Unterschiede fälschlicherweise einer durchgeführten Behandlung oder Maßnahme zu. Saint-Monts Lösung an dieser Stelle ist es, entweder die Daten der Gruppen bedingt auszuwerten, beispielsweise bei einem starken Ungleichgewicht von Männern und Frauen in den Gruppen die jeweiligen bedingten Mediane zu vergleichen, oder bereits im Vorfeld eine systematische Einteilung der Einheiten auf die Gruppen vorzunehmen, wie dies beispielsweise bei Zwillingsstudien getan wird. Jedoch hat die Randomisierung auch Vorteile gegenüber einer systematischen Einteilung, denn „[d]er Vorteil der Randomisierung überwiegt, wenn man sich in einer Situation mit potenziell vielen, unbekannten Faktoren befindet, man also mit der Forschung noch am Anfang steht und wenig weiß.“ (Saint-Mont 2011, 234. Aus dieser Vorteilsposition heraus wird der Randomisierungstest in der vorliegenden Arbeit verwendet, da gerade, wenn (wie in Lehrsituationen typisch) mit nicht selbsterhobenen Daten gearbeitet wird und über ein Experiment nicht viel mehr bekannt ist, als dass die Gruppen zufällig eingeteilt wurden, der Randomisierungstest trotz unbekannter Störfaktoren ein gutes Verfahren zum Schlussfolgern über mögliche Erklärungen bietet.

6 Randomisierungstests in der didaktischen Literatur

In der neueren (vorzugsweise anglo-amerikanischen) didaktischen Literatur wird ein Einstieg in die beurteilende Statistik über Randomisierungstests vorgeschlagen (vgl. Kapitel 5.3). Dabei steht die Annahme im Vordergrund, dass durch das Verfahren des Randomisierungstests die zugrunde liegende Denkweise der beurteilenden Statistik leichter erlernt werden kann, weil keine schwierigen Berechnungen oder Annahmen über Wahrscheinlichkeitsverteilungen getroffen werden müssen. Im Zentrum der Argumentationen steht dabei häufig ein Vortrag von George Cobb von 2005, der 2007 in Form eines Artikels publiziert wurde, in welchem er darauf hinweist, dass existierende (amerikanische) Curricula unnötig kompliziert seien, weil hauptsächlich die Normalverteilung im Fokus stehe, anstatt „the core logic of inference" (G. Cobb 2007, 4). Diese viel zitierte „Kernlogik des Schlussfolgerns" und seine Forderung führt Cobb wie folgt aus:

> „We need to throw away the old notion that the normal approximation to a sampling distribution belongs at the center of our curriculum, and create a new curriculum whose center is the core logic of inference. What is that core logic? I like to think of it as three Rs: randomize, repeat, reject. Randomize data production; repeat by simulation to see what's typical and what's not; reject any model that puts your data in its tail." (G. Cobb 2007, 11f)

Diese drei Rs sind seitdem zum Schlagwort der beurteilenden Statistik geworden. Viele Lehrende haben diese Forderung aufgenommen und verwenden nun Randomisierungstests, um in die Logik des inferentiellen Denkens einzusteigen. Auf Cobbs Forderung bauen einige Unterrichtsvorschläge und zugehörige Studien auf, die durch seine Schlagworte „three Rs" und „the core logic of inference" inspiriert wurden und Randomisierungstests in das Zentrum ihres Curriculums stellen. Mit diesen „three Rs" hat Cobb ein griffiges Schema aufgeführt, nach dem Inferenzstatistik unterrichtet werden kann. Ausführlich erläutert wird dies in Abschnit 6.3.

Zunächst wird Literatur betrachtet, in denen Beispiele zu Randomisierungstests ausgeführt sind, die entweder mit konkreten Unterrichtshinweisen oder mit Erfahrungswerten belegt sind. Daran anschließend werden empirische Studien vorgestellt, in denen Lernende beim Bearbeiten von Aufgaben zu Randomisierungstests beobachtet wurden.

Aus einer normativen Perspektive werden dazu anschließend verschiedene Schemata vorgestellt und in einer Zusammenschau verglichen. Hieraus wird ein eigenes Schema abgeleitet. Dies dient zum einen für die Verwendung in der in dieser Arbeit später beschriebenen Lerneinheit, zum anderen als Katalog notwendiger Konzepte, die bei einem Randomisierungstest adressiert werden müssen.

Da Randomisierungstests zur Einführung immer über Computersimulationen realisiert werden (dies ist ein wesentlicher Grund für die „Einfachheit" dieses Ansatzes), wird in

einem weiteren Unterkapitel explizit nach verwendeter Software geschaut. Welche Anforderungen werden genannt und ggf. welche Schwierigkeiten treten bei der Verwendung durch Lernende auf?

An didaktischer Literatur zu Randomisierungstests werden hier drei Stränge aufgenommen. Zum einen gibt es (vor allem anglo-amerikanische) Lehrwerke für *introductory statistics courses*, die die Einführung in die beurteilende Statistik über Randomisierungstests an den Anfang des Curriculums stellen (G. Cobb 2007, Fitch & Regan 2014, Lock et al. 2013, Rossman, Chance, Cobb & Holcomb 2008, Tintle, VanderStoep & Swanson 2009, Zieffler & Catalysts for Change 2013). Diesen ist gemeinsam, dass sie durchgängig Simulationen für die gesamte Inferenzstatistik verwenden und das Testen von Hypothesen am Beispiel von Randomisierungstests einführen. Diese Lehrwerke kommen (ob induziert oder nicht) Cobbs Forderung nach, beurteilende Statistik komplett mit empirischen Daten zu unterrichten[35]. Darüber hinaus existieren als zweiter Strang Unterrichtsvorschläge, die entweder ein einzelnes Beispiel zum Randomisierungstesten theoretisch beleuchten oder eine kurze Unterrichtssequenz mit mehreren Beispielen vorstellen (Biehler, Kombrink & Schweynoch 2003, Erickson 2006, Frischemeier & Biehler 2014, Gould, Davis, Patel & Esfandiari 2010, Lock et al. 2014 Pfannkuch et al. 2011). Hier finden sich auch einige konkrete Hinweise auf die Gestaltung von Lehrsituationen, in denen Randomisierungstests bzw. die Logik des schlussfolgernden Denkens verwendet werden sollen. Als letzter Strang finden sich empirische Studien, in denen Schüler einer Highschool oder Studierende an einem konkreten Beispiel zu Randomisierungstests arbeiten und deren Lösungsprozess näher untersucht wird (Biehler et al. 2015, Budgett et al. 2012, Budgett, Pfannkuch, Regan & Wild 2013, DelMas, Garfield, Ooms & Chance 2007, Frischemeier 2017, Noll & Kirin 2017, Pfannkuch et al. 2015, Rossman & Chance 2014, Swanson et al. 2014).

6.1 Lernerfolge, Schwierigkeiten und Unterrichtshinweise bei Randomisierungstests

In diesem Abschnitt wird der Frage nachgegangen, welche Ergebnisse empirische Untersuchungen zu Lernerfolgen und zu Schwierigkeiten beim Durchführen von Randomisierungstests durch Lernende zu beobachten sind. Desweiteren geht es um Designideen, z. B. welche Beispiele ausgewählt wurden, oder wie Hands-on Aktivitäten genutzt wurden. Dabei ist das Vorgehen chronologisch am Erscheinen der Literatur orientiert. Begonnen wird mit einigen Beispielen aus der Literatur, die bestimmte Aspekte für das Unterrichten von Randomisierungstests beleuchten.

Biehler et al. (2003) stellen ein frühes Beispiel zur Anwendung eines Randomisierungstests im Anschluss an eine Einheit zu Gruppenvergleichen in der Oberstufe vor. Im Mit-

[35] Diese Forderung findet sich bereits bei Gnanadesikan et al. (1987) oder auch bei Konold (1994).

telpunkt steht der selbsterhobene Datensatz MUFFINS (Medien- und Freizeitgestaltung für interessanten Stochastikunterricht), der für verschiedene Zwecke im Stochastikunterricht, bevorzugt in der 11. Klasse, eingesetzt werden kann. Im Artikel werden der Datensatz und exemplarische Fragen mit zugehörigen Auswertungen diskutiert. Eine der Auswertungen behandelt einen ausführlichen Gruppenvergleich zu der Frage, wie sich Jungen und Mädchen hinsichtlich der wöchentlichen Lesezeit unterscheiden. Ein Ergebnis ist, dass Mädchen tendenziell in der Woche eine Stunde länger lesen als Jungen. „An den Verteilungsvergleich schließt sich die natürliche Frage an: Wie zuverlässig sind denn die festgestellten Unterschiede zwischen Jungen und Mädchen?“ (Biehler et al. 2003, S. 19). An dieser Stelle findet sich zunächst der Hinweis, dass es sich bei dem Datensatz nicht um eine Zufallsstichprobe aus der Population der Elftklässler aus Nordrhein-Westfalen handelt und deshalb auch keine verallgemeinernden Schlüsse gezogen werden dürften. Hierauf folgt jedoch der entscheidende Hinweis für den Unterricht, dass

> „[es] in der Praxis aber auch in solchen Fällen durchaus üblich [ist], einmal hypothetisch von der Annahme auszugehen, es würde eine Zufallsstichprobe vorliegen. Erweisen sich unter dieser Annahme Unterschiede als nicht signifikant, so wird von vorneherein von einer Verallgemeinerung abgesehen. Falls unter dieser Annahme Signifikanz ermittelt wird, ist dies ein positiver Hinweis, man darf aber trotzdem nicht von einem definitiven statistischen Nachweis ausgehen. Wenn man diesen differenzierten Hinweis zu kompliziert findet, kann man sich in der Schule durchaus legitimiert fühlen, den Fall unter der Annahme, es wäre eine Zufallsstichprobe, einmal durchzuspielen.“ (Biehler et al. 2003, S. 20)

DelMas et al. (2007) berichten über Ergebnisse aus dem CAOS Test (Comprehensive Assessment of Outcomes in Statistics), an dem Studierende nach einem (traditionellen) Einführungskurs in Statistik teilnahmen. Dieser Test wurde für Studierende eines *introductory statistics course* mit dem Ziel erstellt, ein zuverlässiges Instrument zu haben, das eine Bandbreite an Items abdeckt, zu denen nach dem Besuch der Veranstaltung Verständnis erwartet wird. Gleichzeitig sollen Gebiete identifiziert werden können, in denen Studierende nach einem Einführungskurs in Statistik große Schwierigkeiten oder auch große Lernerfolge haben.

Hier wird auf zwei Ergebnisse eingegangen, die für die vorliegende Arbeit Relevanz haben. Berichtet wird über die Durchführung des Tests im Herbst 2005 und im Frühjahr 2006 mit insgesamt 1470 Studierenden.

Zum einen wird berichtet, dass circa die Hälfte der Studierenden (49,2%) Schwierigkeiten bezüglich des Items „Understanding of the purpose of randomization in an experiment“ (DelMas et al. 2007, 54) hat. Allerdings wurden hier Studierende betrachtet, die „traditionell“ unterrichtet wurden, das heißt Randomisierungstests im Speziellen waren nur eine von mehreren Methoden und wurden erst in der zweiten Hälfte des Kurses thematisiert. Ein weiteres Ergebnis bezieht sich auf die Interpretation einer Nullhypothese. Fast ein Drittel der Teilnehmer zeigt Schwierigkeiten bezüglich der Schlussfolgerung in Bezug auf die Nullhypothese und bewertet die Aussage „Rejecting the null hy-

pothesis means that the null hypothesis is definitely false“ als richtig (DelMas et al. 2007, 46). Knapp unter 60% der Teilnehmer zeigen jeweils Verständnis für die Items 24, 25 und 26 (DelMas et al. 2007, 44), welche lauten „24: Understanding that an experimental design with random assignment supports causal inference“ (59,5% richtige Antworten), „25: Ability to recognize a correct interpretation of a p-value“ (54,5% richtige Antworten) und „26: Ability to recognize an incorrect interpetation of a p-value (58,6% richtige Antworten). Schwierigkeiten, die hier aufgezeigt werden, sind also vor allem die Interpretation von P-Werten, mit denen gut 40% der Studierenden Schwierigkeiten hatten.

Die hier berichteten Schwierigkeiten bezüglich der Interpretation der Nullhypothese und in Bezug auf den P-Wert sind ähnlich zu Schwierigkeiten, die ich im Rahmen meiner Examensarbeit aufgearbeitet habe (Podworny 2007, 28f). Die dort identifizierten Problembereiche bezüglich des P-Werts sind *P-Wert als bedingte Wahrscheinlichkeit* (das ist P(beobachtete Daten|H_0 ist wahr) was von Lernenden häufig verwechselt wird mit P(H_0 ist wahr|beobachtete Daten)), *P-Wert als Bereichswahrscheinlichkeit* (das ist $P(X \geq h_{beob})$ anstelle nur die Wahrscheinlichkeit $P(X = h_{beob})$ zu betrachten), *P-Wert im Kontext interpretieren* (häufig wird von Lernenden der Kontext bei der Interpretation komplett außer Acht gelassen), *P-Wert als Entscheidungshilfe* (vor allem die Frage nach der Interpretation großer P-Werte), *Begrifflichkeiten* (das ist vor allem die Vermischung von Fachsprache und von Alltagssprache).

Holcomb, Chance, Rossman, Tietjen, et al. (2010) stützen sich auf die drei Rs von Cobb und entwickeln Aktivitäten für den Schulunterricht, in deren Zentrum das Verständnis über statistische Signifikanz steht. Am Ende stellen sie verschiedene Forderungen an die Auswahl eines Einführungsbeispiels. Ein Einführungsbeispiel sollte ihrer Meinung nach ein statistisch signifikantes Ergebnis zeigen, denn „an advantage of starting with a significant example is that students may find it easier to judge when an observed result is very unlikely to occur, as opposed to making a judgement about whether it is not unlikely to occur“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, S. 4). Als möglichen Nachteil für die Verwendung eines Beispiels mit einem nicht signifikanten Ergebnis führen sie aus, dass „this may reinforce students’ natural inclinations to regard a non-small *p*-value as evidence in support of the null model, rather than a lack of evidence against the null model (…)“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, S. 4). Sie führen ebenfalls aus, dass es aus ihrer Sicht besser ist, mit einem Beispiel zu beginnen, bei dem ein Vergleich zwischen zwei Gruppen durchgeführt wird, in welchem sich zunächst auf bestimmte Anteile konzentriert wird, statt eine komplette 2x2 Tabelle zu verwenden, bei der nicht auf den ersten Blick ersichtlich ist, weshalb es genügt, eine Zelle zu betrachten. Hierzu berichten sie jedoch, dass es einigen Lernenden Schwierigkeiten bereitet, entsprechende Anteile zu berechnen und die Differenz zu bilden, so dass „some students get bogged down in calculating the two proportions and the difference

between them, (…) distracting their attention from the inferential reasoning process we are focussing on“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, S. 4). Darüber hinaus plädieren sie dafür, dass vor dem Simulieren mit Software eine Simulation von Hand geschehen soll, um so den Vorgang in der Software besser zu verstehen. Hierzu liefern sie eigene Erfahrungswerte. Es wurde ein Experiment durchgeführt, bei dem die Lernenden in zwei Gruppen eingeteilt wurden, von denen die erste Gruppe (n=20) eine händische Simulation durchführte, indem in einer Bonbonstichprobe vom Umfang n=25 der Anteil orangener Bonbons bestimmt wurde. Dies wurde graphisch dargestellt und später mit Software der gleiche Prozess für viele Stichproben durchgeführt. Die zweite Gruppe (n=43) arbeitete direkt mit einer computergestützten Simulation. Als Ergebnis halten die Autoren fest, „an interesting aspect of this study was that the students in both the tactile group and the other group appeared to finish the activity in about the same amount of time, suggesting that the tactile aspect does not take more time and does not hinder learning“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, S. 6). Es wird jedoch auch berichtet, dass Fragen zum Verständnis von beiden Gruppen ähnlich gut beantwortet wurden, es also keinen Vorteil durch die zuvor händisch ausgeführten Simulationen zu geben scheint. In einem Fragebogen wurde den Lernenden u. a. die Frage gestellt „Do you think that the hands-on simulation with the cards added to your understanding of the randomization process, in addition to the computer applet?“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, 6). 50% der Befragten fanden die Karten hilfreich. Sowohl G. Cobb (2007), als auch Rossman (2008) und Rossman et al. (2008) schlagen einen solchen, wie sie es nennen, „intuitiven“ Zugang über Randomisierungstests als Einstieg in die Logik des Schlussfolgerns vor. Der zugehörige Prozess des Schlussfolgerns kann ihrer Meinung nach gut von Lernenden an konkreten Beispielen nachvollzogen und verstanden werden, sofern er durch eine hands-on Aktivität gestützt wird (Gould et al. 2010, Roy et al. 2014; Tintle et al. 2014; West 2014).

Ein eigener Strang der Forschung zu Randomisierungstests sind die aufeinander aufbauenden Arbeiten von Budgett et al. (2012) und Pfannkuch et al. (2015), die im Folgenden vorgestellt werden. Bevor die empirische Untersuchung von Budgett et al. (2012) vorgestellt wird, folgt ein Einblick in die vorangegangene Forschungsarbeit von Pfannkuch et al. (2011), da die Studie von Budgett et al. (2012) in diesen Rahmen eingebettet ist. Pfannkuch et al. (2011, 903) sprechen von der „revolution in teaching inference that must begin“ und stellen im Sinn von G. Cobb (2007) Randomisierungstests in den Mittelpunkt ihres Designs von Unterricht. Als Begründung führen Pfannkuch et al. (2011, 903) an, dass „the randomization method offer simplicity and direct access to the logic of inferential reasoning. Furthermore, they open up an opportunity to give *all* students access to understanding inference“. Sie schlagen an dieser Stelle vor, dass Randomisierungstests die Einführung in die beurteilende Statistik darstellen sollen. Die Argumentation beim Randomisierungstesten führen Pfannkuch et al. (2011) näher an der

“Fishoil and Blood pressure study” (Abbildung 157) aus. Da dies in weiteren Veröffentlichungen der Autoren genutzt und empirisch mit Lernenden untersucht wird (vgl. Budgett et al. 2012), wird diese Aufgabe hier ausführlich dargestellt.

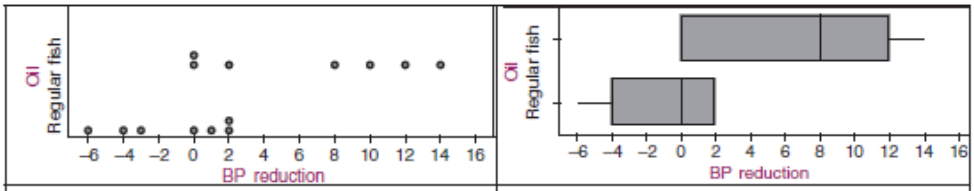
Fish Oil and Blood Pressure Study (Chance and Rossman [15, p. 153])
(based on a study by H. Knapp and G. Fitzgerald, *The antihypertensive effects of fish oil: A controlled study of polyunsaturated fatty acid supplements in essential hypertension*, N. Engl. J. Med. 320(16) (1989), pp. 1037–1043)

Prior to conducting this study the researchers conjectured that those on a fish oil diet would tend to experience greater reductions in blood pressure than those on a regular oil diet. Researchers randomly assigned 14 male volunteers with high blood pressure to one of two four-week diets: a fish oil diet and a regular oil diet. Therefore the treatment is the fish oil diet while the regular oil diet is the control.

Each participant's diastolic blood pressure was measured at the beginning and end of the study, and the reduction was recorded. The resulting reductions in diastolic blood pressure, in millimetres of mercury, were:

Fish oil diet:	8	12	10	14	2	0	0
Regular oil diet:	–6	0	1	2	–3	–4	2

Plots of the data and summary statistics were:



Dotplots of reductions in blood pressure | Box plots of reductions in blood pressure

Variable	Oil	N	Mean	SD	Min	LQ	Median	UQ	Max
BP reduction	Fish	7	6.57	5.86	0	0	8	12	14
	Regular	7	–1.14	3.18	–6	–4	0	2	2
Difference			7.71				8		

Summary statistics comparing blood pressure reductions of the fish oil and regular diet groups

Abbildung 157: Fishoil and Blood Pressure Study (Aufgabe entnommen aus Pfannkuch et al. (2011, S. 906))

In dieser „Fischöl und Blutdruck Studie“ geht es also um die Daten eines realen Experiments (veröffentlicht von Knapp & Fitzgerald 1989) zum Zusammenhang zwischen der Einnahme von Fischöl oder von „normalem Öl“ und der Senkung des Blutdrucks. Dazu wurden 14 männliche freiwillige Teilnehmer zufällig auf zwei Gruppen aufgeteilt, von denen eine Gruppe Fischöl bekam und die andere „normales Öl“. Abbildung 157 zeigt die Einbettung und die Ergebnisse dieses Experiments, so wie sie von Pfannkuch et al. (2011) verwendet wurden.

Die Ausführungen zu der Aufgabe beginnen damit, dass die Autoren festhalten, dass ein Unterschied der Mittelwerte zwischen beiden Gruppen von 7,71 mmHg zwischen der Fischölgruppe und der Kontrollgruppe gefunden wurde. Als erstes wird nach möglichen Erklärungen für diesen beobachteten Unterschied gefragt. Dazu werden zwei genannt:

„(1) The variability can be entirely explained by chance factors (differences between individuals who happened to be randomly assigned to each group and measurement errors) alone. That is, in this example, the fish oil diet is no better than the regular oil diet.

(2) The variability requires explanation by both chance factors and the treatment factor. That is, in this example, the fish oil diet treatment is effective." (Pfannkuch et al. 2011, 905)

In dieser und den folgenden Forschungsarbeiten sind dies die zwei zentralen Erklärungen, die die Autoren für das Auftreten eines Unterschieds zwischen zwei Gruppen als Norm setzen. Hinter der zweiten Erklärung stecken zwei mögliche Erklärungen, die sich etwas versteckt wiederfinden lassen. Die zusammengefasste zweite Erklärung lautet, dass die Behandlung mit Fischöl wirksam ist. Dies kann entweder daran liegen, dass tatsächlich Fischöl wirksam ist; das wären dann die „treatment factors". Oder es gibt weitere, unbekannte Faktoren, die „chance factors", die dazu führen, dass der Blutdruck der Teilnehmer der Fischöldiät stärker gesenkt wird. Die in Kapitel 5.4 vorgestellte dritte Erklärung, dass ein beobachteter Unterschied an so genannten konfundierenden Variablen liegt, die eine systematische Verzerrung der Ergebnisse hervorrufen, wird von den Autoren nicht betrachtet.

Im weiteren Argumentationsprozess wird von Pfannkuch et al. nun standardmäßig die Versprachlichung „chance alone" genutzt. Sie setzen ein eigenes Schema mit drei Rs an, das auf dem Schema von G. Cobb (2007) basiert. Die drei Schritte sind hier „*Re-randomize* the assignment of the units to the groups, *Repeat* this re-randomization many times, *Review* the chance alone explanation (note that this differs from the three Rs of Cobb, which relate to a broader scope)" (Pfannkuch et al. 2011, 906, Hervorhebung im Original). Die Beobachtung der Unterschiede zwischen den beiden Gruppen und das Aufstellen der Nullhypothese oder des Nullmodells zu „chance alone" gehört als Schritt vor das erste der drei Rs. Es folgt das zufällige Neuzuordnen der Werte zu den Gruppen (*re-randomize the assignment of the units to the groups*). Hierzu erfolgt der explizite Hinweis, dass bei dieser Neuzuordnung jeder Teilnehmer der Studie „seine" Blutdrucksenkung behält, da es unter der „chance alone" Erklärung keine Auswirkung der Gruppenzugehörigkeit auf die Blutdrucksenkung gibt. Dies erscheint den Autoren so wichtig, dass ein expliziter Hinweis auf mögliche Schwierigkeiten von Lernenden zu diesem Punkt gegeben wird: „That is, he [the volunteer of the study] will have the same BP [blood pressure] reduction irrespective of the oil diet group he is assigned to, a point we think that students may not realize" (Pfannkuch et al. 2011, 906). Mit anderen Worten soll erkannt werden, dass unter der „chance alone" Erklärung die Blutdrucksenkungen unabhängig von der Zuordnung zu einer Gruppe, und damit unabhängig von der Einnahme des Fischöls oder des „normalen Öls" sind. Wenn dies so dargestellt wird, ist dies wohl ein Hinweis, dass im Unterrichtsgeschehen explizit darauf eingegangen werden sollte. Dieser Prozess des zufälligen Neu-Zuordnens soll die Zuordnung der Teilnehmer zu den Gruppen wie im ursprünglichen Experiment nachahmen. Danach soll der Unterschied der neuen Gruppenmittelwerte berechnet

werden, so wie dies auch für das ursprüngliche Experiment getan wurde. Anschließend folgt das zweite R: *Repeat this re-randomization many times*. In ihren Ausführungen verwenden sie eine Wiederholungszahl von n = 1000 ohne nähere Angaben, aus welchem Grund diese Wiederholungszahl gewählt wird. Das Erstellen der zugehörigen Verteilung und das Auffinden des P-Werts wird nicht detailliert ausgeführt, die Verteilung wird gezeigt und der P-Wert mit 0,005 angegeben, wobei die Vokabel P-Wert explizit vermieden wird, stattdessen wird dies „tail proportion“ genannt. Als Erläuterung wird hinzugefügt „*under chance alone*, an observed difference of at least 7.71mmHg occurred only 5 times out of 1000 re-randomizations“ (Pfannkuch et al. 2011, 906, Hervorhebung im Original). Für den dritten Schritt *Review the chance alone explanation* wird als erstes festgehalten, dass dies der schwierigste Schritt für Lernende sei (Pfannkuch et al. 2011, 907). Es erfolgt eine detaillierte Interpretationslinie. Es werden noch einmal die zwei möglichen Erklärungen notiert. Entweder „Chance is operating alone” oder „Treatment is effective“. Ausgehend von dem erhaltenen (simulierten) P-Wert von 0,5% („tail proportion“) wird wie folgt argumentiert:

> „(1) If chance were operating alone, then it would be extremely surprising to observe a difference in means of 7.71mmHg or greater (happens about 0.5% of the time). Chance alone virtually always leads to a difference in means of 7.71mmHg or less (99.5% of the time).
>
> (2) But, we observed a difference of 7.71 mmHg.
>
> (3) Therefore, we find the chance alone explanation difficult to believe, i.e. we find it implausible.
>
> (4) Therefore, we have evidence that chance is not operating alone.
>
> (5) Thus, we have evidence that the fish oil diet treatment is effective.” (Pfannkuch et al. 2011, 907)

In analoger Weise wird ausgeführt, wie diese Argumentationslinie aussehen würde, wenn die „tail proportion“ 20% betragen würde:

> „(1)If chance were operating alone, then it would not be surprising to observe a difference in means of 7.71mmHg or greater (happens about 20% of the time). Differences this large or even larger happen quite often (20% of the time) under chance alone.
>
> (2) We observed a difference of 7.71 mmHg.
>
> (3) The chance is operating alone explanation is plausible.
>
> (4) Therefore, we have no evidence against the chance alone explanation. Both explanations remain standing; chance could be operating alone or the treatment could be effective.” (Pfannkuch et al. 2011, 908)

Als Schlussfolgerung aus einem großen P-Wert wird hier gezogen, dass beide Erklärungen (*chance alone* oder *treatment is effective*) möglich sind. Als besonderen Vorteil dieses Ansatzes stellen sie heraus: „Note that the beauty of this method is that it is formula-free, with no assumptions about population distributions and sample size” (Pfannkuch et al. 2011, 908). In diesem Sinn wird weder die anfängliche Stichprobengröße von n = 14 Versuchspersonen des Experiments, noch die

Wiederholungszahl für die Simulation des Randomisierungstests von n = 1000 thematisiert.

Den Ausführugen von Pfannkuch et al. (2011) entlang des Beispiels zur Fischölstudie lässt sich gut folgen. Als besonders lässt sich hier herausstellen, dass die Autoren für den beobachteten Unterschied zwischen den beiden Gruppen bezüglich der Blutdrucksenkung zwei Erklärungen anführen, von denen aus meiner Sicht die zweite Erklärung weiter spezifizert werden könnte, um Lernenden einen ersten Kontakt mit dem Gedanken von konfundierenden Variablen zu ermöglichen. Ein Effekt des Verzichts auf statistische Sprache, wie beispielsweise das Benennen der Verteilung der realen Daten mit „dot plots“ anstatt mit „distribution“ (Pfannkuch et al. 2011, 909) und der Verzicht, die Nullhypothese als solche zu benennen, lässt sich nicht zeigen. Ob die sprachlich neuen Konstrukte tatsächlich einfacher für Lernende sind, lässt sich hier schwer feststellen. Die ausführliche Ausführung der Argumentationslinie zu einem erhaltenen P-Wert ist logisch aufgebaut und kann für die vorliegenden Arbeit adaptiert werden.

Darauf aufbauend entwickeln Budgett et al. (2012) Lerntrajektorien für das Unterrichten von statistischer Inferenz für die 13. Klasse, welche das letzte Schuljahr in Neuseeland ist und für Studierende im ersten Semester an der Universität, die einen Einführungskurs zu Statistik besuchen. Dabei wählen sie nach eigenen Worten diesen innovativen Zugang über Randomisierungstests,

> „we generated a more natural form of argumentation based on the Fisherian framework. We limited the learning situations to comparative experiments in order to link how the randomization method mimics the data production process, which in turn determines the type of inference that can be drawn, that is, causal.“ (Budgett et al. 2012, 3)

Berichtet wird über eine Pilotstudie zu einer eintägigen Schulung, von der ein halber Tag Randomisierungstests zum Inhalt hat und ein halber Tag dem Bootstrapping gewidmet ist (dieser Inhalt wird jedoch im Artikel nicht berichtet). Ausgewertet werden die Ergebnisse von fünf Schülern aus dem 13. Schuljahr und von fünf Studierenden im ersten Semester. Von den fünf Schülern der 13. Klasse wird lediglich berichtet, dass sie gute Noten haben. Weiteres Vorwissen wird nicht dargestellt. Drei der fünf Studierenden haben bereits Vorerfahrungen durch einen Einführungskurs in Statistik. Welche Inhalte ihnen bekannt sind, wird jedoch auch nicht berichtet. Die Schulung bildet eine Einführung in das Themengebiet zur beurteilenden Statistik, berichtet wird im Artikel zu dem Teil, in welchem Randomisierungstests behandelt werden. In der Pilotstudie wird ein Vor-/Nachtest durchgeführt mit einem dazugehörigen Interview, und Beobachtungen und Reflektionen des Projektteams werden ausgewertet.

Im Vortest werden zwei Fragen im Kontext der „Fishoil and Blood Pressure Study“ gestellt (siehe Abbildung 157). Die erste Frage zielt über eine Multiple-Choice-Frage auf

das Verständnis bezüglich der zufälligen Zuordnung der Versuchsteilnehmer zu den Gruppen, die zweite zielt darauf, welche zwei möglichen Erklärungen abgegeben werden für den beobachteten Unterschied zwischen den beiden Gruppen. Der genaue Wortlaut der ersten Frage ist:

> „What was the purpose of the random assignment of the 14 male volunteers to one of the two groups? Which ONE of the following statements gives the best response to this question?
>
> A. To increase the accuracy of the research results.
>
> B. To ensure that all male participants with high blood pressure had an equal chance of being selected for the study.
>
> C. To reduce the amount of sampling error.
>
> D. To produce treatment groups with similar characteristics.
>
> E. To prevent skewness in the results." (Budgett et al. 2012, 5)

Als richtige Antwort stufen die Autoren Antwort D ein, die drei der zehn Teilnehmer der Pilotstudie im Vortest ankreuzen. Wobei sich hier die Frage stellt, ob durch die Multiple-Choice Situation nicht Informationen verloren gehen, da mit den richtigen Begründungen versehen auch Antwort A möglicherweise als richtig eingestuft werden kann. Dies decken die Autoren jedoch in einem anschließend durchgeführten Interview auf, da weitere drei Teilnehmer Antwort A angekreuzt haben und in ihren dann gelieferten Begründungen darauf zielen, dass dadurch Verzerrungen vermieden werden („all of these students talked about avoiding bias and wanting comparisons to be fair", Budgett et al. 2012, 5). Daraus schließen die Autoren, dass eher die genaue Formulierung der Antwortmöglichkeiten als die dahinterstehenden Begründungen eine Schwierigkeit darstellen.

In Bezug auf die zweite Frage des Vortests nach zwei möglichen Erklärungen für die beobachteten Unterschiede interessierten sich die Autoren dafür, ob „Zufall" als mögliche Erklärung genannt wird oder nicht. Als Ergebnis halten sie fest, dass bis auf zwei Teilnehmer alle (also 80% der Teilnehmer) als eine mögliche Erklärung notieren, dass Fischöl wirksam ist, um den Blutdruck zu senken. Als zweite mögliche Erklärungen werden verschiedene genannt, z. B. dass sich die Teilnehmer der Fischöl-Gruppe gesünder ernähren, dass psychologische Faktoren zu einer Verbesserung führen, weil sie wussten, dass sie behandelt wurden, usw. Ein Beispiel stellt S3 dar:

> „S3: The participants in the fish oil group had a lifestyle better suited to reducing blood pressure than the regular oil group." (Budgett et al. 2012, 5)

In den Begründungen werden vor allem Kausalfaktoren genannt, ohne zu erkennen, dass die zufällige Aufteilung der Versuchspersonen auf die Gruppen dazu führen soll, dass sich die Gruppen darin nicht unterscheiden. S3 argumentiert also in die Richtung der dritten Erklärung mit konfundierenden Variablen, die von mir bereits zu Pfannkuch et al. (2011) hinzugefügt wurde, die jedoch hier von den Autoren nicht weiter kommen-

tiert wird. Ein interessantes Ergebnis des Vortests von Budgett et al. (2012) ist, dass keiner der zehn Teilnehmer als mögliche zweite Erklärung den Zufall nennt.

Daran anschließend wird die Sequenz der eintägigen Schulung mit den verschiedenen Aktivitäten berichtet. Ein genauer Verlaufsplan ist nicht angegeben. Begonnen wird mit einem Experiment zum Merken von Buchstaben, das im Klassenraum durchgeführt wird, zu dem die Anwesenden in zwei Gruppen eingeteilt werden sollen. Diese Einteilung wird durch den Lehrenden vorgegeben und zwar bewusst nicht zufällig. Alle über 18 Jährigen sollen in eine Gruppe, alle anderen in die zweite. Dieses Szenario wird zur Diskussion gestellt, von den Lernenden jedoch zunächst nur zögerlich in Frage gestellt. Verschiedene Einteilungen werden durch den Lehrenden vorgeschlagen, bis durch die Lernenden eine zufällige Einteilung in zwei Gruppen vorgeschlagen wird mit der Begründung, dass „there would need to be random assignment to each group so that a fair comparison could be made“ (Budgett et al. 2012, 6). Aus der Darstellung wird ersichtlich, dass den Lernenden das Ziel der Aktivität nicht klar zu sein scheint. In einer zweiten Aktivität wird das statistische Argumentieren demonstriert am Beispiel „Has Sara brushed her teeth?“. Dazu wird folgende Situation beschrieben und durch die beiden Argumentationsprozesse in Tabelle 10 beschrieben.

Tabelle 10: Entwicklung einer „alltäglichen“ Argumentation (entnommen aus Budgett et al. 2012, 6)

		Scenario One	**Scenario Two**
1.	Statement to test	She has brushed her teeth.	She has brushed her teeth.
2.	Collect data (information)	The toothbrush is dry.	The toothbrush is wet.
3.	Consider 1. and the data: *If 1 is true, then what are the chances of getting data like that in 2.?*	*The toothbrush-is-dry* would be *highly unlikely* if she had brushed her teeth.	*The toothbrush-is-wet* would *NOT* be surprising if she had brushed her teeth.
4.	Review the statement in 1. in light of 3. together with the data in 2.	Therefore, it’s a fairly safe bet she has <u>not</u> brushed her teeth. I have evidence that she has not brushed her teeth.	Therefore she <u>could</u> have brushed her teeth. *Or she could have just run the brush under the tap.* I have <u>no</u> evidence that she has NOT brushed her teeth.

In der dritten Aktivität wird nun ein probabilistischer Kontext thematisiert, um den Lernenden zu demonstrieren, welche Ergebnisse zufällig vorkommen können und ihnen ihre Erwartungshaltung dazu offenzulegen. Es wird gefragt, wie wahrscheinlich es ist, eine rote Karte aus einem Kartenstapel zu ziehen. Bei einem Standardspiel wäre diese Wahrscheinlichkeit 0,5. Der von den Lehrenden verwendete Kartenstapel enthält jedoch nur schwarze Karten, was die Lernenden nicht wissen. Nach Fragen zu der Erwartungshaltung der Lernenden, welche Anzahl roter Karten sie beim Ziehen mit Zurücklegen von 10 Karten erwartet, werden die Karten gezogen. Da alle gezogenen zehn Karten

schwarz sind, werden verschiedene Fragen zu möglichen Ergebnissen und deren Wahrscheinlichkeiten gestellt unter der Annahme, dass ein faires Kartenspiel vorliegt.

Die vierte Aktivität beinhaltet nun einen Randomisierungstest. Es wird eine Studie verwendet, in der geprüft werden soll, ob bestimmte Maßnahmen das Laufenlernen von Kleinkindern beschleunigen. Es liegen Daten zu zehn Kleinkindern vor, die zufällig in zwei Gruppen aufgeteilt wurden, von denen eine Gruppe täglich ein 12-Minuten Programm absolvierte und die andere Gruppe nicht. Es ergibt sich ein Unterschied zwischen den Medianen der beiden Gruppen von 2,5 Monaten, die die Kinder der Behandlungsgruppe schneller laufen lernten. Als mögliche Erklärungen mit den Lernenden werden diskutiert, ob diese Unterschiede allein am Zufall liegen oder ob sie am Zufall und an der Behandlung liegen. Hierzu wird eine Hands-on Aktivität durchgeführt, in der die Daten der zehn Kleinkinder (Gruppenzugehörigkeit und Alter beim Laufenlernen) auf Zetteln notiert werden. Diese Zettel werden dann zwischen den beiden Merkmalen zerrissen und daraus zwei Stapel gebildet, die gemischt werden. Aus beiden Stapeln wird dann zufällig je ein Zettel gezogen und so zwei „neue“ Gruppen definiert, die wiederum hinsichtlich des Medians ausgewertet und verglichen werden (der Begriff Teststatistik wird nicht verwendet). Dieser Prozess wird anschließend per Computersimulation häufig durchgeführt, um eine Referenzverteilung zu erzeugen (die ebenfalls nicht so benannt wird). Als Software wird die von den Autoren speziell für die Durchführung von Randomisierungstests entwickelte VIT-Umgebung verwendet[36], die den Prozess dieser Re-Randomisierung sichtbar macht. Anschließend wird der P-Wert von 0,02 ermittelt (der „tail-proportion“ genannt wird). Mit einer Verbindung zur „Has Sara brushed her teeth?“-Geschichte wird nun vorgestellt, welche Schlussfolgerungen sich ziehen lassen.

Ein ähnliches Beispiel wird als letzte Aktivität in der Lernsequenz diskutiert, bei dem der P-Wert allerdings 0,15 beträgt, um auch hier zu verdeutlichen, welche Schlussfolgerungen gezogen werden können. Während dieser Einheit wird als Orientierung zur Interpretation der „tail proportion“ gegeben, „if the tail proportion was less than 10%, we had evidence against *chance acting alone*“ (Budgett et al. 2012, 11, Hervorhebung im Original). Somit wird eine scharfe Entscheidungsgrenze vorgegeben bezüglich der die „chance alone“ Erklärung beurteilt werden soll.

Im Nachtest, der eine Woche nach der Schulung stattfindet, werden unter anderem die gleichen Fragen bezüglich des Fischölexperiments wie im Vortest gestellt. Hier zeigt sich, dass bezüglich der Frage nach dem Sinn der zufälligen Zuteilung der Versuchsteilnehmer zu den Gruppen keine Verbesserung in den Antworten zu erkennen ist. Auch hier zeigt jedoch ein anschließendes Interview mit den Teilnehmern der Studie, dass anscheinend Sprachbarrieren in der Formulierung der Antwortmöglichkeiten zu der

[36] https://www.stat.auckland.ac.nz/%7Ewild/VIT/ (abgerufen am 29.06.2018).

schlechten Antwortrate geführt haben, da die Lernenden im Interview durchaus die richtigen Begründungen liefern können. Bezogen auf die zweite Frage des Nachtests zu den zwei möglichen Erklärungen für die beobachteten Unterschiede nennen wieder alle Teilnehmer als eine Erklärung, dass die Behandlung mit Fischöl wirksam ist. Allerdings führen nun auch sieben der zehn Teilnehmer den Zufall als mögliche zweite Erklärung an, wie die Beispiele der folgenden drei Teilnehmer zeigen.

> „S2: Chance may or may not be acting on the reduction in blood pressure.
>
> S3: That the fish oil diet group was made up of people who were more likely to have their blood pressure reduced for whatever reason.
>
> S4: Chance is acting alone. The observed data has resulted in the way it has by chance." (Budgett et al. 2012, 10)

Weiterhin werden die Teilnehmer im Nachtest aufgefordert, zwei weitere Fragen zu beantworten. Einmal sollen sie die (fiktive) Zeitungsaussage „People can lower their blood pressure with a fishoil diet" kommentieren. Zum zweiten sollen sie einen P-Wert („a tail proportion") von 0,27 interpretieren. Das Interesse der Forschung von Budgett et al. in Bezug auf die erste Frage ist, dass die Lernenden die Aussage korrekt umschreiben sollen. Dies geschieht jedoch nicht, es wird lediglich von den Lernenden darauf hingewiesen, dass „people" zu allgemein sei oder dass die Studie nur an Männern und nicht an Frauen stattfand. Auch im anschließenden Interview wird die Aussage nicht in der erhofften Richtung präzisiert. Bezüglich der zweiten Frage wird die Interpretation des neuen P-Werts von 0,27 durch die Lernenden streng entlang des 10% Kriteriums durchgeführt. Die weiteren, im Interview erfragten Begründungen sind hierzu sehr uneinheitlich und reichen von der (falschen) Akzeptanz der *chance alone explanation* bis zu der Aussage, dass Fischöl definitiv nicht wirksam sei (Budgett et al. 2012, 11). Die schlechten Ergebnisse bezogen auf diese letzte Frage erklären die Autoren damit, dass die gesamte oben beschriebene Schulung nur einen Zeitumfang von zwei Stunden hatte und deshalb nicht genügend auf große P-Werte eingegangen werden konnte.

Ohne dass eine genaue Aufgabenstellung berichtet wird, gibt es im Nachtest offensichtlich noch einen Teil, in dem die Lernenden einen Randomisierungstest in Zweiergruppen mit Software durchführen sollen. Hierzu bekommen sie Daten eines Experiments ähnlich wie in der Fischöl-Studie. Die Daten zu zwei randomisierten Gruppen mit der Frage „Does added calcium intake reduce blood pressure?" (Budgett et al. 2012, 11) werden präsentiert. Hierzu wird nur knapp berichtet, dass die Lernenden die Software auf technischer Ebene gut bedienen können. Zu den statistischen Argumentationen wird leider nichts berichtet.

Es ergeben sich Schwierigkeiten der Lernenden bezüglich der Bedeutung der Aussage *chance is acting alone*. Dies machen die Autoren daran fest, dass es einigen Lernenden nicht möglich ist, diesen Sachverhalt in eigenen Worten auszudrücken.

Insgesamt ziehen die Autoren ein positives Fazit aus dieser Studie zum Lernen der statistischen Schlussweisen innerhalb einer zweistündigen Schulung zu Randomisierungstests. Sie nennen kleinere Änderungen vor allem bezüglich der verwendeten Sprache, die sie in einem neuen Durchgang umsetzen wollen. Auch die Hands-on Aktivität in Verbindung mit der anschließenden Computersimulation schätzen sie als wertvoll ein: „the technology coupled with the hands-on acitivity seemed to be essential for their understanding of the randomization method“ (Budgett et al. 2012, 12).

Ohne dass vorher die Begriffe Null- und Alternativhypothese jemals erwähnt wurden, kommen sie zu dem Fazit, dass „after much reflection we have modified the learning trajectory so that the two explanations for an observed difference are not defined in the manner of a null and alternative hypothesis“ (Budgett et al. 2012, 12). Leider werden weder die hierzu führenden Gedanken dargelegt, noch wird erklärt, wie das Konzept der beiden Erklärungen im konkreten Unterricht mit dem Randomisierungstest verbunden wurde.

Die vorgestellte Forschungsarbeit von Budgett et al. (2012) ist in zwei Richtungen interessant. Zum einen zeigen sie ein Unterrichtsexperiment auf, in welchem in einer sehr kurzen Zeitspanne (zwei Stunden) in die Logik des inferentiellen Schließens mit Randomisierungstests eingeführt wird. Einzig der Einstieg über ein Beispiel zum zufälligen Aufteilen der Lernenden auf zwei Gruppen scheint nicht erfolgreich zu sein, weil die Lernenden den Sinn dahinter nicht zu verstehen scheinen. Die Autoren stufen ihr Experiment selbst als gelungen ein und die Ergebnisse des Nachtests zeigen, dass zumindest einige wichtige Elemente von den Lernenden genannt werden konnten. Es ist also möglich, in einer kurzen Zeitspanne in das Thema der beurteilenden Statistik einzuführen. Zum zweiten sind die Erkenntnisse bezüglich des Vor-/Nachtests interessant. Ein Fokus der Forschungsarbeit lag darauf zu untersuchen, welche möglichen Erklärungen Lernende für beobachtete Ergebnisse eines Experiments, das war hier der beobachtete Unterschied zwischen zwei Gruppen, angeben. Im Vortest zeigte sich, dass die Lernenden quasi ausschließlich die Erklärung bezüglich der Wirkung der Behandlung nennen. Nach der Lerneinheit können immerhin sieben von zehn Lernenden beide gewünschten Erklärungen nennen, also insbesondere auch die Erklärung, dass „chance is acting alone“. Es zeigt sich hier noch einmal, wie wichtig gerade in diesem Bereich der Stochastik das genaue Formulieren ist. Ob nun die bewusste Vermeidung der Begriffe Nullhypothese, Teststatistik, Referenzverteilung und P-Wert den erhofften Effekt hatte, nämlich dass genau deshalb die zugrunde liegende Logik des inferentiellen Schließens besser verstanden wurde, sei einmal dahingestellt. Aus meiner Sicht erscheint es wenig Mehraufwand, die Konzepte, die unterrichtet wurden, auch mit den entsprechenden Bezeichnungen zu versehen. Dadurch ließe sich leichter auf die jeweiligen Konzepte referenzieren und das spätere Anknüpfen an eine formale Sprache würde weniger Hürden bieten. Außerdem wird von den Autoren kritisch angemerkt, dass die Lernenden die 10%

Grenze als strikte Entscheidungsgrenze verwenden. Dies ist aus meiner Sicht nicht so negativ einzustufen, wie die Autoren das tun. Als Interpretation bleibt ansonsten, dass sich die Autoren vielleicht erhofften, dass die Lernenden die Unsicherheit in ihre Erklärungen mit einbeziehen, dass, auch wenn die Entscheidung für eine Erklärung gefallen ist, die jeweils andere Erklärung möglich, wenn auch nicht plausibel ist. Die Schwierigkeiten, die berichtet werden bezüglich der Formulierung der Multiple-Choice Antworten und den Erläuterungen, die Lernende dazu geben, stellen in Frage, ob dies eine geeignete Form ist, um Annahmen bezüglich der möglichen Erklärungen für einen beobachteten Unterschied aufzudecken. Vielleicht wäre eine offene Antwortmöglichkeit zielführender gewesen.

In einer letzten Forschungsarbeit der neuseeländischen Autoren über die Lernsequenz zu Randomisierungstests mit der Fischöl-Aufgabe als Vor-/Nachtest identifizieren Pfannkuch et al. (2015) Aspekte beim Randomisierungstesten, in welchen Unsicherheit eine Rolle spielt. Sie legen den Schwerpunkt nun auf das Schlussfolgern bezüglich einer Ursache mittels Randomisierungstests und nennen diesen Ansatz *experiment-to-causation*. Erneut wird das Setting aus Budgett et al. (2012) verwendet, in welchem Erstsemesterstudierende an einem Einführungskurs in Statistik teilnehmen und in einer zweistündigen Lerntrajektorie in das Themengebiet der Inferenzstatistik über Randomisierungstests eingeführt werden. Die Autoren identifizieren zwölf Elemente, zu denen Verständnis bezüglich Unsicherheit und Kausalität nötig ist beim Durchführen eines Randomisierungstests. Anhand der Auswertungen eines Vor- und Nachtests von sechs Studierenden geben sie Hinweise, wie Lernsituationen besser gestaltet werden können, um in die Inferenzstatistik und speziell in das Schlussfolgern auf mögliche Ursachen einzuführen.

Das Schema aus Pfannkuch et al. (2011) wird weiter entwickelt zu „(1) thinking about the data obtained from an experiment; (2) conducting the randomization test by modeling random behavior; and (3) making a claim about the data" (Pfannkuch et al. 2015, 99). Diesen drei Schritten ordnen die Autoren sechs Konzepte zu, die sie für die jeweiligen Schritte als notwendig ansehen. Diese sechs Konzepte werden von ihnen als theoretisches Framework entwickelt, dabei bauen sie auf den vier von Konold und Kazak (2008) identifizierten Konzepten für „sample-to-population inferences" auf und ergänzen diese, um dem von ihnen so genannten „experiment-to-causation inference" Ansatz zu entsprechen. Die Konzepte sind *Model Fit; Signal-Noise; Principles of Experimental Design and Causation; Distribution; Law of Large Numbers, Random Process and Independence* und *Inference Argumentation*. Sie werden den drei Schritten zugeordnet wie in Tabelle 11 zu sehen.

Das Verständnis der Autoren zu diesen sechs Konzepten wird im Folgenden kurz ausgeführt.

Tabelle 11: Framework der zugrundeliegenden Konzepte (entnommen aus Pfannkuch et al. 2015, 100)

Thinking about the data	Modeling random behavior	Making a claim about the data
Model Fit	Model Fit	Signal-Noise
Signal-Noise	Distribution	Inference Argumentation
Principles of Experimental Design and Causation	Signal-Noise	Principles of Experimental Design and Causation
	Law of Large Numbers, Random Process, and Independence	

Bei dem Konzept *Model fit* gehen sie davon aus, dass zu einer durchzuführenden Studie im Vorfeld schon Erwartungen bzw. Modelle vorliegen bezüglich des Erfolgs einer Behandlung in vergleichenden Experimenten. Auch Schüler oder Studierende hätten solche Erwartungsmodelle. Die beobachteten Daten eines Experiments müssten sich nun diesen inneren Modellen stellen. Zu einem Randomisierungstest hätten Lernende ebenfalls ein Erwartungsmodell bezüglich der Referenzverteilung, nämlich „They typically expect to see a unimodal, symmetric distribution that is centered at zero“ (Pfannkuch et al. 2015, 101). Ob dies tatsächlich so ist, wird nicht ausgeführt, es wird auch keine Studie zitiert, in der die Erwartungen an die Referenzverteilung von Lernenden erforscht werden. Auch bezüglich des aufgestellten „probability models“ hätten Lernende Erwartungen, dies wird nur konkretisiert mit „expectation of distributional features and expectation of the possible range of chance variation alone“ (Pfannkuch et al. 2015, 101). Es ist nicht klar, wie diese Erwartungen gemessen werden und ob diese von Lernenden tatsächlich konkret ausformuliert werden. Grundsätzlich ist aber mit *Model Fit* gemeint, dass zu bestimmten Abläufen oder Ergebnissen im Vorfeld bereits innere Modelle bei Lernenden vorliegen und dass diese nun hier mit einem Randomisierungstest überprüft werden.

Zu dem Konzept *signal-noise* (basierend auf der Arbeit von Konold & Kazak 2008) stellen die Autoren die These auf, dass dieses Konzept an drei Stellen im Inferenzprozess auftaucht. Als erstes in den beobachteten Daten, die hauptsächlich aus *noise* bestünden und vielleicht einem *signal*, wenn denn ein Effekt der Behandlung beobachtbar ist. Falls nur der Zufall herrsche, dann wäre dies die Ursache für beobachtete Unterschiede und die Daten wären komplett *noise*. Alternativ könne die Ursache für beobachtete Unterschiede im Zufall und in der Behandlung liegen, dann hätten die Daten sowohl *signal* als auch *noise* (vgl. hierzu die beiden Erklärungen für beobachtete Daten aus Budgett et al. 2012). Als zweites müsse dieses Konzept Anwendung finden beim Analysieren der Referenzverteilung, ob der Effekt der Behandlung als *signal* in den Daten erkennbar ist, die unter Annahme des Zufalls als komplettem *noise* entstanden sind. Schließlich müsse der P-Wert (als *tail proportion*) mit diesem Konzept interpretiert

werden. Wenn die Behandlung effektiv ist, dann liege eine Komposition aus *signal and noise* vor, andernfalls ließe sich keine Schlussfolgerung ziehen, weil innerhalb des *noise* kein *signal* entdeckt werden könne.

Im Konzept *Principles of experimental design and causation* geht es um das Verstehen des hinter dem Experiment liegenden Designs. Neben dem Verständnis zu Blindversuchen, Placebogaben etc. sehen die Autoren vor allem das Verständnis über die zufällige Zuordnung zu den Gruppen als wichtig an und führen hierzu drei Erklärungen an, die im Vergleich zu den bisherigen Arbeiten weiter entwickelt wurden.

> „(1) eliminating bias that may result if the researcher assigns the units; (2) balancing the groups on variables that are known to affect the response; and (3) balancing the groups on confounding variables that may be unknown to the researchers." (Pfannkuch et al. 2015, 102)

Auch dass meistens Freiwillige als Versuchspersonen genommen werden, soll hier explizit bedacht werden, da dadurch keine Verallgemeinerungen auf breitere Populationen möglich seien.

Für das Konzept *Law of Large Numbers, Random Process and Independence* sehen die Autoren als wichtig an, dass Lernende durch die Verwendung der VIT-Software erkennen, dass die Unabhängigkeit zwischen jedem einzelnen Versuch dadurch erreicht wird, dass das „label" (also die Gruppenzugehörigkeit) entfernt wird und dann die Daten neu auf zwei Gruppen aufgeteilt werden. Weiter soll erkannt werden, dass die Unterschiede zwischen den neu zugeordneten Gruppen sich kontinuierlich ändern, weil ein Zufallsprozess zugrunde liegt. Und letztlich, dass das Gesetz der großen Zahlen Anwendung findet durch die 1000fache Wiederholung des Prozesses und sich die entstehende Referenzverteilung stabilisiert. Es wird darauf hingewiesen, dass das Gesetz der großen Zahlen nicht explizit in der Unterrichtssequenz erwähnt wird. Hier stellt sich die Frage, aus welchem Grund diese Verbindung nicht gezogen wird und ob dies das Lernen nicht erleichtern würde.

Unter dem Konzept *Distribution* fassen die Autoren die konzeptuelle Einheit, die aus einzelnen Fällen in einer Repräsentation gebildet wird und führen dies mit dem Gedankengut des Randomisierungstests zusammen. „The reasoning is centered on the idea that the group taken as a whole may benefit from the treatment but it may not be the case for every unit" (Pfannkuch et al. 2015, 103). Für die Referenzverteilung nennen sie weiterhin, dass verstanden werden müsse, was jeder einzelne Datenwert bedeute. Vor allem die „tail proportion" müsse in der Referenzverteilung mit Bezug zur gesamten Verteilung verstanden werden und mit einem stochastischen Modell, also der Referenzverteilung als mögliche Ergebnisse eines Zufallsvorgangs, verbunden werden.

Schließlich im Konzept von *Inference argumentation* müsse Verständnis über die Logik des indirekten Beweisens vorliegen. Verbunden damit soll Missverständnissen beim Interpretieren des P-Werts vorgebeugt werden. Um diesen Missverständnissen zu begeg-

nen, nutzen sie einen anderen Begriff, statt P-Wert reden sie ausschließlich von „tail proportion“ und statt Nullhypothese nutzen sie ausschließlich die Sprechweise „chance is acting alone“ (Pfannkuch et al. 2015, 104).

Die Nutzung dieser Konzepte im Rahmen der drei Schritte des Schemas wird ausführlich von sechs Studierenden im Rahmen eines Interviews analysiert. Dabei stehen diese Studierenden repräsentativ für 60% der 2000 Studierenden, die an dem Kurs und den zwei 50-minütigen Einheiten zum Randomisierungstest teilgenommen haben. Ausgewertet werden der Vor-/Nachtest aller Teilnehmer (mit n = 1886 Teilnehmern im Vortest und n = 868 Teilnehmern im Nachtest) und ein sich daran anschließendes Interview, das mit den sechs oben erwähnten Teilnehmern vom selben Interviewer durchgeführt wurde.

Das Verständnis der Lernenden zu den sechs Konzepten des theoretischen Frameworks (Tabelle 11) wird anhand des Vor-/Nachtests der Studie analysiert mit dem Ziel, empirisch ein Framework zu entwickeln, in welchem Elemente angeführt werden, die bei „experiment-to-causation inferences“ benötigt werden. Die erste Frage sowohl des Vor- als auch des Nachtests zur Fischöl-Aufgabe lautet: „Write down the TWO MAIN possible explanations for this observed difference as shown in Figures 2 and 3“ (Pfannkuch et al. 2015, 126, vgl. hierzu Abbildung 157). Hierbei erwarten die Autoren die beiden folgenden Erklärungen: „(1) the treatment is effective, and (2) chance is acting alone“ (Pfannkuch et al. 2015, 107). Das Vorgehen der Autoren ist, dass sie die Antworten aller Teilnehmer einordnen, aber nur diejenigen der sechs ausgewählten Studierenden näher betrachten.

Die mögliche Erklärung, dass die Behandlung effektiv ist, nennen im Vortest 50% aller Teilnehmer (n = 1886) und im Nachtest 69% der Teilnehmer (n = 868). Als zufriedenstellende Formulierungsbeispiele geben sie „Fish oil replacing regular oil really reduces blood pressure“, „A fish oil diet does reduce blood pressure more than a regular diet“ und „The fish oil diet is effective in lowering blood pressure“ (Pfannkuch et al. 2015, 107). 78,8% der Teilnehmer geben eine zweite mögliche Erklärung im Vortest an. Nur 5% hiervon erwähnen den Zufall als mögliche Ursache. Es werden (ohne Quantifizierung) häufig Erklärungen genannt, die letztlich das Design der Studie außer Acht lassen und auf der Idee konfundierender Variablen basieren. So hätten beispielsweise die Versuchsteilnehmer der Fischöl-Gruppe mehr Sport gemacht als die anderen und deshalb ihren Blutdruck gesenkt, ist der Inhalt einer Studierendenantwort. Außerdem gab es Bedenken hinsichtlich der Gruppengröße (zur Erinnerung: in beiden Gruppen waren jeweils sieben Versuchspersonen), was an den Beispielen von S1: „Samples are too small and the observed difference is a result of biased sampling or big sampling error“ und S3: „External factors are influencing the results due to a small sample size“ (Pfannkuch et al. 2015, 107) verdeutlicht wird. Auf die Stichprobengröße des Experiments wird we-

der während der Beschreibung der Lernsequenz eingegangen, noch findet sich etwas in den Analysen. Im Nachtest geben 93% der Teilnehmer eine mögliche zweite Erklärung an, wovon 61% die gewünschte „chance is acting alone“ Erklärung nennen. Hier geben die Autoren jedoch zu, dass die schriftlichen Antworten häufig nur kurz aus „chance“ oder „chance is acting alone“ (Pfannkuch et al. 2015, 108) bestehen, so dass eine Interpretation bezüglich des Verständnisses der Lernenden schwierig sei. Jedoch sei nun der Zufall Bestandteil des Vokabulars und des Denkens, welches für sich ein positives Ergebnis sei.

Für den zweiten Schritt *conducting the randomization test by modeling random behavior*, in welchem der Randomisierungstest mit Hilfe einer Simulation durchgeführt werden soll, erstellen die Autoren eine Liste mit Elementen, die zur Argumentation während dieses Schrittes nötig sind. Dies sind die Schritte T1 bis T5, zu sehen in Tabelle 12.

Tabelle 12: Nötige Schritte für Aktion 2: *conducting the randomization test by modeling random behavior* (entnommen aus Pfannkuch et al. 2015, 112)

Element	Beschreibung
T1: Purpose of test	Test observed difference if chance is acting alone (Assumption treatment has no effect)
T2: Simulating random reallocation, uncertainty	Notion of randomness of who gets into which group
T3: Measuring under uncertainty	Record and interpret differences in center
T4: Distribution of possible measures under uncertainty (probability model)	Repeat T2 and T3 many times (repeatable process) Build distribution of a statistic (interpret what is measured)
T5: Quantification of uncertainty (tail proportion)	Purpose of putting observed difference on re-randomization distribution Read the tail proportion

An kurzen Beispielen führen die Autoren aus, wie die Lernenden die einzelnen Elemente durchführen. Hier geht es vor allem um den Gebrauch der Software, die einen Randomisierungstest in wenigen Klicks nahezu automatisch durchführt und auch den P-Wert (als tail proportion) direkt anzeigt. Das zusammenfassende Fazit hierzu lautet: „we believe that the dynamic imagery of the VIT software facilitated the ability of students to recall many of the ideas underpinning the randomization test” (Pfannkuch et al. 2015, 113).

Für die Begründungen, die zum dritten Schritt *making a claim about the data* des Schemas gehören, werden nur einige kurze Beispiele genannt. Keiner der Teilnehmer legt sich auf eine Interpretation fest, sondern formuliert die Unsicherheit bezüglich eines möglichen Schlusses mittels des gefundenen P-Werts von 0,5% „The Fishoil diet probably lowers blood pressure“, „it is highly probable that there exists a relationship

between fishoil and blood pressure“ (Pfannkuch et al. 2015, 115). Die Teilnehmer der Studie argumentieren im Wesentlichen damit, dass die Gruppengröße des Experiments zu klein sei und deshalb keine sicheren Aussagen gemacht werden können oder mit konfundierenden Variablen, die ebenfalls einen Einfluss auf den Effekt haben können. Bei der Interpretation eines weitere P-Werts von 30%[37] finden sich Aussagen wie „chance is probably acting alone“ (Pfannkuch et al. 2015, 115) – was eine Tendenz zeigt, die Nullhypothese fälschlicherweise anzunehmen. Aber auch „the researchers could not conclude anything“ wird als Interpretation genannt, allerdings mit der Argumentation, „there is a 30% chance that chance is acting alone” (Pfannkuch et al. 2015, 115) – womit eine falsche Hypothesenwahrscheinlichkeit anzugeben versucht wird. Insgesamt werden die Antworten aller Teilnehmer ausgewertet (n = 865) und vier Elemente herausgearbeitet, die in einer Begründung adressiert werden. Diese sind „Rare occurrence (R), Causal evidence (C), Tendency (T), and Generalization (G)“ (Pfannkuch et al. 2015, 116). Als erwünschtes Beispiel hierfür führen die Autoren an:

> „We are pretty sure (R) that a fish oil diet causes (C) males with high blood pressure (G) to tend (T) to have a higher reduction in blood pressure than those on a regular oil diet. We need to be careful about generalizing beyond the group in the study (G). (Pfannkuch et al. 2015, 116)

Aus all den Auswertungen schließen die Autoren auf zwölf Elemente, die während einer Lerntrajektorie adressiert werden sollen. Die folgende Tabelle zeigt die Auflistung dieser Elemente.

Tabelle 13: Zusammenfassung von Elementen, die bezüglich Unsicherheit während eines Randomisierungstests adressiert werden sollten (entnommen aus Pfannkuch et al. 2015, 118)

Causal evidence	Understanding that in a properly executed randomized comparative experiment causality can be established if the values of the explanatory variable (treatment) are randomly assigned to the units.
Randomization Test	Understanding the purpose of the test and reasoning and ideas underpinning the quantification of uncertainty towards experiment-to-causation inference
Tail Proportion	Understanding that the aim is to detect a signal, the treatment effect, under the obscuring effects of noise or chance variation. A small tail proportion indicates a signal has been detected, while a large tail proportion indicates a signal has not been detected suggesting that noise could be obscuring the signal or there could be no signal, just noise, implying that a claim cannot be made.
Treatment is effective	Understanding that the treatment is effective element is composed of a chance component and a treatment effect component.
Rare occurrence	Realizing the possibility, although small, that a difference in centers at least as large as that observed could happen by chance alone. That is,

[37] Die genaue Frage hierzu wird nicht berichtet.

	the observed difference may be a rare occurrence and the wrong inference may have been made (Type 1 error—not covered in our two-lecture introductory instruction).
Generalization	Understanding that care must be taken with any generalization to a wider group than those in the study who were volunteers with particular characteristics (e.g., male, high blood pressure). The population is all those who participated in the experiment. Inappropriate to think about a wider population.
Tendency	Understanding that the inference is about the tendency of the treatment group as a whole to improve, not every individual.
Confounding variables	Understanding that unknown or potential confounding variables can be treated as chance explanations, which are accounted for in the method of random assignment and in the re-randomization distribution.
Design issues (e.g., group size)	Realizing a design issue such as group size is not a problem. Understanding that smaller group sizes require a large observed difference in centers in order to detect whether the treatment is effective under the obscuring effects of chance variation compared to larger group sizes.
Sample-to-population inference	Realizing that a designed experiment uses volunteers, does not take a sample from the population, and does not aim to make an inference about a population; rather it aims to make an inference about an intervention.
Contextual knowledge	Understanding that claims are based on the data in hand and that contextual knowledge, for example, about possible biological mechanisms for the observed difference in centers is used for the next stage of an investigation. Realizing that one's own contextual knowledge and beliefs can bias perceptions or leads one's thinking astray.
All knowledge is uncertain	Acknowledging that there are other sources of uncertainty such as quantification of uncertainty for statistical inference as well as the uncertainty about current knowledge being overturned in the future.

Mit der Einstufung der Interaktionen aus Budgett et al. (2012) und den hier definierten Elementen ergibt sich für die vorliegende Arbeit ein Katalog an Elementen, die in einer Lerneinheit aufgenommen werden können. Hiermit enden die Forschungsarbeiten der neuseeländischen Autorengruppe zu Randomisierungstests.

Auf Basis der Ideen und der VIT Software von Pfannkuch et al. (2011) und Budgett et al. (2012) stellen Fitch und Regan (2014) kurz dar, dass sie in ähnlicher Weise über Randomisierungstests in die Inferenzstatistik einsteigen. Nach einer sehr kurzen Darstellung leiten sie einige konkrete Hinweise für das Unterrichten ab, die für die Zwecke dieser Arbeit nützlich sind:

1. „Use hands-on simulations before introducing computer simulations. The randomisation method lends itself to tactile and visual experiences.

2. Use software which presents visualisations of simulations and take time to explicitly link the software visualisation and the hands-on activity so that students understand what the software visualisations are showing. See Recommendation 4. below.
3. Demonstrate what is meant by "chance alone". e.g., use the "Randomisation Variation" module in VIT to show visually how random allocation alone can generate apparent differences. Budgett et al. (2013) reported that some students have difficulty in understanding the concept of "chance alone".
4. Distinguish clearly between the observed data and the data simulated under chance alone." (Fitch & Regan 2014, 3)

Eine Forschungsarbeit, die sich mit Einstellungen zu Statistik von Studierenden, die an einem Randomisierungstest-Curriculum teilgenommen haben, befasst, ist diejenige von Swanson et al. (2014). Diese Autoren vergleichen die Einstellungen von Studierenden in einem traditionellen Einführungskurs in Statistik (Stats 101), in welchem formale Methoden zur statistischen Inferenz im letzten Drittel thematisiert werden, mit einem umgestalteten Kurs, bei dem über den Simulationszugang Themen der statistischen Inferenz über Randomisierungstests von Beginn des Kurses an unterrichtet werden. Insgesamt zeigt sich beim Vergleich, dass im Randomisierungstestkurs eine stärkere Verbesserung in kognitiven Kompetenzen stattfindet, eine Zunahme im Bereich der Wahrnehmungen über die Schwierigkeit der Statistik, eine geringere Abnahme des Interesses und eine geringere Abnahme des wahrgenommenen Wertes der Statistik im Vergleich zu denjenigen, die am traditionellen Kurs teilgenommen haben. Es zeigt sich, dass der neue Zugang keine wesentlichen Verbesserungen in den Einstellungen der Studierenden zeigt, positiv ausgedrückt sich jedoch auch nicht schlechter auf die Einstellungen zur Statistik auswirkt als der traditionelle Kurs. Statistisch signifikante Unterschiede ergeben sich jedoch nicht zwischen den beiden Kursen. Die Autoren halten fest, dass

> „we do not observe substantial differences between the attitudes of students in randomization-based curricula as compared to students taking the consensus curriculum. This suggests that the randomization-based curriculum is neither a panacea nor does it harm student attitudes any differently than does the traditional curriculum." (Swanson et al. 2014, 5)

Die Autoren räumen jedoch ein, dass die Lehrenden des Randomisierungstestcurriculums selbst wenig Erfahrung hatten und deshalb die Ergebnisse durch weitere Studien gestützt werden sollten.

Rossman und Chance (2014) haben ebenfalls eine Lerntrajektorie zum Einstieg in Inferenzstatistik über Randomisierungstests entwickelt und berichten allgemein über einige Schwierigkeiten, die bei ihnen im Kurs auftreten. Da gibt es zum einen die Lernenden, die bis zum Ende des Kurses nicht verstehen, was der Sinn der Simulationen ist: „Some [students] mistakenly believe that simulation aims to provide replication of the research study, in order to strengthen the findings through replication." (Rossman & Chance 2014, 218). Außerdem erwähnen sie noch, dass „defining the parameter of interest [of a

randomization test] can be very nontrivial for students" (Rossman & Chance 2014, 218).

Biehler et al. (2015) stellen ebenfalls die Analyse von Studierendenbearbeitungen zu Randomisierungstests vor. Die Autoren entwickeln ein Sechs-Schritte-Schema zur Durchführung eines Randomisierungstests und nutzen dies als Auswertungswerkzeug zur Analyse von Studierendenaktivitäten. Es wird eingesetzt während Interviews, die am Ende zweier Stochastikvertiefungskurse an der Universität Paderborn stattfanden. Daran teilgenommen haben Studierende der Grund-, Haupt-, Gesamt- und Realschule, nachdem sie die Vorlesung „Elementare Stochastik" besucht haben. In den beiden Kursen werden Simulationen mit TinkerPlots behandelt und jeweils am Ende in die Inferenzstatistik über Hypothesentests und schließlich Randomisierungstests eingeführt. Die Teilnehmer führen in Zweiergruppen selbständig einen Randomisierungstest durch (Abbildung 158).

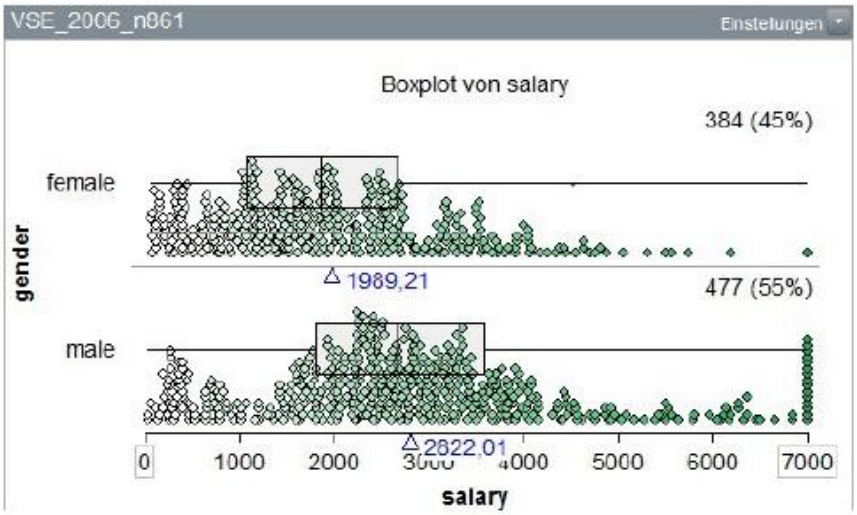


Figure 5.1. Group comparison for the VSE task.

In the dataset (see above) you can see the monthly salaries of 861 women and men of the year 2006. The display suggests that women are way behind men concerning their salary. Someone argues against the result of the group comparison between women and men that only 861 employees were asked. Therefore the differences could have emerged due to the selection of our sample.

YOUR TASK: Now check if there is evidence against the assumption that there is no difference between women and men in the population with regard to their average salary. (This would mean that we can expect similar differences for all employees)

Abbildung 158: Randomisierungstestaufgabe zum Gender Pay Gap, entnommen aus Biehler et al. (2015, 137)

Zur Bearbeitung der Aufgabe erhalten die Studierenden einen Randomsierungstestplan, den sie ausfüllen sollen. Dieser Plan enthält die folgenden Schritte und wurde bereits vorher im Unterrichtsgeschehen eingesetzt:

1. „Observation
   What is the difference in the dataset?
2. Null hypothesis
   Formulate the null hypothesis, that will be assumed as true.
3. Simulation of null hypothesis
   Describe the modeling process, modeled with the sampler of TinkerPlots.
4. Test statistic
   Define the test statistic.
5. P-Value
   Read off the p-value.
6. Conclusions
   Draw conclusions concerning your null hypothesis with the p-value." (Biehler et al. 2015, 141)

Zur Analyse der Bearbeitungen dieser Aufgabe durch die Studierenden verwenden Biehler et al. (2015) die folgenden Kategorien, abgeleitet aus dem an die Studierenden gegebenen Randomisierungstestplan.

„Step 1 Reading off the difference of the means of the groups in the dataset
Step 2 Formulating an adequate null hypothesis
Step 3 Describing the null model
Step 4 Formulating the test statistic
Step 5 Determining the p-value
Step 6a Drawing statistical conclusions from the p-value
Step 6b Drawing contextual conclusions from the p-value" (Biehler et al. 2015, 143)

Das Schema wird als ja/nein-Schema verwendet. In jedem Schritt wird die zugehörige Aktivität überprüft, ob die Lernenden sie erfolgreich ausführen oder nicht. Im ersten Schritt soll ermittelt werden, wie groß der Unterschied der arithmetischen Mittel im gegebenen Datensatz zwischen Männern und Frauen ist. Dieser Unterschied liegt bei 833€, die Männer mehr verdienen als Frauen. Im zweiten Schritt soll die Nullhypothese formuliert werden. Als mögliche Lösung nennen die Autoren „H_0: There is no difference between the average salary of men and women in the population" (Biehler et al. 2015, 139). Im Nullmodell soll die Unabhängigkeitsannahme beschrieben werden. Hierzu wird festgehalten, dass „The simulation model needs to include a sampling device (e.g., mixer) for each attribute (gender and salary). The sampling from each device needs to be independent and without replacement" (Biehler et al. 2015, 144). Die Teststatistik soll als Differenz der Mittelwerte zwischen Männern und Frauen festgelegt und gesammelt werden. Das Erzeugen einer Referenzverteilung findet sich in dem Schema von Biehler et al. (2015) nicht explizit, wird allerdings benötigt, um im fünften Schritt den P-Wert zu bestimmen. Dieser beträgt ca. 0,0000%. Im letzten Schritt schließlich

müssen daraus Schlussfolgerungen gezogen werden. Als mögliche korrekte Lösung für Schritt 6a wird hier „With a p-value of approximately 0%, we can argue against the null hypothesis and infer that there is an effect of gender, in which case, men earn, on average, more than women.” genannt (Biehler et al. 2015, 146). Von den insgesamt 18 Paaren, die an der Studie teilgenommen haben, haben 94% den ersten Schritt, das Bestimmen des Unterschieds in den beobachteten Daten, und 89% den zweiten Schritt, das Aufstellen der Nullhypothese, erfolgreich durchgeführt. Die Beschreibung der Nullmodelle in Schritt 3 ließ sich laut den Autoren nicht mit ihrem Schema auswerten, da hier eine reine ja/nein- bzw. richtig/falsch-Kategorisierung nicht möglich war. Gleiches gilt für die Formulierung der Teststatistik in Schritt 4. 78,3% der Teilnehmer ermittelten den richtigen P-Wert und 72% haben als richtige Schlussfolgerung gezogen, dass starke Evidenz gegen die Nullhypothese vorliegt. Das Formulieren einer richtigen Schlussfolgerung im Kontext geschieht von 61% der Teilnehmer. Wobei die Qualität der Schlussfolgerungen nicht weiter unterteilt wird.

Es wird über folgende Schwierigkeiten berichtet: Die Formulierung der Schritte 3 (Beschreiben des Nullmodells) und 4 (Formulieren der Teststatistik) geschieht sehr unterschiedlich durch die Teilnehmer, so dass keine Entscheidung über erfolgreiches Durchführen getroffen wird. Die Formulierung der Nullhypothese in Schritt 2 wird als eine Schwierigkeit genannt, so formuliert beispielsweise ein Paar „Men earn more than women on average“ (Biehler et al. 2015, 147), was einer Formulierung der Alternativhypothese entspricht. Trotzdem verwendet dieses Paar in der weiteren Arbeit ein Nullmodell, das Unabhängigkeit zwischen den Merkmalen Geschlecht und Verdienst ausdrückt, also Bezug nimmt auf eine „richtige“ Nullhypothese. Als schwierig entpuppt sich auch das Auffinden des P-Werts in der Referenzverteilung, was allerdings im gewählten Beispiel daran liegen mag, dass der P-Wert 0,0000 beträgt und der beobachtete Wert wesentlich extremer ist als jeder unter der Nullhypothese simulierte Fall. In der durch TinkerPlots automatisch erzeugten Achse der Referenzverteilung kommt der beobachtete Wert nicht vor.

Die Simulation wird mit TinkerPlots durchgeführt. Hierzu entwickeln die Autoren ebenfalls ein Schema, das die durchzuführenden Schritte in der Software auflistet. Diese sind mit TP1 bis TP7 bezeichnet.

„TP1 Populating the mixers with the correct labels/values to mimic the original sample.

TP2 Setting the number of repetitions (how many cases should be randomly selected from each mixer) to the original sample size.

TP3 Setting the number of repetitions (how many cases should be randomly selected from each mixer) to the original sample size. [sic]

TP4 Plotting the new, randomized sample and depicting the measure of deviation from the null hypothesis (e.g., mean difference) in the plot.

TP5 Collecting the chosen measure from many different rerandomizations using “Collect Statistic” and the history function.

TP6 Plotting the collected statistics to examine the distribution of the "test statistic".
TP7 Computing the p-value" (Biehler et al. 2015, 147)

Bezüglich der in TinkerPlots durchzuführenden Schritte TP1 bis TP 7 wird als erstes berichtet, dass es bereits beim ersten Schritt, dem Befüllen der Zufallsmaschine vorkommt, dass jeweils ein Anteil von 0,5 Männern und Frauen verwendet wird, was nicht den realen Daten entspricht. Außerdem wird von den Studierenden teilweise mit Zurücklegen gezogen, so dass weder die Ausgangsdaten randomisiert werden (durch die nicht korrekte Anzahl von Männern und Frauen), noch ein Randomisierungstest durchgeführt wird. Ansonsten wird die Software TinkerPlots auf der technischen Ebene von den Lernenden gut beherrscht. Bezüglich der beiden Schwierigkeiten zur Befüllung der Zufallsmaschine lässt sich eher schließen, dass hier die Schwierigkeit im Verständnis zum Nullmodell besteht.

In einer weiteren Analyse werden einzelne Transkriptausschnitte analysiert. Dabei ist bei einem Pärchen auffällig, dass sie, obwohl sie die Nullhypothese falsch formulieren (women earn on average less than men, s. o.), die Zufallsmaschine trotzdem richtig bestücken, also die Merkmale Geschlecht und Einkommen unabhängig voneinander modellieren. Hier zeigt sich, dass die Verbindung zwischen der Nullhypothese und der zu modellierenden Situation für die Studierenden nicht offensichtlich ist. Am Ende revidiert dieses Paar die anfangs aufgestellte Nullhypothese, weil sie einen sehr kleinen P-Wert erhalten haben und anscheinend wissen, dass sie damit die Nullhypothese verwerfen sollen. Deshalb fügen sie ein „not" ein, so dass die Nullhypothese nachträglich lautet „women earn on average not less than men" (Biehler et al. 2015, 157). Insgesamt schließen die Autoren, dass die meisten Studierenden ein gutes Verständnis bezüglich der nötigen Schritte in TinkerPlots haben, „the technical features of TinkerPlots do not seem to be problematic for the preservice teachers" (Biehler et al. 2015, 157), wohl einräumend, dass Randomisierungstests erst am Ende der Kurse thematisiert wurden und schon ein ganzes Semester vorher mit der Software gearbeitet wurde. In keinem der beiden Kurse wurden Hands-on Aktivitäten als Vorbereitung oder Einführung verwendet. Die Autoren schlagen vor, solche Aktivitäten zu implementieren, um besseres Verständnis zu erzeugen. Als weiteren Hinweis nennen sie „that courses need to put more emphasis on relating the statistical and the contextual world" (Biehler et al. 2015, 159), da hier die meisten Schwierigkeiten der Teilnehmer auftreten. Die Nützlichkeit des verwendeten Randomisierungsplans stufen sie als hoch ein, räumen aber auch ein, dass dieser nicht alle Probleme löst:

> „Observations in the video-study and in the two courses made it evident for us that the participants made substantial use of the randomization test scheme when structuring their activities. Nonetheless, there are aspects and misconceptions that cannot be addressed by the use of a randomization test scheme alone" (Biehler et al. 2015, 160).

Man beachte, dass in der Unterrichtssequenz die Vokabeln Nullhypothese, Teststatistik, Referenzverteilung und P-Wert verwendet wurden, ohne dass sprachliche Schwierigkeiten hierzu berichtet werden. Im Vergleich zur Einheit von Budgett et al. (2012) und Pfannkuch et al. (2015), in denen diese Vokabeln explizit zur Vereinfachung ausgelassen wurden, kann also festgehalten werden, dass deren Verwendung in der vorliegenden Studie nicht zu Schwierigkeiten führten.

Frischemeier (2017) beschreibt die Verwendung von Randomisierungstests am Ende eines Statistikkurses, „um die in den Datensätzen (z.B. im MUFFINS-Datensatz) entdeckten Unterschiede (bezüglich der arithmetischen Mittelwerte) zwischen den Verteilungen auf Verallgemeinerbarkeit zu untersuchen“ (Frischemeier 2017, 290). In diesem Kurs wird versucht, innerhalb einer Doppelstunde sowohl das Simulieren mit TinkerPlots, als auch in die Logik des Hypothesentestens mit P-Werten und schließlich in Randomisierungstests einzuführen. Bei der retrospektiven Analyse kommt der Autor zu dem Fazit, dass „eine schnelle und kurze Einführung in Randomisierungstests verständnisorientiert nicht möglich zu sein scheint“ (Frischemeier 2017, 302). Als besonders kritisch stellt er die Schritte „Aufstellen der Nullhypothese“ und „Finden und Interpretieren des P-Werts“ heraus. Als Lösung schließt er eine weitere Sitzung an, in der noch einmal ausführlich Randomisierungstests behandelt werden. Nach eigenen Worten wurden in dieser letzten Sitzung Randomisierungstests sehr dozentenzentriert und schematisch abgehandelt, jedoch schließt er positiv, dass „die Lernenden zurückgemeldet [haben], dass sie nun den Ablauf und die Struktur eines Randomisierungstests verstanden haben“ (Frischemeier 2017, 318). Für eine erneute Durchführung sieht er vor allem das Miteinbeziehen der Teilnehmer als wichtig an. Ähnlich wie bei Pfannkuch et al. (2015) zeigt sich dennoch, dass eine Einführung in Inferenzstatistik in ungefähr zwei Sitzungen möglich ist.

In einer Teilstudie verwendet Frischemeier (2017) das Schema aus Biehler et al. (2015) ebenfalls zur Auswertung eines Randomisierungstests zur Analyse von Studierendenbearbeitungen. Im Rahmen seiner Dissertation hat Frischemeier (2017) ein Seminar für Studierende des Grund-, Haupt-, Real- und Gesamtschullehramts entwickelt, in dem in das statistische Denken eingeführt werden soll. In diesem Seminar werden vielen Datenanalyseprojekte mit der Software TinkerPlots durch Studierende durchgeführt und am Ende findet sich eine kurze Einheit, in der zunächst über Hypothesentests als Fortführung von ausführlich im Seminar behandelten Gruppenvergleichen Randomisierungstests thematisiert werden. Im Anschluss an sein Seminar führt er eine Studie durch, in der von den 22 Teilnehmern neben einer umfangreichen Aufgabe zum Gruppenvergleich auch ein Randomisierungstest selbständig mit der Software TinkerPlots durchgeführt werden soll. Die zu bearbeitende Frage dürfen die Teilnehmer frei in Bezug auf den MUFFINS-Datensatz wählen (Biehler et al. 2003), so gibt es zum Beispiel

Bearbeitungen zu der Frage, „ob die Merkmale Zeit_Lesen und Geschlecht in diesem Datensatz unabhängig sind“ (Frischemeier 2017, 401).

Zur Auswertung wird geprüft, ob die Schritte des Schemas korrekt ausgeführt werden. Dabei wird die übersetzte Version der Schritte verwendet:

> „Schritt 1: Bestimmung des Unterschieds der arithmetischen Mittelwerte der beiden Verteilungen im vorliegenden Datensatz
> Schritt 2: Aufstellen der Nullhypothese
> Schritt 3: Simulation der Nullhypothese (Befüllen der Zufallsmaschine)
> Schritt 4: Definition und Simulation der Teststatistik (und Erstellung der Referenzverteilung)
> Schritt 5: Ablesen des p-Werts
> Schritt 6: Interpretation des p-Werts“ (Frischemeier 2017, 403ff)

Dabei gibt eine Tabelle einen Überblick, wann ein Schritt als korrekt gelöst bewertet wird. Die Definition der einzelnen Schritte mit Beispielen lässt sich dort nachlesen.

Die Häufigkeitsauswertung in Tabelle 14 gibt einen Überblick über die Lösungsraten der einzelnen Schritte. 11 Paare von Studierenden der Lehrveranstaltung haben an der abschließenden Aufgabe teilgenommen.

Tabelle 14: Zusammenfassung der Auswertung der Randomisierungstests aus Frischemeier (2017)

	Schritt 1	**Schritt 2**	**Schritt 3**	**Schritt 4**	**Schritt 5**	**Schritt 6**
Lösungsrate	11 von 11 (100,0%)	8 von 11 (72,7%)	10 von 11 (90,9%9	10 von 11 (90,9%)	6 von 11 (54,6%)	5 von 11 (45,5%)

Auffällig bei der Auswertung ist, dass, obwohl Schritt 2, das Aufstellen der Nullhypothese, nur von 8 Paaren richtig gelöst wurde, trotzdem bei 10 von 11 Paaren die nächsten beiden Schritte als richtig gelöst bewertet werden. Hierzu gibt Frischemeier keine Erklärung, es lässt sich jedoch erschließen, dass in den in Schritt 3 durchgeführten Simulationen eine im Sinne der Aufgabe richtige Nullhypothese modelliert wurde, diese jedoch im Schritt 2 falsch formuliert wurde. Dieses Verhalten wird nicht weiter kommentiert, ist jedoch meiner Meinung nach sehr interessant, da sich hier zeigt, dass das durch die Zufallsmaschine ausgedrückte Modell in gewisser Weise zeigt, dass verstanden wurde, welche Situation modelliert werden soll, dies sprachlich jedoch nicht in der Nullhypothese ausgedrückt wird. Vielleicht ist den Teilnehmern der Studie auch der Zusammenhang zwischen der Nullhypothese und dem zu simulierenden (Null-)Modell nicht klar. Ähnliches ließ sich bereits in Biehler et al. (2015) beobachten.

In der Studie von Frischemeier (2017) zeigt sich weiter, dass das technisch prozedurale Wissen über die Simulationsumsetzung in TinkerPlots den Teilnehmern keine Schwierigkeiten bereitet, 10 von 11 Paaren führen diese Schritte korrekt aus. Frischemeier kommt hier zu dem Schluss, dass Schwierigkeiten vor allem im Bereich des P-Werts liegen, denn „die auftretenden Probleme in Schritt 5 [Ablesen des P-Werts] und Schritt

6 [Interpretation des P-Werts] geben Anlass zur Vermutung, dass die Definition des p-Werts [sic] von einigen Teilnehmern noch nicht verstanden worden ist, da es Schwierigkeiten gab, diesen zu finden und zum anderen diesen zu beurteilen“ (Frischemeier 2017, 407). Sechs der elf Paare konnten den P-Wert korrekt ablesen, fünf davon konnten ihn korrekt mit dem eingeübten Sprachgebrauch der Evidenz interpretieren.

Die vorgestellte Studie von Frischemeier zeigt ein Schema, mit dessen Hilfe sich auswerten lässt, welche Schritte im Prozess des Durchführens eines Randomisierungstests von Studierenden richtig gelöst werden. Für einen Überblick über das Lösungsverhalten ist dies ein nützliches Werkzeug. Es lässt sich damit jedoch wenig darüber aussagen, welche Argumentationsweisen von den Studierenden verwendet werden und welches die tatsächlichen Schwierigkeiten bei den einzelnen Schritten sind. Es muss allerdings gesagt werden, dass dies nur eine Teilstudie von Frischemeier ist, seine Hauptforschung liegt im Bereich des Gruppenvergleichs, so dass für seine Zwecke die grobe richtig/falsch-Auswertung ausreichend ist.

Noll und Kirin (2017) bauen ebenfalls auf dem Framework von Biehler et al. (2015) auf und untersuchen mit diesem Schema, wie College-Studierende aus „introductory statistics classes, a ten-week course designed for students prior to entering our traditional introductory statistics sequence” (Noll & Kirin 2017, 218) einen Randomisierungstest mit TinkerPlots durchführen. Die Studie wird im Rahmen des Abschluss-Assessments durchgeführt, nachdem die Teilnehmer eine zweiwöchige Einführung unbekannten Umfangs in Randomisierungstests und Bootstraps erhalten haben. Die Ergebnisse von acht Teilnehmern in vier Gruppen werden vorgestellt. Als Aufgabe verwenden sie den „Dolphin-Therapy task” (nach Zieffler & Catalysts for Change 2013, 196ff) mit folgender Aufgabenstellung:

> „Swimming with dolphins can certainly be fun, but is it also therapeutic for patients suffering from clinical depression? To investigate this possibility, researchers recruited 30 subjects aged 18-65 with a clinical diagnosis of mild to moderate depression. Subjects were required to discontinue use of any antidepressant drugs or psychotherapy four weeks prior to the experiment, and throughout the experiment. These 30 subjects went to an island off the coast of Honduras, where they were randomly assigned to one of two treatment groups. Both groups engaged in the same amount of swimming and snorkeling each day, but one group (the animal care program) did so in the presence of bottlenose dolphins and the other group (outdoor nature program) did not. At the end of two weeks, each subject’s level of depression was evaluated, as it had been at the beginning of the study, and it was determined whether they showed substantial improvement (reducing their level of depression) by the end of the study (Antonioli and Reveley, 2005). Research Question: Is swimming with dolphins therapeutic for patients suffering from clinical depression? The researchers found that 10 of 15 subjects in the dolphin therapy group showed substantial improvement, compared to 3 of 15 subjects in the control group.
>
> The above descriptive analysis tells us what we have learned about the 30 subjects in the study. But can we make any inferences beyond what happened in this study? Does the higher improvement rate in the dolphin group provide convincing evidence that the dolphin therapy is effective? Is it possible that there is no difference between the two treatments and that the difference observed could have arisen just from the random nature of putting the 30 subjects into groups (i.e., the luck of the draw)?

> We can't expect the random assignment to always create perfectly equal groups, but is it reasonable to believe the random assignment alone could have led to this large of a difference?
>
> **The key statistical question is**: If there really is no difference between the therapeutic and control conditions in their effects of improvement, how unlikely is it to see a result as extreme or more extreme than the one you observed in the data just because of the random assignment process alone?" (Noll & Kirin 2017, 219, Hervorhebung im Original)

Hier werden die Schritte des Schemas aus Biehler et al. (2015) verwendet, um die Bearbeitungen der Teilnehmer zu kategorisieren. Dabei legen die Autorinnen den Fokus auf den Übergang von der statistischen Ebene zur Softwareebene, um genauer zu untersuchen, wie das in der statistischen Ebene aufgestellte Nullmodell in Form der Nullhypothese in der Software umgesetzt wird. Aus diesem Grund werden nur die ersten drei TinkerPlots-Schritte untersucht (TP-Step1: Populating the mixers with the correct labels/values to mimic the original sample, TP-Step2: Setting the number of repetitions (how many cases should be randomly selected from each mixer) to the original sample size, TP-Step3: Setting the number of repetitions (how many cases should be randomly selected from each mixer) to the original sample size [sic] (vgl. Biehler et al. 2015, 148). Als Erläuterung für den Übergang zwischen der Nullhypothese und dem in der Simulation ausgedrückten Nullmodell führen die Autorinnen hier an,

> „we want to think of the sampler devices as randomly allocating a person from one of the therapy groups to a result of improved or not improved and repeating this process until all 30 patients have been allocated a result at random. This random shuffling of group to outcome allows us to see if chance alone could produce the outcomes reported in the original experiment." (Noll & Kirin 2017, 222)

Zu Beginn der Bearbeitung stellen alle vier Gruppen von untersuchten Studierenden eine Zufallsmaschine auf, die im ersten Bauteil die beiden Gruppen Dolphin und Control mit jeweils fünfzehn Einträgen modelliert. Zwei der Gruppen setzen die Modellierung korrekt um, wie in Abbildung 159 zu sehen.

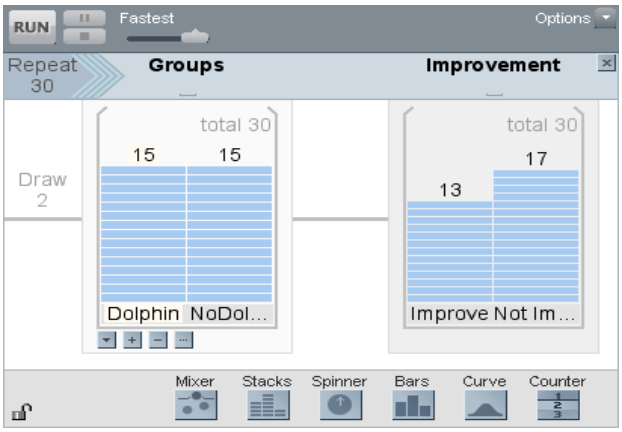


Abbildung 159: Richtig aufgestellte Zufallsmaschine der Teilnehmer der Studie von Noll und Kirin (2017, 226)

Im ersten Bauteil wird die Gruppenzugehörigkeit abgebildet, jeweils 15 Personen waren bei dem ursprünglichen Experiment in der Delphingruppe und 15 Personen waren in der Kontrollgruppe. Im zweiten Bauteil wird unabhängig davon die Anzahl der Personen, bei denen eine Verbesserung auftritt (13) modelliert bzw. die Personenanzahl, bei der keine Verbesserung auftritt (17).

Im Vergleich hierzu findet sich bei zwei der vier Paare eine Zufallsmaschine, wie in Abbildung 160 zu sehen, bei der sowohl die Anzahl der Etiketten für beide Gruppen gleich ist als auch die Anzahl von Improvement/No Improvement.

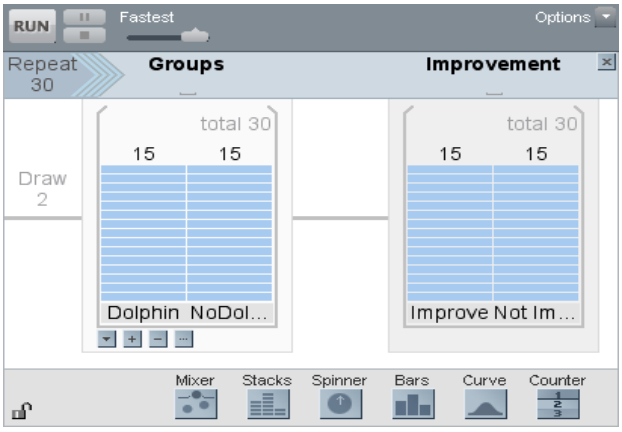


Abbildung 160: Eine der von Studierenden aufgestellten Zufallsmaschinen in Noll & Kirin 2017, 227

Interessant bei der Bearbeitung dieser Gruppe ist, dass die Teilnehmer die Aufgabenstellung zitieren und damit begründen, dass entgegen der Daten des Experiments (für 13 Personen liegt eine Verbesserung vor, für 17 Personen liegt keine Verbesserung vor) das zweite Merkmal der Zufallsmaschine ebenfalls gleichverteilt sein muss.

Kate: Hold on. So it says [she re-reads the problem out loud] if there really is no difference between therapy treatment how likely it is that you would see a result as extreme or more extreme than the one you observed in the data... *So it should be fifty-fifty* [Improved/Not improved]. This should be fifty-fifty.

Joe: The improvement.

Kate: Yeah.

Joe: Why?

Kate: Because *this model is supposed to represent the null hypothesis*.

Joe: Okay.

Kate: This model is supposed *to represent the conception that there is no difference*. (Noll & Kirin 2017, 227)

Nachdem der Interviewer interveniert, verändern die Teilnehmer ihr Modell von 15/15 für Improvement/noImprovement zu 13/17, allerdings schreiben die Autoren hierzu, dass „her responses appeared to be equivalent, suggesting that the distinction between

the two models is not entirely clear to her" (Noll & Kirin 2017, 228). Für diese Gruppe scheint die zufällige Zuordnung der Versuchspersonen zu den Gruppen für den Argumentationsprozess problematisch zu sein, denn „there was no further discussion and no evidence to suggest that they understood the random assignment in their model or how it related to the null hypothesis" (Noll & Kirin 2017, 228). Ähnliche Schwierigkeiten werden auch für das vierte Paar berichtet. Die Autoren beziehen Interpretationen immer wieder auf die Nullhypothese, allerdings werden an keiner Stelle die von den Gruppen aufgestellten Nullhypothesen berichtet, so dass unklar bleibt, inwiefern vielleicht bereits beim Aufstellen der Nullhypothesen Schwierigkeiten bestehen. Hier scheint jedoch ein wesentlicher Knackpunkt der vorliegenden Studie zu liegen. Da die Lernsequenz dem Ansatz von Zieffler und Catalysts for Change (2013) folgt, könnte ein möglicher Erklärungsansatz hierfür in der ausschließlichen Fokussierung beim Formulieren der Nullhypothese mit dem Sprachgebrauch von „no difference" liegen. Da hierzu jedoch keine Details berichtet sind, ist dies meine eigene Interpretation.

Bezüglich der weiteren benötigten TinkerPlots-Schritte schreiben die Autoren, dass es bezüglich der zu wählenden Ziehungsanzahl keine Schwierigkeiten gibt. Dies liegt wohl auch daran, dass die Ziehungsanzahl automatisch auf 2 gesetzt wird, wenn zwei Bauteile hintereinander geschaltet werden. Bezüglich der Wiederholungsanzahl setzen drei der vier Gruppen diese direkt auf 30, als Begründung wird zitiert

> Mike: The repeat is set at thirty to simulate thirty individuals (Noll & Kirin 2017, 232)

Die vierte Gruppe schwankt zwischen einer Wiederholungszahl von 15 oder 30, was allerdings in vorherigen Schritten begründet ist und an der unglücklichen Benennung der Variablen liegt. Nach der Klärung durch die Interviewerin hierzu ist auch für diese Gruppe klar, dass 30 Wiederholungen gewählt werden müssen. Hieraus schließen die Autoren jedoch, dass im Gegensatz zu der Studie von Biehler et al. (2015), „in our study we found that the naming of attributes was actually important for students" – dies wurde in Biehler et al. (2015) als unwichtig angesehen und führte dort auch nicht zu Schwierigkeiten.

> „We argue that the naming of the attributes is an important aspect of the modeling process. For example, the names students give to the attributes can serve as a bridge (or concrete component) between the context world they began in and the more abstract world of the computer within which they are trying to build their model." (Noll & Kirin 2017, 236)

Als letzter Schritt wird das Verhalten der Teilnehmer bezüglich der Einstellung *ohne Zurücklegen* ausgewertet. Drei der vier Gruppen wählen dies sofort, die vierte Gruppe einigt sich nach kurzer Diskussion darauf. Hier allerdings vermuten die Autoren, dass den einzelnen Teilnehmern nicht klar ist, inwiefern dies eine wichtige Entscheidung bezüglich der Modellierung ist. Sie argumentieren

> „but there is no evidence from either of these groups that they understood that setting their samplers up in this way created the random allocation of treatments to responses, rather they appeared to be relying on their memory of the sampler set up from the previous activity" (Noll & Kirin 2017, 235).

Am Ende geben die Autorinnen einen Ausblick darauf, dass sie den Ansatz über das Modellieren per Simulation für erfolgversprechend halten, jedoch mehr Forschungsarbeit hierzu benötigt wird, gerade darauf, wie in TinkerPlots ausgedrückte Modelle von Studierenden interpretiert werden.

> „The models that students create in TinkerPlots are key parts of the statistical content in courses that focus on modeling and simulation. We believe that a modeling and simulation approach using the CATALST materials and TinkerPlots software has great potential in supporting students' development into statistical thinkers and modelers. At the same time, we also see this work as in its infancy in that we need more research focused on why students create the models that they do and what the various features of the TinkerPlots models they create mean to them." (Noll & Kirin 2017, 240)

Wie Noll und Kirin sehe auch ich, dass in diesem Ansatz verbunden mit der Software TinkerPlots Potential für das Einführen in schlussfolgerndes Denken liegt. Gerade TinkerPlots eignet sich gut, um durch die Modellierung der Zufallsmaschine auf nicht versprachlichte Nullmodelle von Lernenden Rückschlüsse ziehen zu können.

6.2 Verschiedene Schemata zu Randomisierungstests

Nachdem in 6.1 verschiedene empirische Studien vorgestellt wurden, in denen Lernende beim Bearbeiten eines Randomisierungstests beobachtet und analysiert werden, folgt hier nun ein Blick in Literatur, in welcher eine normative Perspektive eingenommen wird. Hierzu findet sich wesentlich mehr Literatur, dies und die vorigen Ausführungen sollen einen Überblick über notwendige Konzepte für das Durchführen eines Randomisierungstests liefern. Aus diesem Überblick wird theoriegeleitet ein eigenes Schema entwickelt, zu dem auch potentielle Schwierigkeiten für die einzelnen Konzepte besprochen werden. Wie zuvor wird die Literatur chronologisch aufgearbeitet.

Das Bearbeiten eines Randomisierungstests erfordert bestimmte Abläufe und das Verwenden bestimmter Konzepte. Welche Schritte hierbei aufeinander folgen und wie diese in der Literatur beschrieben werden, wird in diesem Abschnitt nachgegangen.

Da hier das Bearbeiten eines Randomisierungstests ausschließlich mit Software geschieht, spielt stochastische Simulationssoftware eine entscheidende Rolle. Aus diesem Grund wird zunächst ein allgemeines Schema zum Vorgehen bei stochastischen Computersimulationen betrachtet. Bereits 1987 haben Gnanadesikan et al. in ihrem Buch *The Art and Techniques of Simulation* ein acht-Schritte Schema aufgestellt, das in abgewandelter Form immer wieder als Vorbild für allgemeine stochastische Simulationen diente und auch für Randomisierungstests die Grundlage bildet (z.B. Erickson 2006; Maxara 2009). Dieses beinhaltet die Schritte

„Step 1: State the problem clearly
Step 2: Define the key components
Step 3: State the underlying assumptions
Step 4: Select a model to generate the outcomes for a key component
Step 5: Define and conduct a trial
Step 6: Record the observation of interest
Step 7: Repeat steps 5 and 6 a large number of times (at least 50)
Step 8: Summarize the information and draw conclusions" (Gnanadesikan et al. 1987, 3f)

Mit acht Schritten ist dies ein sehr ausführliches Schema, das eine erste gute Orientierung bildet. Nähere Ausführungen dazu lassen sich in Gnanadesikan et al. (1987) nachlesen. Dies wird hier nicht weiter ausgeführt, da die Bezeichnungen einigermaßen selbst erklärend sind und der Fokus hier auf speziellen Schemata zum Randomisierungstest liegen soll. Explizit aus diesem Schema entwickelt 19 Jahre später Erickson (2006) ein Schema, das er generell für Simulationen, aber auch insbesondere als geeignet für Randomisierungstests ansieht. Sein Schema besteht aus fünf Schritten, in denen er bestimmte Konzepte formuliert:

1. „Collect data from the situation of interest – data that seem to reflect some phenomenon.
2. Design a *measure* of that phenomenon that you can calculate from the data. Ideally, this measure is a large number if the phenomenon is strong and present, and small – even zero – when the phenomenon is absent. The value for this measure, using the real data, is the *test statistic*.
3. Simulate the condition of the null hypothesis, and collect those data.
4. Compute the measure from simulated data, and repeat to build up a sampling distribution for the measure in question.
5. Compare the test statistic to the sampling distribution. The empirical *P*-value is the fraction of cases in the sampling distribution that are at least as extreme as the test statistic." (Erickson 2006, 3)

Im Vergleich sieht man, dass Erickson im Wesentlichen den Schritten von Gnanadesikan et al. (1987) folgt, diese allerdings zu fünf Schritten zusammenfasst und präzisere Prozessanweisungen gibt. Dabei weist Erickson auch darauf hin, dass gerade seine Schritte 2 und 3 umfangreicher sind: „both require some craft" (Erickson 2006, 3). Dahinter steht wohl, dass vor allem in Schritt 3 ein Nullmodell, respektive eine Nullhypothese aufgestellt werden und in Simulationssoftware abgebildet werden muss. Bei näherer Betrachtung fehlt bei den fünf Schritten von Erickson allerdings ein wichtiger Schritt am Ende, der klar Bezug nimmt auf das Ziehen von Schlussfolgerungen, also den achten Schritt von Gnanadesikan et al. 1987. Implizit lässt sich erkennen, dass vor oder in Schritt 3 eine Nullyhpothese aufgestellt werden muss, denn deren Bedingungen sollen simuliert werden. Als Maß für die Extremheit der beobachteten Werte verwendet Erickson den P-Wert. In diesem fünfschrittigen Schema werden konkrete Anweisungen gegeben, welche Aktionen durchzuführen sind. Aus Sicht eines Lernenden bietet das Schema eine gute Orientierung, welcher Ablauf einzuhalten ist. Wie in Schritt 2 werden

zum Teil nicht nur Anweisungen gegeben, sondern auch Hinweise, auf was geachtet werden kann, hier zum Beispiel darauf, in welchem Fall eine Teststatistik eher einen großen oder einen kleinen Wert haben sollte.

Noch kürzer zusammenfassend ist das eingangs erwähnte prominente Schema der drei „Rs“ von G. Cobb (2007). Es lohnt sich ein weiterer Blick in den Artikel von Cobb, um seinem Schema der drei Rs folgen zu können, das wegweisend für viele spätere Publikationen ist. Dieses Schema besteht aus den genannten „three Rs of inference: randomize, repeat, reject“ (G. Cobb 2007, 12).

Im ersten Schritt *randomize data production* geht es darum, zu beschreiben wie die beobachteten Daten eines Experiments erhoben wurden und welche Schlüsse später gezogen werden dürfen (vgl. hierzu auch Ramsey & Shafer 2013 und Kapitel 5.4.3). Cobb beschreibt die Reichweite der möglichen Schlussfolgerungen sehr kurz, dass Zufallsstichproben Verallgemeinerungen auf Populationen zulassen und dass zufällige Einteilungen von Versuchspersonen auf Gruppen Schlussfolgerungen bezüglich der Kausalität zulassen. In seinem ersten Schritt steckt also bereits der Blick auf die Interpretation des gesamten Tests. Hinter dem zweiten Schritt *repeat by simulation to see what's typical,* der mit der Erklärung „randomized data production lets you re-randomize, over and over, to see which outcomes are typical, which are not“ (G. Cobb 2007, 12) so einfach erscheint, steht das gesamte Erstellen einer Simulation und damit zugehörig eines Nullmodells, das in einer Nullhypothese ausgedrückt werden muss. Dieser zweite Schritt beinhaltet für den Anwender eines Randomisierungstests also wesentlich mehr (auch mehr Schwierigkeiten), als Cobb in diesem einen Satz zunächst so scheinbar kurz darstellt. Schließlich muss im dritten Schritt *reject any model that puts your data in its tail* eine Schlussfolgerung gezogen werden. Wesentlich bei dieser Formulierung ist vor allem „reject“, wodurch angedeutet wird, dass ein Modell zwar verworfen, aber nicht unbedingt bestätigt werden kann. Cobb verwendet hier eine im Amerikanischen gängige Metapher für den P-Wert. Im *tail*, also im „Schwanz“ oder am Rand einer Verteilung finden sich die Ergebnisse, die genau so extrem oder noch extremer sind als der beobachtete Wert, also die Ergebnisse, die der P-Wert umfasst. Im Deutschen ist mir keine ähnliche, gleich bekannte Metapher bekannt, was auch daran liegen mag, dass das gesamte Konzept des P-Werts weniger verwendet wird. Auch hier wird nicht auf bekannte Schwierigkeiten, zum Beispiel beim Interpretieren eines P-Werts, denn darauf laufen die Ausführungen von G. Cobb (2007) hinaus, eingegangen. Als grundlegendes Gerüst zur Einführung ins Hypothesentesten scheint dieses Schema sehr gut geeignet zu sein, die Logik des Schlussfolgerns in der Inferenzstatistik ist sehr gut zusammengefasst. Allerdings erscheint die Leichtigkeit, mit der Cobb sein Konzept darstellt, bei näherem Hinschauen etwas übertrieben. Das Paper von G. Cobb (2007) liest sich jedoch durchaus provokativ und möchte vor allem anregen, den Einstieg in die beurteilende Statistik zu überdenken.

Im Gegensatz zu Erickson (2006) hat G. Cobb (2007) sein Schema mit drei Schlagworten versehen, die aus Sicht eines Lernenden sehr griffig sind durch die Alliteration. *Randomize – repeat – reject* lässt sich als Kurzform eines Schemas leicht lernen. Hier müssen allerdings die passenden Konzepte verstanden bzw. verinnerlicht sein, um daraus einen kompletten Randomisierungstest durchlaufen zu können. Auch die längere Ausführung *randomize data production*, *repeat by simulation to see what's typical*, *reject any model that puts your data in its tail* setzt meiner Meinung nach gute Kenntnisse zu einem Randomisierungstest voraus. Als Zusammenfassung des Prozesses des Randomisierungstestens ist dieses Schema gut geeignet und lässt sich sehr gut merken. Als Hilfe für die Durchführung erscheint es zu knapp.

Aufbauend auf Cobbs Ideen haben Rossman et al. (2008) einige Module zu Randomisierungstests entwickelt, auf denen ein Zugang zum inferentiellen Denken aufgebaut werden soll. Um den Lernenden den Zugang zu dieser Denkweise bei Randomisierungstests zu erleichtern, schlagen sie ein vierschrittiges Schema vor, das die Logik verdeutlicht: „The logic is built of big pieces, each composed of smaller pieces. My intuition suggests that in a first course, we'd help students most by starting with big pieces and names for them" (Rossman et al. 2008, 6). Dies ist das ausführlichste Schema, das in der gesamten Literatur gefunden wurde, mit vielen Unterschritten und Erklärungen:

1. „Observed Data
    a. **Observed data** (case x variable format) as a set of tickets (label response)
    b. **Data production**: where is the randomization?
2. Null Model
    a. Research question
    b. **Null hypothesis**: There's no difference in the way treatment and control affect the response.
    c. **Conditioning**: Take the set of observed outcomes (response values) and labels as given and fixed.
    d. **Null model**: Tear each ticket in half. Put the labels in one box, outcomes in a second box, mix each box thoroughly, and create random pairs, choosing one label and one outcome per pair, to simulate the data production process. (All possible pairings are equally likely.)
3. Statistical Test
    a. **Tetst statistic**: Either the number of Yes outcomes in the treatment group (initially), or (equivalently and eventually) treatment average (proportion) minus control average (proportion). The difference in group averages will work for all two-sample comparisons. For regression and correlation, there's a different test statistic, but for it, too, the test will be based on tearing tickets in half and re-randomizing the pairing.
    b. **Reference distribution**: Use the null model to re-randomize; for each re-randomization, compute the value of the test statistic.
    c. **Observed value** (of the test statistic)
    d. **P-value**: chance of a test statistic more extreme than the observed value.

   e. **Reject**? Is $p < 0.05$?
4. Scientific Inference
   a. **Reject**? Can we rule out chance as a plausible explanation?
   b. **Conclusion**: What's the scope of inference? If not chance, what explanations remain plausible? (Here we come full circle: Answering the question about scope of inference completes the chain of logic that begins with randomized data production and ends with a conclusion about the applied context.)" (Rossman et al. 2008, 6f)

Dieses Schema scheint sehr gut geeignet als direkte Vorlage für die Verwendung in einer Lehrsituation zur Orientierung der Lernenden. Zu den vier Hauptschritten *Observed Data, Null Model, Statistical Test* und *Scientific Inference* gibt es jeweils zwischen zwei und fünf erklärende Unterschritte, die in Form von Fragen oder Anweisungen weitere Hilfestellungen geben, was genau auszuführen ist. Dabei fokussieren die Autoren auf das Bild des „ticket splitting", das also beobachtete Daten aufgefasst werden als Ticket oder Etikett, das durch den Vorgang des Rerandomisierens der Daten durchgerissen wird und somit die Daten unabhängig von der Gruppenzugehörigkeit aufgefasst werden. Dies wird im Unterschritt 2d *null model* im Schema erläutert. Somit entstehen zwei Variablen, zum einen die beobachteten Daten, zum anderen die ursprünglichen Werte der beiden Gruppen (zum Beispiel Behandlungs- und Kontrollgruppe). Diese werden nun zufällig zueinander gelost. Eventuell problematisch an diesem Bild ist, dass somit quasi zwei „Pseudogruppen" entstehen. Es gibt dann erneut eine Behandlungsgruppe und eine Kontrollgruppe, allerdings haben die mit den ursprünglichen Gruppen des Experiments nichts mehr zu tun, auch wenn sie deren Namen tragen. Hier entsteht möglicherweise ein kognitiver Konflikt bei Lernenden, weil somit die Gefahr besteht, die randomisierten Daten während der Durchführung des Randomisierungstests mit den ursprünglich beobachteten zu verwechseln. Oder das Konzept wird nicht klar, was eigentlich beim Randomisieren der beobachteten Daten passieren soll.

Am Ende verwenden Rossman et al. (2008) den P-Wert für die Beurteilung der Ergebnisse und ergänzen dies durch das Vokabular von G. Cobb (2007) und fragen nach *reject*?, allerdings mit der Frage danach, ob der Zufall als plausible Erklärung ausgeschlossen werden kann.

Zusammengefasst wird Cobbs „core logic of inference" von den Autoren in einer Darstellung, die die drei Rs von Cobb aufnimmt. Abbildung 161 zeigt diese Darstellung, die verdeutlichen soll, dass „Inference is based on randomization" (Rossman et al. 2008, 7). Es geht bei diesem Schaubild also eher um die Verdeutlichung, wie bzw. wieso Schlussfolgerungen gezogen werden können als um eine Zusammenfassung des von ihnen aufgestellten Vier-Schritte-Schemas.

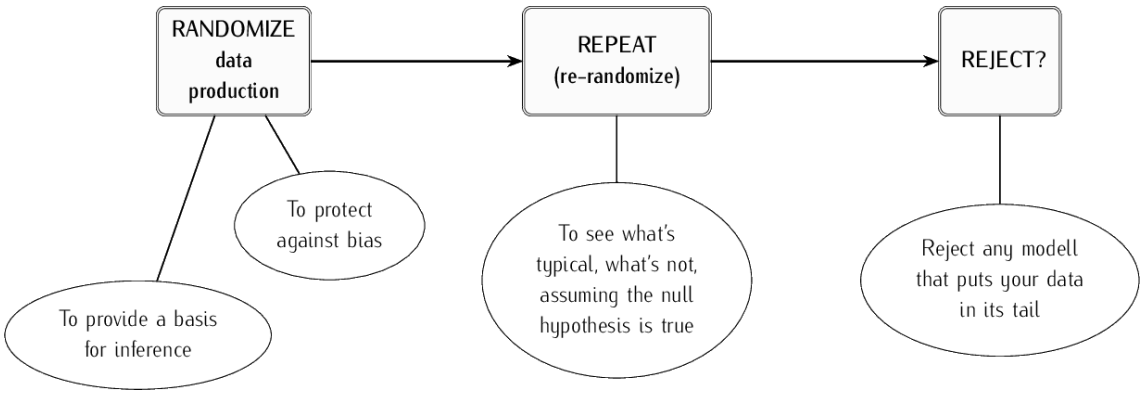


Abbildung 161: Logik des Schlussfolgerns nach Rossman et al. (2008, 8), eigene Darstellung

Herausgestellt wird hier vor allem der Sinn des Randomisierens zu Beginn eines Experiments. Zum einen soll es gegen systematische Fehler (bias) schützen, zum anderen bildet es eben die Grundlage für die spätere Schlussfolgerung. Interessanterweise wird in Abbildung 161 komplett Bezug genommen auf die drei Rs von Cobb, die jedoch in das aufgestellte vier-Schritte-Schema vorher keinen direkten verbalen Einzug gehalten haben. Scheinbar sehen die Autoren hier die Notwendigkeit, ein wesentlich ausführlicheres Schema für die Durchführung des Randomisierungstests zu erstellen, und nutzen für die Logik dann den Ansatz von G. Cobb (2007).

Garfield und Ben-Zvi (2008a) schlagen ebenfalls einen Zugang zum inferentiellen Schließen über Randomisierungstests vor. Hier findet sich auch der Hinweis, dass es zum Zeitpunkt des Erscheinens des Buchs kaum Literatur zu empirischen Studien in diese Richtung gab. Am Ende ihrer ersten Unterrichtseinheit *Learning to reason from data* findet sich ein Vorschlag, Schlussfolgerungen über mögliche Ursachen anhand von Randomisierungstests zu ziehen. Hierfür stellen sie ein fünfschrittiges Schema auf, das sie an einem Beispiel zum Hypothesentesten vorstellen, jedoch auch zum Randomisierungstesten verwenden. Die fünf Schritte dieses Schemas sind:

1. „Assume no one can really tell the difference and is just guessing (…)
2. Generate data based the model that everyone is guessing and can't really tell the difference. How many would they get right just by chance? How many would they get wrong?
3. Compute the proportion of correct guesses for each sample
4. Graph these proportions (show the distribution)
5. Show your result compared to the distribution of guesses. Does it seem likely that you were just guessing?" (Garfield & Ben-Zvi 2008a, 141)

Auch wenn die Autoren explizit sagen, dass dieses Schema zum Randomisierungstesten verwendet werden soll, so ist es doch eher ein allgemeines Schema zum Hypothesentesten, das am Beispiel eines typischen Geschmackstests erklärt wird. Die spezifischen Eigenheiten beim Randomisierungstest, wie zum Prüfen, ob randomisierte Daten vorliegen und welche Schlüsse deshalb am Ende gezogen werden dürfen, fehlen hier. Im ersten Schritt dieses Schemas finden sich die Schritte 1 bis 4 von Gnanadesikan et al.

(1987) wieder. Vor allem wird hier implizit die Nullhypothese aufgestellt und mit dem Vokabular „just guessing“ oder „just by chance“ belegt. Dass der Schritt 3 dieses Schemas durch Simulationen durchgeführt werden soll, wird durch den Kontext klar. Dieses Schema beinhaltet, ähnlich wie das von Erickson (2006), kurze Anweisungen, welche Aktion durchgeführt werden soll. Gleichzeitig werden Fragen gestellt, die im Laufe des Prozesses durchdacht und beantwortet werden müssen. Dies geschieht wohl in der Hoffnung, dass dadurch klarer wird, was in den einzelnen Schritten zu tun ist.

Erkenntnisse zu Bearbeitungen durch Lernende werden hier nicht geliefert. Im Gegensatz zu fast allen anderen Ausführungen in ihrem Buch führen die Autoren zum Thema Randomisierungstests an, dass sie dieses Konzept wohl als nützlich betrachten, hierzu jedoch keinerlei Studien oder konkrete Aufgaben vorstellen können. Allerdings findet sich später in ihrem Buch, mit losem Bezug zu Randomisierungstests, ein allgemeines Schema zum Gebrauch von Modellen in der Statistik, welches für die vorliegende Arbeit als nützlich eingestuft wird. Abbildung 162 zeigt einen Ausschnitt dieses Modells, der sich auf das vorgestellte Schema zum Hypothesentesten durch Simulation bezieht. Der Rest des Modells bezieht sich auf reine Datenanalyse und wird deshalb hier nicht betrachtet.

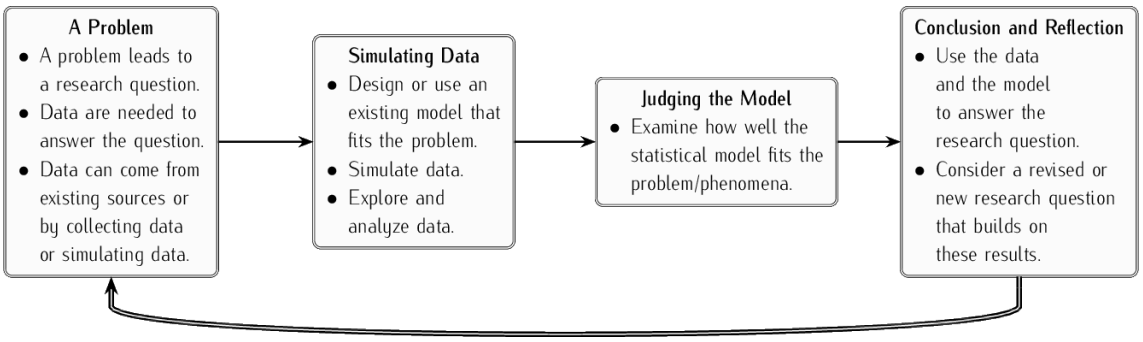


Abbildung 162: Ausschnitt aus dem Schema „Use of models in statistical analysis“ von Garfield und Ben-Zvi (2008a) (eigene Darstellung)

In gewisser Weise ähnelt dieses Schema dem vierschrittigen Schema von Rossman et al. (2008). Garfield und Ben-Zvi (2008a, 146) stellen ein allgemeines Schema zum Gebrauch von Modellen in der statistischen Analyse auf, in welches sie Randomisierungstests einordnen. Abbildung 162 zeigt den Strang des Schemas als Ausschnitt, der sich auf Randomisierungstests bezieht. Es lässt sich erkennen, dass hier ein vierschrittiges Schema zugrunde liegt, das jedoch noch einmal in Unterpunkten erklärt wird. Ähnlich wie bei Erickson (2006) werden auch hier in den Unterschritten genaue Anweisungen gegeben, was für den jeweiligen Schritt auszuführen oder zu bedenken ist. Hier finden sich große Parallelen zu dem Schema von Rossman et al. (2008). Beide Schemata bestehen aus vier Schritten, die im wesentlichen ähnliche Schritte beinhalten, wobei das

von Rossman et al. vor allem durch die Unterpunkte stärker auf Randomisierungstests zugeschnitten ist.

Tintle et al. (2009) haben ein komplettes Curriculum entwickelt, das über Randomisierungstests in statistische Inferenz einführt und aufbauend auf Cobbs Forderung Randomisierungstests in das Zentrum stellt. Sie entwickeln zunächst ein sechsschrittiges Schema für den allgemeinen statistischen Untersuchungsprozess, das in Abbildung 163 zu sehen ist. Dieses Schema erinnert stark an den PPDAC-Cycle von Wild und Pfannkuch (1999) und ist wie dieser zyklisch angelegt.

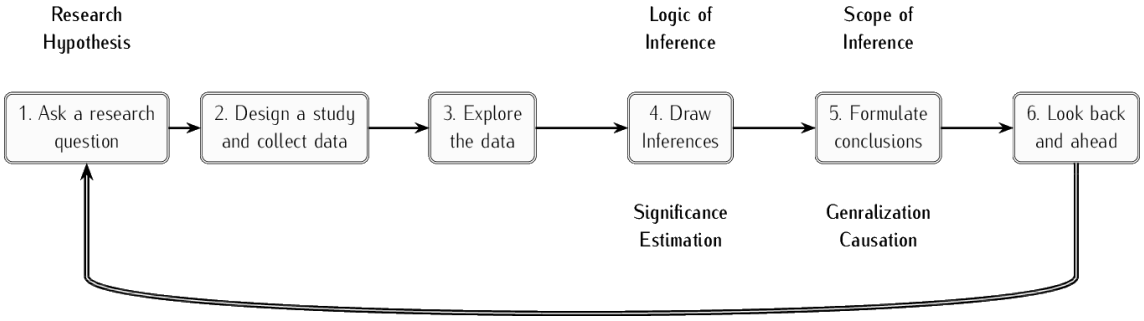


Abbildung 163: Der "Six step statistical investigation process" von Tintle et al. (2009, 2), eigene Darstellung

Im Gegensatz zu den vorangegangenen Schemata beginnen die Autoren hier mit einer Forschungsfrage, aus der sich alle weiteren Schritte ergeben. Im Sinne des PPDAC-Cycles wird also das komplette Durchlaufen eines Datenanalysezyklusses angeregt. Die bisher betrachteten Schemata setzen letztlich mit der Durchführung eines Randomisierungstests an, wenn bereits ein Experiment durchgeführt ist. Wünschenswert für eine Lernsituation wäre natürlich wie hier angeregt, mit einer Forschungsfrage zu beginnen und alle weiteren Schritte selbständig durchzuführen. In der Realität jedoch scheinen die Schemata, die zwei Schritte später beginnen, also mit dem Explorieren vorhandener Daten starten, wesentlich häufiger anzutreffen zu sein und im Unterricht mit wesentlich weniger Aufwand realisierbar zu sein.

Darüber hinaus entwickeln die Autoren noch die von ihnen sogenannte „3S Strategy for measuring strength of evidence“. Diese Strategie „will serve as the foundation for addressing the question of statistical significance” (Tintle et al. 2009, 17) und besteht aus den Schritten

> „1. **S**tatistic: Compute the statistic from the observed sample data.
> 2. **S**imulate: Identify a “by chance alone” explanation for the data. Repeatedly simulate values of the statistic that could have happened when the chance model is true.
> 3. **S**trength of evidence: Consider whether the value of the observed statistic from the research study is unlikely to occur if the chance model is true. If we decide the observed statistic is unlikely to occur by chance alone, then we can conclude that the observed data provide strong evidence against the plausibility of the chance model. If not, then we consider the chance model to

> be a plausible (believable) explanation for the observed data; in other words what we observed could plausibly have happened just by random chance." (Tintle et al. 2009)

Nicht zufällig erinnert diese 3S Strategie an die drei Rs von G. Cobb (2007), wenn auch das dritte R (reject) von Cobb aus dem Gedankengut von Neyman und Pearson stammt und sich „Strength of evidence" eher auf die Argumentation nach Fisher stützt (vgl. Kapitel 5.4). Auch diese Autoren verfolgen offensichtlich die Strategie, den Prozess auf drei einfache Schlagworte zu reduzieren. Wie allerdings bei Garfield und Ben-Zvi (2008a) ist dieses Schema nicht nur für Randomisierungstest, sondern allgemein für Hypothesentests anwendbar. Dadurch geht eventuell etwas von der Klarheit verloren, die Cobbs drei R in Bezug auf das Randomisieren mit sich bringen. Konform mit G. Cobb (2007) und mit Rossman et al. (2008) setzen Tintle et al. (2009) dieses Schema für die Logik des Schlussfolgerns bzw. hier für das Messen der Evidenz an. Somit ist explizit klar, dass das Schema der drei S für den Test gedacht ist und nicht für das allgemeine Modell der statistischen Analyse.

Wie bereits im vorherigen Abschnitt erwähnt, entwickeln Pfannkuch et al. (2011) ebenfalls ein Schema von drei Rs, das sicherlich inspiriert ist von G. Cobb (2007), legen allerdings bewusst einen anderen Fokus. Die drei Rs sind hier: „Re-randomize the assignment of the units to the groups, Repeat this re-randomization many times, Review the chance alone explanation" (Pfannkuch et al. 2011, 906). Wie auch bei Cobb beginnt dieses Schema mit einem Schritt, der direkt auf den Simulationsprozess beim Randomisierungstesten abzielt, und nicht mit dem Stellen einer statistischen Frage wie zum Beispiel bei Garfield und Ben-Zvi (2008a) oder Tintle et al. (2009). Das erste R erklären sie als die Beobachtung „to see what is typical and what is not typical under the chance explanation" (Pfannkuch et al. 2011, 906). Hier gehört also das Aufstellen der „chance explanation", die erst im dritten Schritt genannt wird, entweder zu diesem Schritt oder vor diesen Schritt, ohne dass dies explizit aufgenommen wird. Mit der gegebenen Erklärung erinnert dieser Schritt eher an das zweite R von G. Cobb (2007), an *repeat by simulation to see what's typical*. Eine häufige Wiederholung der Simulation geschieht dann sowohl bei Pfannkuch et al. als auch bei Cobb. Der letzte Schritt bei Pfannkuch *review the chance alone explanation* unterscheidet sich nun wieder von Cobbs drittem Schritt: *reject any model that puts your data in its tail.* Pfannkuch et al. geben im Grunde durch den Zusatz *chance alone explanation* das Modell vor, das geprüft werden soll. Dies konkret in einem Kontext anzuwenden, bleibt dem Lernenden überlassen. Allerdings wird hier die zu verwendende Sprache stark vorgegeben, welches so auch von ihnen intendiert ist. Cobb hingegen erwähnt allgemein „Modelle", die sprachlich nicht konkretisiert sind. Auch wenn es nicht explizit in den Schritten steht, so laufen doch beide Schemata auf das Bestimmen des (durch Simulation geschätzten) P-Werts hinaus, wobei Pfannkuch et al. den Sprachgebrauch des P-Werts bewusst vermeidet und nur von

„tail“ der Verteilung sprechen. Dies soll ihrer Meinung nach vereinfachend wirken. Dies muss nun im dritten Schritt sowohl bei G. Cobb (2007) als auch bei Pfannkuch et al. (2011) interpretiert werden. Die Anweisungen unterscheiden sich hier auch stark. Während bei G. Cobb (2007) ganz klar das Ziel ist, zu prüfen, ob ein Modell (in der Nullhypothese ausgedrückt) zurückgewiesen werden kann, wenn die beobachteten Daten stark am Rand der erzeugten Referenzverteilung liegen, so ist dies bei Pfannkuch et al. (2011) offener formuliert, indem die „chance alone explanation“ beurteilt werden soll. Das Ziel, das Nullmodell abzulehnen, wie dies implizit Cobbs drittes R vorschlägt, wird von ihnen nicht vorgegeben.

Das Schema von Pfannkuch et al. (2011) wird weiterentwickelt zu drei „Aktionen“, die beim Randomisierungstest durchgeführt werden sollen. In Pfannkuch et al. (2015, 99) werden diese drei Aktionen beschrieben: „(1) thinking about the data obtained from an experiment; (2) conducting the randomization test by modeling random behavior; and (3) making a claim about the data“. Wiederum besteht das Schema aus drei Schritten, die nun in der Formulierung stark abweichen und inhaltlich stärker an andere Schemata heranrücken. Dabei werden Ziele für die Schritte angeführt. Beim ersten Schritt ist dies das Auffinden von „Explanations for observed difference“, im zweiten Schritt geschieht „Testing observed difference againt chance alone“ und im letzten Schritt schließlich wird die Schlussfolgerung gezogen als „Argument for observed difference“ (Pfannkuch et al. 2015, 100). Generell wird hier wesentlich stärker der Fokus auf das Schlussfolgern durch einen Randomisierungstest auf eine bestimmte Ursache gelegt. Nach Zieffler et al. (2011) handelt es sich hierbei um die einzige mögliche Schlussfolgerung, wenn mit Daten aus Experimenten gearbeitet wird, denen keine Zufallsstichprobe zugrunde liegt. Die drei Aktionen scheinen klarer definiert zu sein, als die vorherigen drei Schritte. Im Gegensatz zu dem Schema von 2011 dieser Autoren zielt nun die erste Aktion des neuen Schemas eindeutig auf das Untersuchen der Daten des Experiments ab, was zuvor gefehlt hat. Auch der zweite Schritt ist nun klarer definiert. Es ist eindeutig der Schritt, in welchem der statistische Test, also der Randomisierungstest durchgeführt werden soll und zwar auf Grundlage von „random behavior“. Somit ist auch das Nullmodell bzw. die Nullhypothese in diesen Schritt implementiert. Der dritte Schritt bleibt ähnlich wertungsfrei wie vorher auch, es soll eine Beurteilung der Ergebnisse stattfinden. Die griffige Verschlagwortung oder Alliteration wird bei dieser Weiterentwicklung verlassen zugunsten von etwas ausdifferenzierteren Anweisungen für die drei Schritte.

Ein weiteres Schema für die Durchführung eines Randomisierungstests stellen Frischemeier und Biehler (2014) vor, das auf dem allgemeinen Schema zum Durchführen von Simulationen mit Software von Maxara und Biehler (2007) aufbaut. Frischemeier und Biehler (2014) verwenden ein Sechs-Schritte-Schema, das aus den Schritten *Observation – Hypothesis* H_0 *– Simulation of* H_0 *– Test statistic – P-value – Conclusions* besteht (Frischemeier & Biehler 2014, 5). Dieses Schema wird an einem

Beispiel paraphrasiert. Hier ist zu beobachten, dass das Schema im Wesentlichen aus einzelnen Worten besteht, im Gegensatz zu allen anderen Schemata werden hier keine genauen Anweisungen gegeben oder Fragen gestellt. Dafür ist dieses Schema explizit als Schema in Tabellenform für Lernende gedacht, in welchem zu jedem Schlagwort etwas notiert werden soll. Die explizite Anweisung, eine Nullhypothese zu formulieren, findet sich ansonsten nur bei Rossman et al. (2008). Dieses Schema und eine Studie dazu wird in Frischemeier (2017) genauer ausgeführt.

Es gibt ein weiteres Curriculum, das komplett auf dem Modellierungsansatz durch Simulationen aufbaut und Randomisierungstests in den Vordergrund stellt, um in inferentielles Denken einzuführen. Dies ist der Catalst-Kurs (Catalst steht für Change Agents for Teaching and Learning Statistics, Garfield et al. 2012; Zieffler & Catalysts for Change 2013). Der Grundgedanke des Modellierens wird beschrieben als „Setting up models, use models to simulate data, examine distributions of simulated data, evaluate an observed result within a distribution, and use that distribution to estimate a standard error" (Garfield et al. 2012, 886). Dies bezeichnen die Autoren als Methode, um „Kochen" zu lernen, anstatt nur „Rezepte" zu befolgen, wie sie in Standardkursen gelehrt würden. Mit Bezug auf G. Cobb (2007) wird diese Methode in einem allgemeinen Schema beschrieben, das ebenfalls wieder dreischrittig ist, wie viele andere aber auch noch weitere Unterschritte enthält. Die drei Schritte sind *Model*, *Randomize and repeat* und *Evaluate* und sind wie folgt definiert:

> „*Model*. Specify a model that will generate data to reasonably approximate the variation in outcomes attributable to the random process - be it in sampling or assignment. The model is often created as a null model that may be rejected in order to demonstrate an effect.
>
> *Randomize and repeat*. Use the model to generate simulated data for a single trial, in order to assess whether the outcomes are reasonable. Specify the summary measure to be collected from each trial. Then, use the model to generate simulated data for many trials, each time collecting the summary measure.
>
> *Evaluate*. Examine the distribution of the resulting summary measures. Use this distribution to assess particular outcomes, evaluate the model used to generate the data, compare the behavior of the model used to generate the data, compare the behavior of the model to observed data, make predictions, etc." (Garfield et al. 2012, 887)

Wie bei Budgett et al. (2012) und bei Tintle et al. (2009) findet sich auch hier die starke Orientierung an G. Cobb (2007). Wiederum startet das Schema beim Testen und lässt eventuelle Analysen von beobachteten Daten außen vor. Es wird bereits im ersten Schritt der Blick auf das mögliche (gewünschte) Verwerfen des Nullmodells gelenkt, wie dies auch bei G. Cobb (2007) angedacht ist. Hier wird das zu erstellende (Null-)Modell konkret im ersten Schritt benannt. Der zweite Schritt ähnelt dem zweiten Schritt der anderen Schemata, hier soll die Simulation häufig durchgeführt werden, wiederum versehen mit dem Hinweis zu prüfen, „whether the outcomes are reasonable" – vermutlich in Bezug auf eine anfängliche Beobachtung in den Daten. Im dritten

Schritt, dem Evaluieren, wird jedoch im Gegensatz zum ersten Schritt das Verwerfen des Nullmodells nicht mehr als explizites Ziel hervorgehoben. Vielmehr geht es wie bei Pfannkuch et al. (2011) und Pfannkuch et al. (2015) um ein wertungsfreies Untersuchen der Ergebnisse. Ein Blick in das Lehrwerk (Zieffler & Catalysts for Change 2013) zeigt, dass die in jedem Schritt formulierten Anweisungen durch genaue Fragen in einzelnen Teilaufgaben detailliert aufgegriffen werden. Dies geschieht jedoch immer an konkreten Aufgaben, so dass das oben zunächst grob erscheinende dreischrittige Schema durch viele Unterpunkte stark verfeinert und angeleitet wird. Das Lehrwerk besteht aus drei Einheiten, von denen die erste *Chance models and simulations* hauptsächlich das Wesen von zufälligen Ereignissen untersucht und zugehörige Modelle aufstellt und simuliert. Erst nachdem dadurch ein gewisser probabilistischer und simulationstechnischer Grundstein gelegt ist, werden das Hypothesentesten und Randomisierungstests in der zweiten Einheit *Models for Comparing Groups* eingeführt. Randomisierungstests stellen hier somit nicht den Anfangspunkt des Curriculums dar. Der Vollständigkeit halber sei erwähnt, dass in der dritten Einheit hauptsächlich Bootstrapping behandelt wird, im Fokus stehen Schätzverfahren.

In ähnlicher Weise wie Rossman et al. (2008) geben auch Fitch und Regan (2014) für einen introductory statistics course der University of Auckland ein Schema vor, das sie wie Frischemeier und Biehler (2014) verschriftlicht an die Lernenden geben. Sie nennen dieses Schema einen Algorithmus, der bei ihnen fünf Schritte umfasst: hypothesis →data → test statistic → P-value → interpretation. Bis auf den ersten Schritt „observation“ ist dies nahezu identisch mit dem Schema von Frischemeier und Biehler (2014), abgesehen davon, dass der zweite Schritt bei Frischemeier und Biehler „simulation“ statt „data“ heißt, allerdings bei Fitch & Regan ebenfalls das Simulieren neuer Daten gemeint ist. Als Ziel beschreiben Fitch und Regan (2014): „we want them [the learners] to show explicitly all the steps involved in order to gain a better understanding of the testing method” (Fitch & Regan 2014, 4). Über die Wirksamkeit des erstellten und an die Lernenden ausgegebenen Schemas wird leider nichts berichtet.

Das Lehrwerk von Lock et al. (2013) für introductory statistics courses verfolgt ebenfalls den Ansatz, über Simulationen konzeptionelles Verständnis der beurteilenden Statistik zu erzeugen. Darin wird nach einer Einheit zu Datenanalyse und nach einer Einheit zu Konfidenzintervallen in Hypothesentests über Randomisierungstests eingeführt ähnlich wie bei Zieffler und Catalysts for Change (2013). Ein explizites Schema findet sich hier nicht, allerdings findet sich dazu bei Lock et al. (2014) ein verbal beschriebenes Vorgehen, aus dem ein Schema abgeleitet werden kann.

> „We start with a set of (null and alternative) hypotheses, some sample data, and a statistic measuring some aspect of that sample data. The critical question of interest is how unusual would that sample statistic be if the null hypothesis were true. A straightforward way to approach this question (and thus estimate a p-value) is to generate lots of samples showing what is likely by random chance if the null

hypothesis is true, and then compute the proportion of statistics from those randomization samples that are as extreme as the statistic from the original sample." (Lock et al. 2014, 2)

Interpretiert man dies als Schritte, so stehen zunächst die Null- und Alternativhypothese am Anfang, zusammen mit den beobachteten Daten und einer betrachteten Statistik. Direkt mit den Hypothesen anzufangen, scheint im Hinblick auf alle anderen hier vorgestellten Schemata ungewöhnlich. Der zweite Schritt ist dann das Fragenstellen. Ähnlich wie bei allen anderen auch wird gefragt, wie ungewöhnlich eine bestimmte Beobachtung ist unter Voraussetzung, dass die Nullhypothese wahr ist. Im dritten Schritt wird die Referenzverteilung erzeugt mit dem Ziel, den P-Wert zu schätzen. Es fehlt am Ende noch eine Beurteilung der erhaltenen Ergebnisse.

Ein weiteres Schema explizit zu Randomisierungstests findet sich in Biehler et al. (2015). In diesem Schema wird zwischen drei „Welten" unterschieden, die jeweils ineinander eingebettet sind: Zum einen die kontextuelle Welt, die statistische Welt und die Welt der Software. In jeder Welt müssen nach diesem Schema zwei Schritte ausgeführt werden, vgl. Abbildung 164.

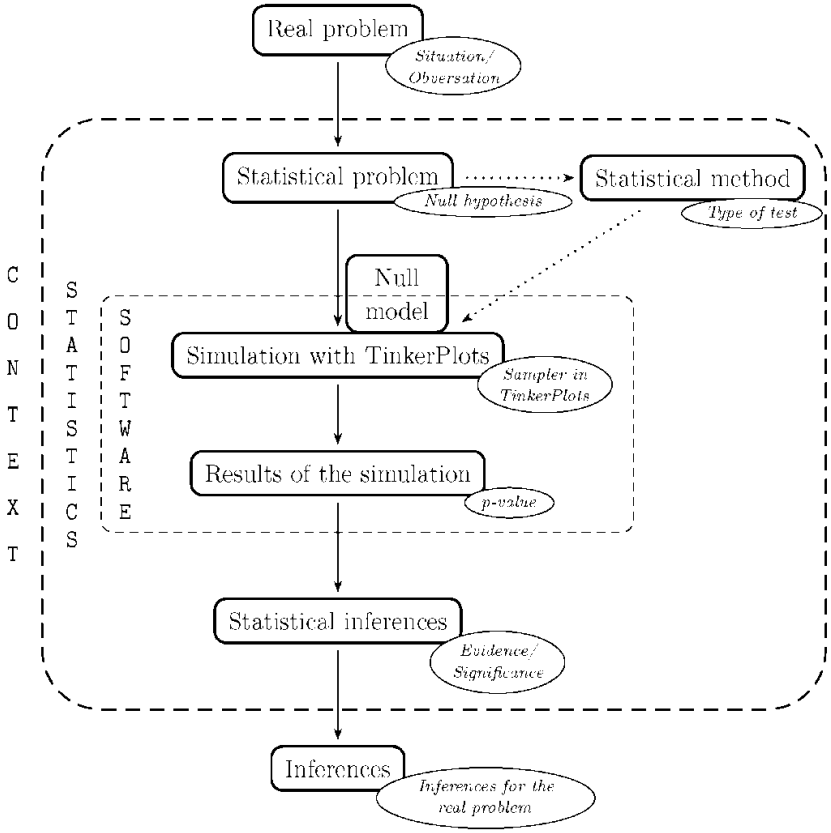


Abbildung 164: Schema zum Randomisierungstesten aus Biehler et al. 2015, 139

Die hierzu gehörigen „6 Major Steps when conducting a randomization test" sind:

„Step 1 Reading off the difference of the means of the groups in the dataset

Step 2 Formulating an adequate null hypothesis
Step 3 Describing the null model
Step 4 Formulating the test statistic
Step 5 Determining the p-value
Step 6 Drawing conclusions from the p-value" (Biehler et al. 2015, 143)

Aus diesem Schema wird ein Sechs-Schritte-Plan für die Studierenden entwickelt, der bei der Durchführung eines Randomisierungstests zur Orientierung dienen soll. Dieser besteht aus den Schritten 1. Observation, 2. Null hypothesis, 3. Simulation of the null hypothesis, 4. Test statistic, 5. P-Value und 6. Conclusion. Jeder Schritt dieses Plans ist einem Schritt des Schemas in der jeweiligen „Welt" zugeordnet. Hier lässt sich deutlich erkennen, in welchem Bereich (Kontext, Statistik oder Software) ein Schritt zu bearbeiten ist. Insofern ist dies eine nützliche Ergänzung zu bisher betrachteten Schemata. Für die Schritte, die in der Software-Welt auszuführen sind, *describing the null modell* und *determining the p-value* wird von den Autoren ein zusätzliches Schema entwickelt, das die nötigen Schritte für die Simulation (in TinkerPlots) auflistet(Biehler et al. 2015, 147).

Für die Lernenden wird aus diesen Schemata der oben erwähnte 6-Schritte-Plan entwickelt, der wiederum aus Stichworten zu den einzelnen Schritten besteht. Die beiden Schemata Step1-Step6 und TP1-TP7 werden zur Auswertung von Studentenaktivitäten verwendet.

Frischemeier (2017) entwickelt für seine Lernumgebungen ebenfalls ein Randomsierungstestschema, welches die Schritte „Beobachtung, Nullhypothese H_0, Simulation von H_0, Testgröße, Wahrscheinlichkeit: P-Wert, Beurteilung, Beurteilung dieser Situation" (Frischemeier 2017, 292) in ähnlicher Weise (da daraus entnommen) wie in Biehler et al. (2015) enthält.

6.3 Die „Kernlogik" des Randomisierungstests

Es wurden nun verschiedene Schemata zu Randomisierungstests vorgestellt. Es lässt sich beobachten, dass diese unterschiedlich ansetzen. Zum einen gibt es Kurz-Schemata, die vor allem die Logik des inferentiellen Schließens, meistens in einem Dreischritt, abbilden möchten. Als zweites gibt es Lang-Schemata, die verschiedene Konzepte auflisten, die bei einem Randomisierungstest benötigt werden. Um die Logik des Schließens bei einem Randomisierungstest zu zeigen, wurden die Kurz-Schemata entwickelt, die nun verglichen werden, um daraus einen eigenen Dreischritt zu entwickeln. Im Anschluss (Kapitel 6.4) werden Konzepte aus der Literatur zusammengestellt, die während der Durchführung eines Randomisierungstests adressiert werden sollen.

Zur Gruppe der Kurz-Schemata lassen sich vor allem Schemata identifizieren, die die „core logic of inference" (G. Cobb 2007, 4) in Kurzform darstellen. Dabei wollen diese

Schemata keinen kompletten Datenanalysezyklus wie zum Beispiel von Wild und Pfannkuch (1999) oder von Garfield und Ben-Zvi (2008a) oder auch Tintle et al. (2009) abbilden, sondern fokussieren auf die Argumentationslinie, die ihrer Meinung nach für inferentielle Schlussfolgerungen nötig ist. Hierzu gehört zunächst einmal das Schema von G. Cobb (2007), das sicherlich als Vorbild für die folgenden, ähnlichen Schemata diente. Des weiteren zählen zu dieser Gruppe die Schemata von Tintle et al. (2009), Zieffler und Catalysts for Change (2013), Pfannkuch et al. (2011) und die revidierte Fassung durch Pfannkuch et al. (2015). Alle diese Schemata bestehen aus drei Schritten. Zum Teil sind diese jedoch in die oben erwähnten kompletten Datenanalysezyklen eingebettet.

In einer Gegenüberstellung der einzelnen Schritte lassen sich hier einige Gemeinsamkeiten, aber auch Unterschiede erkennen. Die Schritte sind in Tabelle 15 zusammengestellt.

Tabelle 15: Zusammenschau der Kurz-Schemata zur Verdeutlichung der Logik beim inferentiellen Schlussfolgern

G. Cobb (2007)	Tintle et al. (2009)	Zieffler und Catalysts for Change (2013)	Pfannkuch et al. (2011)	Pfannkuch et al. (2015)
Randomize data production	Statistic: Compute the statistic from the observed sample data	*Model.* Specifiy a model that will generate data to reasonably approximate the variation in outcomes attributable to the random process - be it in sampling or assignment.	Re-randomize the assignment of the units to the groups	thinking about the data obtained from an experiment
Repeat by simulation to see what's typical	Simulate: Identify a "by chance alone" explanation for the data. Repeatedly simulate values of the statistic that could have happened when the chance model is true	*Randomize and repeat.* Use the model to generate simulated data for a single trial, in order to assess whether the outcomes are reasonable. Specify the summary measure to be collected from each trial. Then, use the model to generate simulated data for *many trials*, each time collecting the summary measure.	Repeat this re-randomization many times	conducting the randomization test by modeling random behavior

Reject any model that puts your data in its tail	Strength of evidence: Consider whether the value of the observed statistic from the research study is unlikely to occur if the chance model is true (…)	*Evaluate.* Examine the distribution of the resulting summary measures. Use this distribution to assess particular outcomes, evaluate the model used to generate the data, compare the behavior of the model used to generate the data, compare the behavior of the model to observed data, make predictions, etc	Review the chance alone explanation	making a claim about the data
Schlagworte: Randomize, repeat, reject	Schlagworte: Statistic, Simulate, Strength of evidence	Schlagworte: Model, Randomize and repeat, Evaluate	Schlagworte: keine	Schlagworte: keine

Während G. Cobb (2007) sein Schema verbal möglichst schlank hält, so wird dies bei Tintle et al. (2009) schon stärker ausformuliert und Zieffler und Catalysts for Change (2013) schließlich fügen in ihr Schema langanmutende Definitionen ein. Dafür verwenden diese drei Autorengruppen jedoch eine Verschlagwortung, die die längeren Anweisungen wieder reduziert und somit leichter merkbar macht. Eine solche Verschlagwortung findet sich weder bei Pfannkuch et al. (2011) noch bei Pfannkuch et al. (2015). Hier allerdings werden die Schritte etwas näher ausformuliert, in der Variante von 2015 noch etwas griffiger als in 2011. Das heißt, dass hier die in den Schritten gegebenen Formulierungen eher Anweisungen sind, die sich konkret umsetzen lassen.

Der erste Schritt ist in den obigen fünf Schemata durchaus unterschiedlich. Alle Schemata setzen voraus, dass ein Experiment mit randomisierter Gruppenzuordnung bereits durchgeführt wurde und Daten vorliegen.

Die Struktur der Argumentation von G. Cobb (2007) setzt bei den beobachteten Daten an. Für diese soll geprüft werden, ob tatsächlich eine Randomisierung vorliegt und somit sollen mögliche (kausale) Schlussfolgerungen in Betracht gezogen werden. Da im zweiten Schritt die Wiederholung des Randomisierens genannt wird, muss das zugehörige Modell des zufälligen Zuordnens wohl direkt im ersten Schritt aufgestellt werden. Im dritten Schritt schließlich soll dieses aufgestellte Modell beurteilt werden mit der konkreten Anweisung, es zu verwerfen, wenn die Daten zu weit am Rand („tail“) der Verteilung liegen. Wie bereits vorher erwähnt, stehen hier viele Konzepte implizit im Schema, die verstanden sein müssen, um die Schritte wirklich ausführen zu können.

Beispielsweise das Aufstellen eines Modells in Form einer Nullhypothese oder eines Nullmodells, das den Zufall als kausale Ursache für die in den Daten beobachteten Unterschiede zugrunde legt, ist ein solches nicht erwähntes Konzept. Auch das Bestimmen einer Teststatistik und daraus resultierend die Referenzverteilung sind Hintergründe, die Lernende parat haben müssen. Prinzipiell ist aber gerade durch die Verschlagwortung *randomize* (wobei vielleicht bewusst offen bleibt, ob geprüft werden soll, ob randomisierte Daten vorliegen oder ob ein Modell zur Randomisierung aufgestellt werden soll), *repeat, reject* eine Triade gefunden worden, die sich leicht merken lässt.

Anders sieht es aus bei dem Schema von Tintle et al. (2009), das ebenfalls eine Verschlagwortung beinhaltet, das jedoch anders ausgerichtet ist: *statistic, simulate, strength of evidence.* Während das Schema von G. Cobb (2007) eher die Logik in den Vordergrund stellt (erst muss randomisiert werden, dies soll wiederholt werden und dann das nicht passende Modell abgelehnt werden), so steht bei Tintle et al. (2009) eher ein statistisches Verfahren im Vordergrund. Zunächst soll eine Teststatistik in den beobachteten Daten berechnet werden. Hier beginnt die Argumentation im Gegensatz zu G. Cobb (2007) explizit mit den Daten. Es folgen die Regeln, nach denen eine „chance alone"-Erklärung im zweiten Schritt aufgestellt werden soll und als Modell häufig simuliert werden soll. Die Verschlagwortung hier mit *simulate* legt den Fokus auf das Nutzen eines bestimmten Vorgehens, nämlich der Simulation. Treffender wäre hier (auch im Licht der anderen Schemata) ggf. das Schlagwort *model* gewesen – was allerdings nicht in die Alliteration gepasst hätte. Auch für den letzten Schritt *strength of evidence* gilt, dass meiner Meinung nach hier erneut das Hauptaugenmerk auf der statistischen Ermittlung des P-Werts liegt. Es wird zwar in der Definition ein Hinweis geliefert, wie nun mit dem zuvor aufgestellten Modell umzugehen ist, allerdings kommt dies in der Verschlagwortung zu wenig zum Ausdruck, so dass Lernende, die sich diese Stichworte merken, ggf. bei dem ermittelten P-Wert stehen bleiben. Es fehlt also hier der Schluss in der Argumentation.

Zieffler und Catalysts for Change (2013) wählen nun ein etwas anders startendes Schema. Mit dem Schritt *model* setzen sie einen Schritt später an als G. Cobb (2007) und Tintle et al. (2009). In der Argumentation fehlt hier der Beginn, das Formulieren des Ausgangspunkt oder der Beobachtung. Einen beobachteten Unterschied in den Daten setzen sie offensichtlich als gegeben voraus und starten direkt mit dem Aufstellen eines Modells, das eine mögliche Ursache für die Variation in den Daten darstellen kann. Im Gegensatz zu G. Cobb (2007) und Tintle et al. (2009) wird der Schritt des Modellierens hier explizit aufgeführt, bei den anderen beiden gehört er als nicht explizit benannter Teil zu einem anderen Schritt. Der zweite Schritt bei Zieffler und Catalysts for Change (2013) ist nun wieder eher technisch, das Wiederholen der Simulation mit dem Identifizieren einer Teststatistik und einer Referenzverteilung. Der letzte Schritt *evaluate* ist nun hier wieder sehr weitgefasst und umfasst das Beurteilen des Modells und das Tref-

fen von Vorhersagen. Dies bringt sowohl die Wortwahl als auch die Beschreibung zum Ausdruck.

Pfannkuch et al. (2011) starten in ihrem Schema ebenfalls direkt mit dem Modell, das allerdings auch hier nicht so bezeichnet wird. Das Formulieren einer Beobachtung als Start der Argumentation fehlt auch hier. Das Modell wird in Schritt drei explizit durch die „chance alone explanation“ zum Ausdruck gebracht, in den Schritten vorher wird dieses Verständnis implizit erwartet. Die drei Schritte dieses Schemas beinhalten kurze Anweisungen, ähnlich wie bei G. Cobb (2007). Im Schema von 2015 lässt sich ein deutlicher Wandel erkennen. Nun wird auch hier im ersten Schritt der Blick auf die beobachteten Daten gelenkt und im zweiten Schritt wird deutlich geschrieben, dass es um das Durchführen eines Randomisierungstests geht. Wie auch im früheren Schema wird der dritte Schritt neutral formuliert. Auch hier wird deutlich, dass es zu einer Beurteilung kommen muss, dass also eine Schlussfolgerung das Ziel ist. Diese Variante erscheint wesentlich verbessert gegenüber der vorherigen. Eine Verschlagwortung nehmen die Autoren in keiner Variante vor.

Tintle et al. (2009) und Pfannkuch et al. (2015) setzen in ihren Schemata bei den beobachteten Daten an. Bei Zieffler und Catalysts for Change (2013) und Pfannkuch et al. (2011) hingegen startet das Schema mit dem Modellierungsschritt, indem ein Modell für das zufällige Zuordnen aufgestellt werden soll. Im Sinne einer Argumentation fehlt auch hier das Formulieren der Ausgangslage, der Beobachtung. Im zweiten Schritt geschieht in allen Schemata eine häufige Wiederholung des Prozesses des zufälligen (Neu-)zuordnens zu den Gruppen. Dabei ist bei G. Cobb (2007) unklar, an welcher Stelle das Modellieren der zufälligen Zuordnung passieren soll, denn im zweiten Schritt soll bereits eine häufige Wiederholung dieses Zuordnungsprozesses stattfinden. Tintle et al. (2009), Zieffler und Catalysts for Change (2013) und Pfannkuch et al. (2015) setzen die Modellierung verbal in den zweiten Schritt. Wie bereits eingangs erwähnt, handelt es sich hier um Kurz-Schemata, die die „Regeln“ nicht explizit als einzelne Schritte ausformulieren, sondern die gesamte Argumentationslinie im Blick haben, oder zumindest den Anschein erwecken, diesen Anspruch zu haben.

Aus diesen Darstellungen lässt sich ein Schema synthetisieren, das die Logik des Randomisierungstests abbildet. Unter Dreischritt wird hier als erstem Schritt das Formulieren des Ausgangspunkts, also der Beobachtung verstanden, das über Modellieren und Simulierem als zweitem Schritt argumentativ zu einer Schlussfolgerung im dritten Schritt führen soll. Das synthetisierte Schema soll sowohl das Untersuchen der beobachteten Daten als auch das Modellieren beinhalten, so wie es in den obigen Schemata nur teilweise explizit beinhaltet ist. Dabei scheint eine verkürzende Darstellung im Sinne von Schlagworten wie bei den ersten drei Schemata sinnvoll zu sein, auf eine Alliteration kann gut verzichtet werden. Im Sinne der *cognitive load theory* (Chandler & Sweller

1991, vgl. Kapitel 2.3) ist eine Verschlagwortung der Argumentation leichter zu merken, dadurch, dass der *extraneous load* verringert wird. Die Verschlagwortung soll dazu dienen, die zugehörigen Konzepte wieder abzurufen.

Als Elemente der Logik beim Randomisierungstest lassen sich ableiten:

Nachdenken über das Experiment und die Daten: In diesem Element soll die Ausgangslage der Argumentation formuliert werden. Dabei gilt es entweder, ein entsprechendes Experiment zu planen, durchzuführen und Unterschiede zu notieren. Oder, falls mit den Daten eines bereits durchgeführten Experiments gearbeitet wird, es soll im Sinn von G. Cobb (2007) und Pfannkuch et al. (2015) über das durchgeführte Experiment nachgedacht werden. Dazu soll betrachtet werden, inwiefern dem Experiment eine Randomisierung der Teilnehmer oder der Objekte auf Gruppen stattgefunden hat, da dies eine direkte Auswirkung auf die Schlussfolgerungen hat (vgl. Kapitel 5.4.1. Außerdem soll zu diesem Element gehören, dass wie bei Tintle et al. (2009) die beobachteten Daten analysiert werden und Unterschiede zwischen den Gruppen herausgearbeitet werden. Nach Biehler et al. (2015) ist dies ein Schritt, der vor allem auf Kontextebene anzusiedeln ist.

Randomisieren und Wiederholen: Dieses Element umfasst mehrere Schritte, die ausgeführt werden müssen. Es geht darum, die Regeln aufzustellen, nach denen im Anschluss Schlussfolgerungen gezogen werden sollen. Zu diesem Element gehört als erstes das Aufstellen eines Nullmodells, so wie es implizit oder explizit bei allen fünf obigen Schemata vorkommt. Es soll jedoch im

Gegensatz zu G. Cobb (2007) deutlich werden, dass dieses Modellerstellen in diesem Schritt geschehen soll. Als Wortwahl soll hier nicht wie bei Tintle et al. (2009) Simulieren verwendet werden, da dies eher ein Vorgehen bezeichnet und durch das Wort Randomisieren der Fokus auf den Randomisierungstest gelegt wird. Wie von Biehler et al. (2015) aufgezeigt befindet sich dieses Element auf Statistik- und Softwareebene.

Schlussfolgerungen ziehen: Auch dieses Element findet sich bei allen Schemata. Es soll jedoch Wert darauf gelegt werden, dass Schlussfolgerungen wie bei Biehler et al. (2015) sowohl statistisch als auch im Kontext zu formulieren sind.

Dabei kann das Schema *Nachdenken über die Daten – Randomisieren und Wiederholen – Schlussfolgerungen ziehen* als solches an Lernende gegeben werden. Es werden, wie bei G. Cobb (2007) und im Gegensatz zu Tintle et al. (2009), Schlagworte gewählt, die Verben, also Aufforderungen etwas zu tun, beinhalten.

6.4 Konzepte für die Durchführung eines Randomisierungstests

Bisher (Kapitel 6.3) wurde beschrieben, wie eine Argumentation für die Durchführung eines Randomisierungstests in der dreielementigen Struktur aussehen kann, um die Logik des Schlussfolgerns zu verdeutlichen. Dabei wurden jedoch die nötigen Konzepte, die für die drei Schritte nötig sind, noch nicht im Detail ausgeführt. Der Blick wie mit einer Lupe auf diesen Teil der Argumentation soll nun in diesem Abschnitt stattfinden. Dabei müssen verschiedene Ebenen unterschieden werden. Zum einen muss auf statistischer Ebene argumentiert werden, zum anderen müssen statistische Annahmen in Software umgesetzt werden und mit der Software gearbeitet werden. Weiterhin müssen Ergebnisse aus Software interpretiert werden. Zusätzlich sollen die Konzepte im jeweiligen Kontext formuliert werden, um den Bezug zum zugehörigen Experiment, respektive der zugehörigen Aufgabe, herzustellen. Eine entsprechende Einstufung findet sich bei Biehler et al. (2015).

In den verschiedenen vorgestellten Schemata und Studien finden sich einige Konzepte für das Durchführen eines Randomisierungstests, die teilweise als normative Schritte, teilweise als Auswertungskategorien verwendet werden. Von diesen nehmen jedoch nur Rossman et al. (2008) und Tintle et al. (2009) einen expliziten Bezug auf ein Schema zur Logik des Randomisierungstests. Die hier vorgestellten Schemata sind eher prozessorientiert in dem Sinne, welche Abfolge von Schritten beim Durchführen eines Randomisierungstests beachtet werden müssen bzw. welche Konzepte jeweils aufge-

nommen sind. Welche Konzepte lassen sich nun finden? Hierzu erfolgt die Übersicht in tabellarischer Form, wobei das Schema von Erickson (2006) der Übersicht halber zu Stichworten zusammengefasst wird, um den Vergleich mit den anderen Schemata zu ermöglichen. Es bleibt anzumerken, dass dieses Schema und das von Rossman et al. (2008) als einzige eine Ausformulierung der einzelnen Schritte vornimmt. Das Schema von Rossman et al. (2008) ist reduziert dargestellt auf die dort formulierten Schlagworte.

In Tabelle 16 wird versucht, die verschiedenen vorgestellten Schemata mit ihren einzelnen Konzepten anzuordnen, wobei gleiche Konzepte auf gleicher Höhe stehen sollen, so dass eventuell einige Zellen leer bleiben, weil das entsprechende Schema das jeweilige Konzept nicht (explizit) beinhaltet. Aus dieser Zusammenschau wird im Anschluss ein Schema entwickelt, das die Konzepte zusammenführt, die für die Durchführung eines Randomisierungstests nötig sind.

Tabelle 16: Zusammenschau der prozessorientierten Konzepte innerhalb einzelner Schemata zum Randomisierungstesten

Erickson (2006)	Rossman et al. (2008)	Fitch und Regan (2014)	Biehler et al. (2015)	Frischemeier (2017)
Collect data				
Design a measure	Observed data		Situation/ Observation	Beobachtung
	Data production			
	Research question			
	Null hypothesis	Hypothesis	Null hypothesis	Nullhypothese
	Conditioning			
	Null model		Null model	
Simulate the null hypothesis		Data		Simulation der Nullhypothese
Compute the measure & build a sampling distribution	Test statistic	Test statistic		Testgröße
	Observed value			
P-value	P-value	P-value	P-value	P-Wert
	Reject?		Evidence/ Significance	Beurteilung
	Conclusion	Interpretation	Inferences	Beurteilung in der Situation

Von all diesen Schemata ist das von Rossman et al. (2008) jenes mit den meisten Schritten. Schaut man genauer hin, so lässt sich jedoch erkennen, dass sich die Schemata sehr stark ähneln. Zu Beginn steht als Konzept das Untersuchen der beobachteten Daten, wobei alle Schemata hier Spielraum zulassen und ggf. die Frage nach der Herkunft der Daten, also ob randomisierte Gruppen zugrunde liegen, mit beinhalten können. Lediglich das Schema von Rossman et al. (2008) fragt explizit nach diesem Aspekt. Abweichend ist noch, dass das Schema von Erickson (2006) einen Schritt vor allen anderen

beginnt, mit dem Datensammeln. Wie jedoch bereits früher erwähnt, ist dies ein interessanter Ansatz, der sich in vielen Unterrichtssituationen jedoch aus Zeitgründen kaum realisieren lässt. Dieser Schritt lässt sich auch interpretieren als das Auffinden der Daten eines interessanten Experiments. Das nächste Konzept, das in allen Schemata vorkommt, ist das Aufstellen einer Nullhypothese, außer bei Erickson (2006), hier wird diese erst im nächsten Schritt erwähnt. Das darauf folgende Konzept ist das Erstellen eines Nullmodells, welches in zwei Schemata explizit aufgenommen ist. Dieses Nullmodell wird standardmäßig durch Software ausgedrückt und findet sich dann in der Simulation, die im nächsten Schritt genannt wird. In keinem Schema findet sich sowohl das Konzept zum Nullmodell als auch der Schritt für die Simulation, was wohl an dem vorgenannten liegen mag. Das Bestimmen der Testgröße findet sich ebenfalls in allen Schemata bis auf das von Biehler et al. (2015). Auch der P-Wert ist ein Konzept, das in allen Schemata vorkommt. Teilweise enden die Schemata an dieser Stelle, einige gehen einen Schritt weiter und fordern eine statistische Beurteilung, womit hier das Bestimmen der Evidenz oder das statistische Ablehnen des Modells gemeint ist. Weder Erickson (2006) noch Fitch und Regan (2014) haben dies in ihr Schema aufgenommen. Erickson (2006) verzichtet auch als einziger in seinem Schema auf das Interpretieren der gefundenen Ergebnisse, alle anderen fordern dies in ihrem letzten Schritt.

Über die einzelnen Konzepte in diesen prozessorientierten Schemata scheint weitgehend Einigkeit zu herrschen. Auch in der Wahl der Schlagworte ist nicht viel Variation zu finden. Hieraus leite ich in Verbindung mit dem Dreischritt zur Logik beim Randomisierungstesten und der aufgearbeiteten didaktischen Literatur des ersten Teils dieses Kapitels neun Konzepte für einen Randomisierungstest ab. Zusätzlich zu den in Tabelle 16 genannten Konzepten führe ich zu Beginn noch das Konzept der Forschungsfrage ein, um in den jeweiligen Kontext einzuführen. Dieses Konzept gehört der Vollständigkeit halber dazu, gerade wenn ein Experiment nicht selbst geplant und durchgeführt wird, sondern mit bereits vorliegenden Daten gearbeitet wird, die Notwendigkeit für dieses Konzept haben die Analysen von Noll und Kirin (2017) gezeigt. Die beobachteten Daten im Hinblick auf einen beobachteten Unterschied und im Hinblick auf das Design des zugrunde liegenden Experiments zu analysieren, ist dann das nächste Konzept. Die zufällige Zuordnung der Versuchseinheiten zu den Gruppen eines Experiments zu erklären, ist das nächste Konzept. Als Konzept gehört weiterhin dazu, mögliche Erklärungen für beobachtete Unterschiede zu formulieren, da dies die Notwendigkeit eines Tests ausmacht (vgl. Pfannkuch et al. 2015). Dies wird gefolgt vom Konzept des Aufstellens eines Nullmodells, worin das Ausformulieren der Nullhypothese und der Alternativhypothese integriert wird. Die Simulation wird als nächstes Konzept aufgenommen, zum einen um klar zu machen, dass nun der Schritt in die Software erfolgt, zum anderen, um eine Beschreibung und somit Erklärung des zufälligen Zuordnens zu fordern. Das nächste Konzept beinhaltet das Bestimmen der Teststatistik und wie bei

Erickson (2006) ausformuliert, das Erstellen der Referenzverteilung. Aus dieser soll dann als nächstes Konzept der P-Wert bestimmt werden. Im letzten Konzept sollen dann Schlussfolgerungen gezogen werden. Die nötigen Konzepte sind in Tabelle 17 zu sehen und erklärt. Es erfolgt eine Einordnung der Konzepte auf die Ebenen Kontext, Statistik und Software, in Anlehnung an Biehler et al. (2015) in der dritten Spalte von Tabelle 17.

Tabelle 17: Abgeleitete Konzepte zur Durchführung eines Randomisierungstests unter Einbezug der zu adressierenden Ebene nach Biehler et al. (2015)

Konzept	Beschreibung	Ebene
Forschungsfrage aufstellen/rekonstruieren	Die Frage, die zur Durchführung des Experiments geführt hat, wird hier genannt	Kontext
Zufällige Zuordnung im Experiment erklären	Das Design des Experiments soll analysiert werden im Hinblick darauf, ob tatsächlich eine zufällige Zuordnung von Versuchseinheiten auf Gruppen geschehen ist. Weiter soll erläutert werden, welche Bedeutung dies im Hinblick auf den Randomisierungstest hat	Kontext*
Beobachtete Daten analysieren	Analyse der beobachteten Daten des Experiments, z. B. kann bei zwei numerischen Merkmalen ein Gruppenvergleich anhand von Mittelwerten oder ein Vergleich bestimmter Anteile stattfinden und als X_{beob} notiert werden. Ein Unterschied zwischen den Gruppen des Experiments muss ausformuliert werden	Kontext
Mögliche Erklärungen für beobachtete Unterschiede nennen	Für die beobachteten Unterschiede sollen zwei mögliche Erklärungen gefunden werden, da hieraus die Motivation für einen Randomisierungstest erfolgt. Eine mögliche Erklärung ist die zufällige Zuordnung der Versuchseinheiten auf die Gruppen, die andere mögliche Erklärung ist die Wirksamkeit einer durchgeführten Maßnahme (die dritte mögliche Erklärung, die auf konfundierenden Variablen beruht, sollte durch das Design des Experiments ausgeschlossen werden)	Kontext*
Nullmodell aufstellen	Null- und Alternativhypothese werden hier formuliert. In der Nullhypothese wird die zufällige Zuordnung zu den Gruppen als Erklärung für die beobachteten Unterschiede ausgedrückt. Die Nullhypothese wird im Folgenden als wahr vorausgesetzt und im Folgenden modelliert.	Kontext/ Statistik
Simulationsmodell aufstellen	Basierend auf der Nullhypothese wird erklärt, wie die Randomisierung der Daten für den Randomisierungstest aussieht. Dies wird in Software übertragen, wodurch das Nullmodell ausgedrückt und getestet wird	Software
Teststatistik und Referenzverteilung darstellen	Der zu überprüfende Wert/die Teststatistik X wird für die simulierten Daten festgelegt, z. B. als *X: Differenz der arithmetischen Mittelwerte*. Eine häufige Wiederholung der Simulation findet statt und aus der Sammlung der Teststatistik wird die Referenzverteilung erzeugt.	Software

P-Wert ermitteln	Der P-Wert als Wahrscheinlichkeit dafür, einen Wert wie den beobachteten oder einen noch extremeren zu erhalten, unter der Voraussetzung, dass die Nullhypothese wahr ist, wird als $P(X \geq X_{beob})$ aus der Simulation geschätzt	Software
Schlussfolgerung ziehen	Es erfolgt das Ziehen von Schlussfolgerungen. Hierbei wird geschaut, ob ein signifikantes Ergebnis vorliegt ($P(X \geq X_{beob}) \leq 10\%$), und die Nullhypothese abgelehnt werden kann. Restunsicherheiten werden diskutiert. Weiter wird geprüft, ob in den beobachteten Daten eine Randomisierung der Teilnehmer zu den Gruppen stattfand und deshalb auf eine kausale Erklärung geschlossen werden darf. Gleichfalls wird diskutiert, ob sich die Ergebnisse verallgemeinern lassen, was nur möglich ist, falls mit einer Zufallsstichprobe gearbeitet wurde	Statistik/ Kontext

* Diese Konzepte sind nicht im Schema von Biehler et al. (2015) enthalten.

Wie bei Erickson (2006) und Rossman et al. (2008) soll aus diesen Konzepten ein Schema für Lernende entwickelt werden, das um kurze Anweisungen bzw. Fragen zur Orientierung angereichert werden soll. Das fertige Schema soll während der Lerneinheit zum inferentiellen Schließen mit Randomisierungstests (Kapitel 8) verwendet werden.

Das Schema von Biehler et al. (2015), welches in Abbildung 164 zu sehen ist, zeigt die Übergänge vom Kontext in die Welt der Statistik und darin eingebettet in die Welt der Software. Noll und Kirin (2017) haben dieses disjunkt dargestellte Schema erweitert, in dem sie darauf hingewiesen haben, dass auch in den der Softwarewelt zugeordneten Schritten statistisch argumentiert wird und ein häufiger Wechsel zwischen der Ebene der Software und der Ebene der Statistik bei Lernenden zu beobachten ist. Entsprechende Wechsel zwischen den Ebenen sollen im Rahmen dieser Arbeit mit analysiert werden im Hinblick auf den Argumentationsprozess.

Der P-Wert im Randomisierungstest

Eine Besonderheit stellt das Konzept des P-Werts dar, zu dem es viel publizierte Literatur gibt. Gerade in der anglo-amerikanischen Tradition findet der P-Wert häufig Anwendung. In Deutschland gibt es wenige Ansätze hierzu, ein Beispiel hierfür ist das Schulbuch „Neue Wege“ (Lergenmüller et al. 2012). In der hier vorliegenden Arbeit soll das Konzept des P-Werts Einzug finden, da sich dieser beim Randomisierungstest anbietet, wie es auch Lock et al. (2014, 3) formulieren:

> “The process to find a p-value in a randomization distribution is also very visually compelling and directly reinforces the definition of a p-value. (…) The process for finding the p-value is more intuitive and concrete than looking up a standardized value in a table for some theoretical approximating distribution.”

Im Folgenden werden einige Schwierigkeiten berichtet, die Lernende mit dem Konzept des P-Werts zeigen, da diese für die Studie dieser Arbeit relevant sind.

Bei Garfield und Ben-Zvi (2008a) findet sich eine Zusammenstellung der typischen Schwierigkeiten, die Lernende beim Interpretieren von P-Werten haben. Diese sind

- „A P-value is the probability that the null hypothesis is true.
- A *P*-Value is the probability that the null hypothesis is false.
- A small *P*-Value means the results have significance (statistical and practical significance are not distinguished).
- A *P*-Value indicates the size of an effect (e.g., strong evidence means big effect).
- A large *P*-value means the null hypothesis is true, or provides evidence to support the null hypothesis.
- If the *P*-value is small enough, the null hypothesis must be false.“ (Garfield & Ben-Zvi 2008a, 270)

Auch Budgett et al. (2013, 2) stellen eine solche Liste auf:

- „regarding a p-value as the probability that the research results are due to chance;
- considering a p-value as the probability that the null hypothesis is true (given the data) rather than the probability of the data (assuming that the null hypothesis is true);
- believing that the size of a p-value is an indicator of the size of any difference or relationship;
- concluding that ruling out a null hypothesis at a particular level of significance, say α, means that the research hypothesis has a probability of 1-α of being true;
- interpreting a statistically significant result as practically important;
- accepting the null hypothesis if the p-value is considered to be ‘large’.”

Eine ähnliche Aufstellung findet sich auch bei Krauss & Wassner 2001.

Bei all den genannten Schwierigkeiten zum Auffinden und Interpretieren des P-Werts stellt dieser trotzdem eine gute Möglichkeit dar, das Maß an Evidenz gegen eine Nullhypothese zu bestimmen. Für die Interpretation des P-Werts geben Tintle et al. (2009) eine Richtlinie vor, an der sich Lernende orientieren können.

„Guidelines for evaluating strength of evidence from p-values

0.10 < p-value	not much evidence against null hypothesis; null is plausible
0.05 < p-value < 0.10	moderate evidence against the null hypothesis
0.01 < p-value < 0.05	strong evidence against the null hypothesis
p-value < 0.01	very strong evidence against the null hypothesis

The smaller the p-value, the stronger the evidence against the null hypothesis.” (Tintle et al. 2009, 26),

Wobei in dieser Richtlinie die Formulierung „null is plausible“ irreführend sein kann, da hier interpretiert werden könnte, dass die Nullhypothese angenommen werden soll. Eine solche Formulierung soll in der vorliegenden Arbeit nicht unterstützt werden.

Eine ähnliche Kategorisierung findet sich auch in anderen Lehrwerken, beispielsweise bei Meyfarth (2009), und auch Frischemeier (2017) verwendet eine solche Aufstellung. Demgegenüber stehen jedoch recht viele der ausgeführten Schemata (Kapitel 6.4), die eine solch differenzierte Richtlinie nicht verwenden. Bezüglich der Vermischung der

Ansätze zum Hypothesentesten von Fisher (nach dem es nur eine Nullhypothese gibt und durch einen Hypothesentest das Maß an Evidenz gegen diese in Form eines P-Werts ermittelt wird) und von Neyman (nach welchem eine Entscheidung zwischen Null- und Alternativhypothese aufgrund des Erreichens eines vorher festgelegten Signifikanzniveaus stattfinden soll) argumentiert Carlton (2003), dass Lernende „can handle both approaches“. Ein solcher Ansatz findet sich beispielsweise bei Pfannkuch et al. (2015), die eine klare Entscheidungsgrenze vorgeben, ab welcher ein P-Wert zum Ablehnen der Nullhypothese führen soll. Eine solche Entscheidungsgrenze soll zur Vereinfachung ebenfalls Einzug halten in das in der vorliegenden Arbeit zu verwendende Konzept.

6.5 Verwendete Software bei Randomisierungstests

Da in dieser Arbeit ausschließlich das Durchführen von Randomisierungstests über Simulationen betrachtet wird, wird nun im letzten Unterkapitel geschaut, welche Software in der Literatur für die Durchführung von Randomisierungstests beschrieben wird und welche Anforderungen dort genannt werden.

Holcomb, Chance, Rossman, Tietjen, et al. (2010) verwenden für ihre Lerneinheit Applets und werfen die Forderung auf, dass diese möglichst den händischen Simulationsprozess wiederspiegeln sollen, „so the technology is not simply a black box“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, S. 4). Außerdem stellen sie die Frage, ob eine Software einen P-Wert automatisch ausgeben soll, oder ob dieser von den Lernenden selbständig identifiziert werden soll. Hierauf liefern sie jedoch keine Antwort. Als generelle Zusammenfassung am Ende berichten die Autoren, dass „the design of the simulations appear effective, but students still struggle using the simulation results to draw appropriate conclusions“ (Holcomb, Chance, Rossman, Tietjen, et al. 2010, 6), ohne dies weiter auszuführen.

In dem Kurs von Tintle, VanderStoep, Holmes, Quisenberry und Swanson (2011) wird in der ersten Hälfte des Kurses Fathom verwendet. In diesem Teil wird über Randomisierungstests in die Inferenzstatistik eingeführt. Im zweiten Teil stehen traditionelle Tests im Vordergrund, wie der t-Test, $\chi 2$-Test, etc. Für diesen zweiten Teil wird SPSS verwendet, wobei als einzige Evaluation dazu genannt wird, dass „most students are comfortable with using two software packages in the course“ (Tintle et al. 2011, 6).

Bei Budgett et al. (2012, 3) findet sich der Anspruch, dass

> „dynamic software can allow students to analyze directly the behavior of a phenomenon, to visualize statistical processes in way that were not previously possible such as viewing a process as it develops rather than analyzing it from the end result. (…) The computer is able to take on the lower level tasks, such as performing many calculations, whilst the student can attend to the higher level tasks of applying the logic of statistical inference to the problem at hand.“

Diese Autoren verwenden dazu die von Chris Wild entwickelte VIT Software, die ein spezielles Modul für Randomisierungstests implementiert hat.

Garfield et al. (2012) entscheiden sich in ihrem entwickelten Curriculum für die Verwendung der Software TinkerPlots, weil sie in ihr einige Vorteile sehen.

> (…) TinkerPlots™ software was chosen instead because of the unique visual capabilities it has, allowing students to see the devices they select (e.g., sampler, spinner) and to easily use these models to simulate and collect data, which allows students to examine and evaluate distributions of statistics in order to draw statistical inferences. Although the software was developed for use in elementary and secondary classes, its capabilities provide a unique and novel way for tertiary students to learn to think statistically as they consider, develop, and use models to draw inferences. (Garfield et al. 2012, 886)

Die Auswertung verschiedener Assessment-Instrumente hinsichtlich der Verwendung von TinkerPlots fassen die Autoren zusammen als „Positive results were found regarding students' attitudes about the use of TinkerPlots" und weiter heißt es „It was also valuable to see that students did not dislike the use of the software and felt it helped them learn statistics" (Garfield et al. 2012, 896). Insgesamt sind die Ergebnisse und Schlussfolgerungen der Autoren sehr positiv hinsichtlich des Gebrauchs der Software TinkerPlots auch für Randomisierungstests.

Rossman und Chance (2014) haben ebenfalls ein Curriculum für den Introductory course an Universitäten entwickelt, in welchem Inferenzstatistik an den Anfang gestellt wird und Randomisierungstests im Fokus sind. Hier finden sich einige Anforderungen an eine geeignete Software für einen solchen Kurs:

- „The technology should be very easy to use. (…)
- The technology should make use of animation and mimic by-hand-simulations that students typically conduct first, using coins or cards. By making the technology-based simulation directly comparable to the by-hand-simulation, students can better understand what the technology is producing and not succumb to the temptation to view the technology as a magical black box. (…)
- The technology should have a consistent look-and-feel across scenarios, partly for ease of use and partly to emphasize common aspects of analyses in different scenarios." (Rossman & Chance 2014, 215f)

Für die spezielle Anwendung von Software zur Simulation bei Randomisierungstests oder bei Bootstraps finden sich einige Anforderungen bei Lock et al. (2014).

> „Ideally, the technology should make heavy use of interactive graphics that help illustrate the main concepts of inference. For example, it should allow students to see and distinguish between the original sample, a single bootstrap or randomization sample, and the distribution of statistics from many simulated samples. Software should strike a balance between making methodologies simple while not being so automated that it becomes simply a mysterious "black box" that cranks away in the background and spits out an answer." (Lock et al. 2014, 5)

Die wesentlichen Anforderungen hier sind demnach, dass zwischen den beobachteten Daten und den einmal randomisierten Daten und der zugehörigen Referenzverteilung

gut unterschieden werden kann. Wichtig ist auch, dass nicht alle Schritte automatisch ausgeführt werden, um der Gefahr entgegenzutreten, dass die Software in zu hohem Maß als „Blackbox“ verwendet wird. Sie nennen einige Tools, die im Web frei verfügbar sind und aus ihrer Sicht die Anforderungen gut erfüllen. Dazu gehören *StatKey* (http://lock5stat.com/statkey/), *Rossman/Chance Applets* (http://www.rossmanchance.com/applets/), *VIT: Visual Inference Tools* (https://www.stat.auckland.ac.nz/~wild/VIT/) und *Mosaic* (http://mosaic-web.org/).

Auch Roy et al. (2014) haben den traditionellen Einführungskurs in Statistik dahingehend verändert, dass sie über den Simulationszugang bereits ab der ersten Woche mit statistischer Inferenz beginnen. Hierfür haben sie eigene Webapplets[38] entwickelt und nennen einige Anforderungen an geeignete Simulationssoftware. Sie beschreiben die von ihnen entwickelten Applets als „the applets have been developed to be self-explanatory, and show students visual representations of what the simulation does” (Roy et al. 2014, 2). Hieraus folgt, dass ähnlich wie die VIT Software die Einstellmöglichkeiten für den Anwender sehr gering sind und die Modellierung beispielsweise der Nullhypothese komplett automatisch erledigt wird. Hier stellt sich die Frage, ob dadurch nicht Lernchancen verloren gehen, weil gerade durch die Modellierung der Nullhypothese die zufällige Zuordnung spezifiziert wird. Weiter nennen die Autoren, dass es von Vorteil ist, „to run one trial of simulation/randomization at a time, before jumping to a large number of trials, say 1000” (Roy et al. 2014, 2).

Biehler und Frischemeier (2016) führen ein Beispiel aus, wie die Software TinkerPlots für die Durchführung eines Randomisierungstests eingesetzt werden kann und heben gerade die Visualisierungen, die durch diese Software möglich sind, hervor.

> „Aus didaktischer Perspektive ist außerdem zu betonen, dass bei diesem Vorgehen und bei der Simulation die Visualisierung des Zufallsprozesses durch die Zufallsmaschine in TinkerPlots sehr gut dargestellt wird: Die Lernenden sehen die zufällige Zuordnung durch die Kugeln sowie das Etikettieren und Bekleben als Idee des Randomisierungstests und können so den Prozess nachvollziehen.“ (Biehler & Frischemeier 2016, 14)

Es zeigt sich, dass eine Vielzahl von unterschiedlicher Software oder Tools eingesetzt wird, die jedoch alle mit dynamischen Visualisierungen arbeiten. Besonders zu Schwierigkeiten mit den jeweiligen Tools wird wenig berichtet. In dieser Arbeit soll nun gezielt analysiert werden, wie Studierende die Software TinkerPlots einsetzen, um einen Randomisierungstest durchzuführen und auch, ob sich an bestimmten Stellen im Prozess Schwierigkeiten bezüglich der Softwarebedienung ergeben.

[38] Diese speziellen Webapplets sind zu finden unter http://rossmanchance.com/ISIapplets.html (abgerufen am 29.06.2018)

6.6 Zusammenfassung

Der Artikel von G. Cobb (2007), in welchem er Randomisierungstests anpreist, um damit in Inferenzstatistik einzuführen, statt dem traditionellen amerikanischen Collegecurriculum zu folgen, hat einen Stein ins Rollen gebracht. Noch vor zehn Jahren haben Garfield und Ben-Zvi (2008a) festgestellt, dass es quasi keine Studien gäbe, in welchen Lernende beim Durchführen von Randomisierungstests untersucht werden. Nach Erscheinen des Artikels von Cobb wurden einige Curricula und Lernumgebungen entwickelt, die Randomisierungstests in das Zentrum stellen. Hier lassen sich zum Beispiel Rossman et al. (2008), Tintle et al. (2009) oder Zieffler und Catalysts for Change (2013) nennen, die komplette Curricula ausgearbeitet haben, in welchen über Randomisierungstests in die Inferenzstatistik eingeführt wird. Budgett et al. (2012) haben eine Lerneinheit dazu entwickelt und erfolgreich mit Schülern im letzten Schuljahr und im ersten Studienjahr durchgeführt. In den verschiedenen Umgebungen werden unterschiedliche Vorgehen favorisiert. Gemeinsam ist allen, dass sie einen Plan bzw. ein Schema vorgeben, welches die Struktur eines Randomisierungstests für Lernende vorgibt in Form von Konzepten, die benötigt werden. Besonderes Augenmerk wird dabei jedoch (fast) immer auf das Verständnis gelegt, so dass, um mit der Metapher von Schoenfeld (1998) zu sprechen, die Lernenden befähigt werden sollen zu kochen, anstatt nur ein Rezept zu befolgen. Die Lerneinheit von Budgett et al. (2012) zeigt, dass über Randomisierungstests innerhalb einer kurzen Zeitspanne in die inferentielle Denkweise eingeführt werden kann.

Für die Einführung in Randomisierungstests zeigt sich, dass vorgeschlagen wird, zunächst mit Hands-on Aktivitäten zu arbeiten und die Lernenden die (Re-)Randomisierung selbst durchführen zu lassen, bevor mit Computersimulationen gearbeitet wird (z. B. Budgett et al. 2012, G. Cobb 2007, Gould et al. 2010, Rossman et al. 2008, Tintle et al. 2014). Holcomb, Chance, Rossman, Tietjen, et al. (2010) räumen jedoch ein, dass Lernende, die zunächst eine Hands-on Aktivität zu einem Randomisierungstest durchführen und erst danach mit Computersimulationen arbeiten, später kein besseres Verständnis zeigen als Lernende, die direkt mit einer Computersimulation beginnen. Allerdings haben beide Gruppen in der Studie von Holcomb, Chance, Rossman, Tietjen, et al. (2010) die gleiche Zeit benötigt, so dass die Autoren den Einsatz von Hands-on Aktivitäten positiv bewerten, da es weder mehr Zeit kosten noch das Lernen behindern würde (Holcomb, Chance, Rossman, Tietjen, et al. 2010, 6).

Für die Wahl von Beispielen zeigt sich, dass hier vor allem herausgestellt wird, dass mit Daten realer Studien gearbeitet wird, die meistens einen recht geringen Stichprobenumfang (zwischen $n = 14$ (Budgett et al. 2012) und $n = 30$ (Noll & Kirin 2017) haben. Holcomb, Chance, Rossman, Tietjen, et al. (2010) diskutieren darüber hinaus, dass ein Einführungsbeispiel einen signifikanten P-Wert haben soll. Über einen besser oder we-

niger geeigneten Kontext der Beispiele findet sich kaum etwas. Jedoch zeigt sich, dass die gewählten Beispiele viele Bereiche abdecken und häufig dem medizinischen oder pädagogischen Fach entnommen sind, ohne dabei spezielles Fachwissen vorauszusetzen.

Das Führen einer Argumentation hinsichtlich eines Randomisierungstests wird explizit bei Holcomb, Chance, Rossman, Tietjen, et al. (2010) und Pfannkuch (2011) thematisiert (auch wenn die Autoren es nicht so nennen). Dabei wird in mehreren Sätzen ausgeführt, wie mit einem durchgeführten Randomisierungstest ein beobachteter Unterschied beurteilt werden kann und welche Schlüsse bezüglich möglicher Erklärungen gezogen werden können. Die oben ausgeführte Zusammenschau verschiedener Schemata zum Durchführen eines Randomisierungstests liefert das Gerüst für eine entsprechende Argumentation. Für den Aufbau einer solchen Argumentation wird das Nutzen der für einen Randomisierungstest nötigen Konzepte, die in diesem Kapitel identifiziert wurden, benötigt.

Als Schwierigkeit im Randomisierungstestprozess wird u. a. genannt, die Nullhypothese aufzustellen (Frischemeier 2017) und mit dem Nullmodell zu verbinden (Biehler et al. 2015, Noll & Kirin 2017), ebenso wie mögliche Erklärungen für beobachtete Unterschiede zu finden und für den weiteren Prozess zu nutzen (Budgett et al. 2012). Eine weitere, häufig genannte Schwierigkeit ist das Auffinden (Biehler et al. 2015) und Interpretieren (z. B. Budgett et al. 2013, Garfield & Ben-Zvi 2008a, Krauss & Wassner 2001,) von P-Werten.

Konkrete Analysen von Lernenden, die einen Randomisierungstest durchführen, sind immer noch selten. Budgett et al. (2012) untersuchen in einer Pilotstudie am Beispiel der Fischöl-Studie einige ausgewählte Aspekte, z. B. welche möglichen Erklärungen für beobachtete Unterschiede in einem Experiment von Lernenden genannt werden und wie die von Ihnen entwickelte VIT-Software die Durchführung eines Randomisierungstests unterstützt. In einer weiteren Publikation zu dieser Studie (Pfannkuch et al. 2015) konzentrieren sie sich darauf, Elemente zu identifizieren, die während des Durchführens eines Randomisierungstests von Lernenden adressiert werden sollten um hieraus Lehrempfehlungen zu generieren. Noll und Kirin (2017) haben sich in ihrer Studie anhand des Beispiels der Delfin-Therapie damit auseinander gesetzt, wie Lernende die Nullhypothese mit der Zufallsmaschine von TinkerPlots verknüpfen und wie sie diese verwenden. Biehler et al. (2015) haben untersucht, welche Schritte während eines Randomisierungstests Lernende erfolgreich durchführen können, und drei Welten (Kontext, Statistik, Software) identifiziert, in denen sich die Lernenden dabei bewegen. Ähnliche Untersuchungen hat auch Frischemeier (2017) durchgeführt mit dem Ergebnis, dass ein Randomisierungstestplan hohes Unterstützungspotential für Lernende birgt.

Die aus diesem Literaturüberblick gewonnen Erkenntnisse hinsichtlich der Wahl von Beispielen, dem Nutzen eines Schemas für die Durchführung eines Randomisierungstests, typischen Schwierigkeiten und Hinweisen für das Unterrichten einer Randomisierungstesteinheit wie beispielsweise von Fitch und Regan (2014) sollen in das Design der Lerneinheit einfließen, die in Kapitel 8 vorgestellt wird und die Grundlage für das Hintergrundwissen der Teilnehmer der Studie der vorliegenden Arbeit bildet.

In allen Studien wird deutlich, dass implizit oder explizit die in Kapitel 6.4 identifizierten Konzepte von Lernenden adressiert werden, allerdings wird dies in keiner der betrachteten Studien systematisch untersucht. Mit der vorliegenden Arbeit soll dies nun geschehen, ebenso wie ein systematischer Blick darauf, wie Software in dem Prozess eingesetzt wird. Die ausführliche Werkzeuganalyse von TinkerPlots für die stochastische Simulation ist hierfür eine Voraussetzung, die mit dem ersten Teil dieser Arbeit geschaffen wurde.

7 Forschungsfragen und Methodik für Teil II

In Abschnitt 7.1 werden die in 1.3 formulierten Ziele des zweiten Teils dieser Arbeit zu Forschungsfragen formuliert. Im folgenden Abschnitt 7.2 wird der Ansatz des Design-Based Research vorgestellt, in den die Studie eingebettet ist. Die methodische Anlage der Studie wird im Abschnitt 7.3 beschrieben und Abschnitt 7.4 befasst sich mit „mathematisch argumentieren“, wie es für diese Studie zugrunde gelegt wird. In Abschnitt 7.5 wird abschließend die Anlage der Studie kurz skizziert.

7.1 Forschungsfragen

Ein in 1.3 formuliertes Ziel für den zweiten Teil dieser Arbeit ist es, theoriegeleitet eine Lerneinheit zu konzipieren, die in die Logik des statistischen Schließens mit Randomisierungstests einführt. Als Forschungsfrage wird dazu formuliert:

1. Wie kann eine konkrete Lerneinheit zur Inferenzstatistik mit Randomisierungstests für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung konzipiert und gestaltet werden, welche die instrumentelle Genese der Lernenden zur Benutzung der Software TinkerPlots für Randomisierungstests anleitet und welche die in Kapitel 6 genannten Konzepte und Ideen umsetzt?

Eine solche Lerneinheit zur Inferenzstatistik soll auf Grundlage didaktischer Literatur zu Randomisierungstest theoriegeleitet erfolgen. Der Einsatz der Software TinkerPlots soll dazu gemäß Kapitel 2.2.2 instrumentell orchestriert werden. Zur Beantwortung der Frage ist es nötig, die Lerneinheit in eine Lehrveranstaltung zu implementieren und durchzuführen. Es wird als Theorie angenommen, dass Lernende nach Besuch der Lerneinheit in der Lage sind, mit dem dort erworbenen Wissen einen Randomisierungstest durchzuführen. Dies soll überprüft werden mit einer Fallstudie, in der die kognitiven Prozesse von Studierenden untersucht werden, die einen Randomisierungstest mit TinkerPlots durchführen. Um diese Prozesse aufzudecken, sollen die Studierenden paarweise an der Studie teilnehmen, da erwartet wird, dass dadurch Kommunikation zwischen den Teilnehmern angeregt wird, die (im Gegensatz zu den ablaufenden kognitiven Prozessen) direkt beobachtet werden kann. Hierzu werden als Forschungsfragen gestellt:

2. Inwieweit treten die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen auf? Welche unerwarteten Problemstellen können identifiziert werden?

3. Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots können festgestellt werden?

Hierzu sollen durch die Analyse der Bearbeitungen durch die Teilnehmer Hypothesen generiert werden. Eine Präzisierung der Forschungsfragen durch Unterforschungsfragen findet in Kapitel 10.3 statt.

In Bezug auf das Design der Lerneinheit sollen aus der Studie Ideen für ein mögliches Redesign gewonnen werden. Hierzu wird die letzte Forschungsfrage gestellt:

4. Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?

Um die Forschungsfragen zu beantworten, wird einerseits ein Ansatz benötigt, der das Designen einer Lerneinheit ermöglicht und anhand einer Studie Schlussfolgerungen für ein Redesign zulässt. Auf der anderen Seite wird ein Forschungsansatz für eine „Fallstudie“ benötigt. Die methodischen Ansätze dazu werden in den folgenden beiden Abschnitten 7.2 und 7.3 vorgestellt.

7.2 Design Based Research

Design-Based Research ist ein Forschungsansatz für Lernforschung, bei dem Theorie und Praxis eng miteinander verzahnt werden. Der Ansatz wird beschrieben durch

> „Educational design-based research (DBR) can be characterized as research in which the design of educational materials (e.g. computer tools, learning activities, or a professional development program) is a crucial part of the research. That is, the design of learning environments is interwoven with the testing or developing of theory.” (Bakker & van Eerde 2015, 430)

Design Based Research wird verschiedentlich auch als Design Experiment oder Design Research bezeichnet. Es handelt sich um einen zyklischen Ansatz, nach dem theoriegeleitet Lernumgebungen und Werkzeuge entwickelt und analysiert werden sollen (P. Cobb, Confrey, diSessa, Lehrer & Schauble 2003). Die folgende Darstellung beruht auf Bakker und van Eerde (2015); P. Cobb et al. (2003) und The Design-Based Research Collective (2003).

Es gibt zwei Hauptziele, die mit dem Design-Based Research Ansatz verfolgt werden. Einerseits sollen Lerntheorie und Praxis miteinander verbunden werden, andererseits sollen, im Unterschied zu vielen anderen Methoden der empirischen Sozialforschung, Design-Experimente durchgeführt werden, um aus diesen Theorien oder „Prototheorien“ (The Design-Based Research Collective 2003, 5) zu entwickeln.

P. Cobb et al. (2003) unterscheiden für Design Experimente zwischen drei Phasen. Zu Beginn steht die Phase „preparing for a design experiment“ (P. Cobb et al. 2003, 11), in der vorhandene Literatur als Grundlage aufgearbeitet werden soll und spezifische Lerninhalte und –ziele festgelegt werden sollen. Für die vorliegende Arbeit ist dies in den Kapiteln 5 und 6 geschehen. Darauf aufbauend soll eine Lerneinheit designt werden, für die u. a. empfohlen wird, Software zu integrieren. Vermutungen über den Lernprozess und mögliche Unterstützungsmaßnahmen durch den Lehrenden sollen hier aufgestellt werden. Dies alles zusammen bildet eine „hypothetische Lerntrajektorie“ (Bakker & van Eerde 2015, 439), die im Forschungsprozess weiterentwickelt werden soll.

Die zweite Phase „conducting a design experiment“ (P. Cobb et al. 2003, 11) beinhaltet die Durchführung der Intervention. Dabei kann es geschehen, dass „the teacher or researcher feels the need to adjust the HLT [hypothetical learning environment] or instructional activity for the next lesson” (Bakker & van Eerde 2015, 439). Dies zielt auf Erkenntnisse, die während der Durchführung gewonnen werden durch permanente Analyse des Prozesses. Daraus ergibt sich, dass eine Lernumgebung während der Durchführung geändert werden darf.

In der dritten Phase „conducting retrospective analyses“ (P. Cobb et al. 2003, 12) findet eine retrospektive Analyse der Lerneinheit statt, für die betont wird, dass diese erst nach Beendigung der Durchführung der gesamten Einheit stattfinden soll. Das Ziel liegt in der Überarbeitung der Lerneinheit und der aufgestellten Theorien. Für die retrospektive Analyse schlagen Bakker und van Eerde (2015, 442) als eine von zwei möglichen Analysen eine „task oriented analysis“ vor, die Daten bezüglich des Lernprozesses der Lernenden vergleichen soll. An dieser Stelle wird in der vorliegenden Arbeit die Fallstudie gesetzt (Abschnitt 7.3), um einen solchen Bearbeitungsprozess einer Aufgabe durch Lernende, die an einer designten Lerneinheit (Kapitel 8.2.2) teilgenommen haben, zu untersuchen. Zu der Fallstudie wird im nächsten Abschnitt 7.3 der Forschungsansatz beschrieben.

Der zyklische Charakter des Design-Based Research Ansatzes manifestiert sich daran, dass aufgrund der retrospektiven Analyse ein Zurückspringen in die erste Phase (preparing for a design experiment) stattfindet und das Design der Lerneinheit angepasst wird, um daraufhin die folgenden Phasen erneut zu durchlaufen.

7.3 Qualitative Forschung

Wie in Kapitel 6 gezeigt wurde, gibt es einzelne Studien, in denen untersucht wird, wie Studierende eine Randomisierungstestaufgabe bearbeiten. In der Studie von Noll und Kirin (2017) wurde TinkerPlots als Werkzeug verwendet, um den speziellen Fragen nachzugehen „How do students connect the null hypothesis with the TinkerPlots model

they create?“ und „How do students select or design TinkerPlots models with the sampler tool when given a comparing two groups problem?” (Noll & Kirin 2017, 214). In ihren Analysen legen die Autorinnen einen starken Fokus auf die technische Umsetzung der Simulation in TinkerPlots. Die Argumentationen der Lernenden beziehen sich größtenteils auf technische Aspekte und werden nicht in einen größeren Zusammenhang eingeordnet, sondern lokal interpretiert. Biehler et al. (2015) haben in ihrer Studie gezeigt, inwieweit Studierende die Schritte eines Randomisierungstests erfolgreich durchführen können, allerdings ebenfalls ohne die Argumentationsprozesse genauer in den Blick zu nehmen. Pfannkuch et al. (2015) verweisen darauf, dass es im Bereich Inferenzstatistik mit Bezug zu Kausalität kaum Studien gibt und untersuchen Lernende, die eine Randomisierungstestaufgabe mit der VIT-Software bearbeiten. Im deutschsprachigen Raum sind die Argumentationsprozesse von Lernenden beim Durchführen eines Randomisierungstests mit TinkerPlots noch kaum erforscht, die vorliegende explorative Fallstudie soll hier einen Beitrag liefern.

In der vorliegenden Studie werden die Kommunikationsprozesse von Lernenden beim Bearbeiten eines Randomisierungstestproblems analysiert. Hierzu sollen die in Kapitel 6 identifizierten Konzepte (beobachtete Daten, Nullhypothese, etc.) genutzt und überprüft werden, um hieraus Hypothesen zu generieren, wie Lernende argumentieren und an welchen Stellen sich Schwierigkeiten ergeben. Aus diesem Grund soll ein qualitatives Design verwendet werden, durch welches intensiv an einer kleinen Personenzahl die Argumentationsprozesse analysiert werden können.

In der qualitativen Forschung sollen im Gegensatz zur quantitativen Forschung die „Beforschten“ selbst zu Wort kommen, da angenommen wird, dass Menschen selbstreflexive Subjekte sind, die als Experten ihrer selbst handeln (Mayring 2016). In der qualitativen Forschung sollen bisher unbekannte Phänomene exploriert werden oder neue Theorien und Modelle entwickelt werden. Dadurch ist sie durch eine induktive Vorgehensweise geprägt. Da der Forscher jedoch bereits vorhandenes Wissen und Annahmen bezüglich des Forschungsfeldes hat, ist die qualitative Forschung eine Verbindung aus induktions- und deduktionsgeleiteten Schritten und die passende Methode für das Forschungsinteresse des zweiten Teils dieser Arbeit.

Ein Überblick über mögliche Designs qualitativer Forschung findet sich zum Beispiel bei Flick (2013). Flick unterscheidet zwischen Fallstudien, Vergleichsstudien, retrospektiven Studien, Momentaufnahmen und Längsschnittstudien. Für die vorliegende Forschung ist das Design der Fallstudie zu wählen, um dadurch eine genaue Beschreibung eines Falls zu ermöglichen.

7.3.1 Fallstudie

Eine Fallstudie „zielt auf die genaue Beschreibung oder Rekonstruktion eines Falls ab“ (Flick 2013, 253) und „Fallstudien können grundsätzlich Antworten auf explorative, deskriptive und/oder explanative Fragen liefern“ definieren Borchardt und Göthlich (2007, 35). Häder (2015) nennt als Untersuchungseinheit entweder eine einzelne Person, eine Personengruppen, eine Organisation oder sogar eine Gesellschaft, die durch Einzelfallstudien analysiert werden können.

Um Einzelfallstudien wissenschaftlich verwertbar zu machen, stellt Mayring (2016) einen Vorgehensplan auf, der die folgenden fünf Punkte enthält:

- Fragestellung der Fallanalyse
- Falldefinition
- Bestimmung der Methode und Sammlung des Materials
- Aufbereitung des Materials, ggf. durch Fallzusammenfassung und/oder Fallstrukturierung
- Einordnen des Einzelfalls in einen größeren Zusammenhang oder Vergleich mit anderen Fällen (nach Mayring 2016, 43f)

In der Fragestellung soll beschrieben werden, was der Zweck der Fallanalyse sein soll. Die Falldefinition bestimmt, was als Fall gilt und welche Fälle ausgewählt werden. Hier können zum Beispiel besonders typische Fälle genommen werden, extreme Fälle, besonders häufige oder auch besonders seltene Fälle. Es können einzelne Personen ausgewählt werden, oder kleine Gruppen von zwei bis drei Personen, um durch die Interaktion zwischen den Gruppenmitgliedern weitere Einsichten zu gewinnen. Dies wird in der vorliegenden Arbeit gewählt. Zur Aufbereitung des Materials gehört dessen Fixierung (z. B. durch Audioaufnahmen) und für die weitere Bearbeitung eine Fallzusammenfassung, in der zum Beispiel die wichtigen Eckdaten zusammengefasst werden. Alternativ oder ergänzend kann das Material durch die Fallstrukturierung in einzelne Kategorien gegliedert werden. Am Ende schließlich soll die Gültigkeit der Ergebnisse abgeglichen werden, indem ein Zusammenhang zu anderen Fällen hergestellt wird. Dieser letzte Punkt, der Vergleich mit anderen Fällen, kann auch direkt als Design angelegt sein:

> „Der Vorteil einer vergleichenden Fallstudie gegenüber einer Einzelfallstudie liegt darin, dass die gewonnenen Erkenntnisse durch Ähnlichkeiten und Unterschiede zwischen den Fällen kritisch beleuchtet werden können. Aus diesem Grund gelten die Ergebnisse vergleichender Fallstudien als überzeugender, vertrauenswürdiger und robuster.“ (Borchardt & Göthlich 2007, 36)

Wobei die Autoren hier keine Vergleichsstudien im Sinne eines Forschungsdesigns von Flick (2013) meinen, sondern eher auf die „Zwischenstufe“ zwischen beidem zielen, die Flick definiert als

> „Eine Zwischenstufe [zwischen Einzelfall und Vergleichsstudie] stellt die Verbindung mehrerer Fallanaylsen dar, die zunächst als solche durchgeführt werden und dann komparativ oder kontrastierend gegenübergestellt werden.“ (Flick 2013, 254)

Durch die detaillierte Untersuchung von drei Einzelfällen und deren Vergleich in Kapitel 11 wird in der vorliegenden Arbeit eine solche Zwischenstufe genutzt. Trotzdem muss mit den Ergebnissen einer Einzelfallstudie vorsichtig umgegangen werden. „Fallstudien erlauben, hierin liegt ihre Limitation, im Gegensatz zu quantitativen Methoden keinen statistischen Induktionsschluss auf eine Grundgesamtheit“ (Borchardt & Göthlich 2007, 36). Die Erkenntnisse sind also zunächst auf die untersuchte Stichprobe begrenzt.[39] Dies kann jedoch auch zum Vorteil genutzt werden. Als besonderen Nutzen von Einzelfallanalysen nennt Mayring, dass „je weniger Versuchspersonen analysiert werden, desto eher kann man auf die Besonderheiten des Falles eingehen, desto genauer kann die Analyse sein“ (Mayring 2016, 42). Häder (2015) sieht ebenfalls das Problem der Generalisierbarkeit der Ergebnisse aus Einzelfallstudien, nennt jedoch weitere Funktionen, die die Nützlichkeit dieser Methode in der qualitativen Forschung betonen. Dazu gehören die Erkundung unklarer Sachverhalte, mit dem Ziel, Hypothesen zu bilden und dadurch standardisierte Erhebungen vorzubereiten (Häder 2015, 359). Auch für die Analyse besonders typischer oder besonders seltener Fälle ist dies die Hauptmethode. Und schließlich lassen sich Zusammenhänge durch Einzelfallstudien anschaulich darstellen. Primäre Anwendungsgebiete von Einzelfallstudien sieht er in der pädagogischen, sonderpädagogischen oder klinischen Forschung (Häder 2015, 360), wodurch sie für die vorliegende Arbeit prädestiniert sind.

Nach diesen Darstellungen ist die Einzelfallstudie eine geeignete Methode um die Argumentationsprozesse Lernender während des Lösungsprozesses einer Randomisierungstestaufgabe mit TinkerPlots zu analysieren. Gerade die „Zwischenstufe“ als Verbindung weniger Einzelfallstudien, die Flick (2013) nennt, ist ein Ansatz, mit dem das Material gut analysiert werden kann und der in Verbindung mit den Erkenntnissen aus anderen Studien genutzt werden kann, um erste generelle Erkenntnisse abzuleiten. Auch der explorative Charakter der Studie wird durch die Methode unterstützt.

Als Erhebungsverfahren bietet sich das qualitative Interview an. Ein Überblick über verschiedene Formen von Interviews findet sich zum Beispiel bei Flick (2016) oder Döring und Bortz (2016). Für die vorliegende Arbeit wird ein Leitfadeninterview ausgewählt und mit Ansätzen der Verwendung des Interviews in der mathematikdidaktischen Forschung kombiniert, die von Beck und Maier (1993) beschrieben werden.

---

[39] Wodurch sich eine interessante Parallele zu den Schlussfolgerungsmöglichkeiten aus einem Randomisierungstest ergibt.

7.3.2 Erhebungsverfahren: Das Leitfadeninterview

In der qualitativen Forschung spielt das Gespräch eine besondere Rolle (Mayring 2016), deswegen werden vor allem Interviews als Datenerhebung verwendet. Grundsätzlich lassen sich Interviews im Grad ihrer Strukturiertheit unterscheiden (Döring & Bortz 2016). Dem unstrukturierten Interview (auch narratives Interview) liegt kein zuvor entwickeltes Instrument zugrunde und dieses verläuft sehr unterschiedlich. Verschiedene Interviews zum selben Forschungsthema sind deshalb kaum vergleichbar. Vollstrukturierte Interviews basieren auf einem Interviewfragebogen mit geschlossenen Fragen mit exakt vorgegebenen Antworten. Subjekte Ansichten oder Argumente des Interviewten werden hier nicht erfasst. Die dritte Form des Interviews ist das halbstrukturierte Interview, dessen Fragen auf einem Leitfaden basieren. Dieser Leitfaden beinhaltet meist offene Fragen, die von den Interviewten in eigenen Worten beantwortet werden sollen. Dabei ist der Leitfaden jedoch nicht als starres Instrument anzusehen, sondern

> „der Leitfaden gibt die Fragen und ihre Reihenfolge vor, erlaubt aber den Interviewenden je nach Interviewsituation in gewissen Grenzen individuelle Anpassungen. So können Fragen vorgezogen, übersprungen oder vertieft werden, um den Interviewfluss am Laufen zu halten.“ (Döring & Bortz 2016, 358)

Auch müssen die Fragen des Leitfadens in der Interviewsituation nicht wörtlich wiedergegeben werden, sondern „können und sollten von den Interviewenden passend zur Situation und zum Gegenüber formuliert werden“ (Döring & Bortz 2016, 372). Bei Mayring (2016) findet sich darüber hinaus noch der Hinweis, dass in einem Interview Aspekte auftreten können, die im Leitfaden nicht bedacht sind, aber für die „Erhaltung des Gesprächsfadens bedeutsam sind“ (Mayring 2016, 70). In solchen Situation solle der Interviewende spontane Fragen formulieren, die ebenfalls den Interviewfluss fördern. Somit kann das Leitfadeninterview individuell auf die jeweilige Situation angepasst werden. Beim Leitfadeninterview handelt es sich deshalb um kein standardisiertes Verfahren, vielmehr „besteht die Standardisierung nicht in exakt vorgegebenen Fragenformulierungen, sondern in der inhaltlichen Vergleichbarkeit des Fragenkatalogs“ (Döring & Bortz 2016, 372).

Beck und Maier (1993) erarbeiten verschiedene Aspekte an Interviews, die für die Verwendung in der Mathematikdidaktik eine Rolle spielen. In Bezug auf Einzelfallstudien, wie sie in der vorliegenden Arbeit durchgeführt werden sollen, handelt es sich nach Beck & Maier um ein typisches Verfahren in der Mathematikdidaktik, deren Zweck das „Auffinden und Herausarbeiten des Typischen von z. B. Lern- und Verstehensprozessen“ (Beck & Maier 1993, 153) ist. Interviewformen und –instrumente sollen der sozialwissenschaftlichen Forschung entlehnt werden, aber „diese Instrumente [sind] eventuell für die speziellen Bedürfnisse eines mathematikdidaktischen Forschungsprojekts abzuändern“ (Beck & Maier 1993, 153). Bezüglich abzuändernder oder neu zu entwickelnder Forschungsinstrumente nennen sie folgende Punkte.

- „*Vorgaben*: Häufig werden der Interviewte bzw. die Interviewten aufgefordert, bestimmte mathematische Aufgaben zu lösen, (…) und eventuell dabei ‚laut zu denken' (…). Die Aufgaben können seitens der Forscher selbst konstruiert werden oder Forschungsberichten anderer Autoren entnommen sein, eventuell in modifizierter Form. (…) Die Benutzung des [optionalen] Arbeitsmaterials bzw. der Zeichnung mag freigestellt oder verpflichtend sein.
- *Spezifität*: Der Interviewte kann genaue Handlungsaufträge erhalten, (…). Oder der Handlungsauftrag kann auch weniger spezifisch gehalten sein (…).
- *Interaktivität*: Der Interviewer kann seine Instruktionen in Abhängigkeit vom Interviewten modifizieren, z. B. werden bei Nicht-Lösung oder Antwortverweigerung zusätzliche Hilfen und Instruktionen gegeben.
- *Strukturierung*: Das Interview kann in verschiedene Phasen gegliedert sein: z. B. spontane Aufgabenlösung als erste Phase, unterstützte Aufgabenlösung als zweite Phase oder in eine Phase der ausführlichen Aufgabenbeschreibung und eine Phase der Aufgabenlösung.
- *Dauer*: Die Dauer des Interviews kann relativ klar auf eine bestimmte Zeit fixiert sein, oder sie kann, je nach Interviewverlauf, innerhalb einer weiten Zeitspanne streuen." (Beck & Maier 1993, 154f)

Für die vorliegende Arbeit sollen die Argumentationen von Lernenden beim Randomisierungstesten analysiert werden. Dazu wird ein Leitfadeninterview erstellt, das die besonderen Punkte berücksichtigt, die Beck & Maier in ihrem Artikel anführen. Hiermit soll versucht werden, Einblicke in die kognitiven Prozesse der Lernenden zu bekommen. Hierzu ist es nützlich, wenn Teilnehmer paarweise an einem Interview teilnehmen, da so das von Beck & Maier angesprochene „laute Denken" Teil des natürlichen Kommunikationsvorgangs beim gemeinsamen Aufgabenbearbeiten ist. Da im Leitfadeninterview keine Antwortmöglichkeiten vorgegeben werden, können die Teilnehmer freie Antworten geben und Argumentationen im gemeinsamen Dialog entwickeln. Dabei soll der Leitfaden für das Interview wie oben ausgeführt Fragen vorgeben, die durch die Interviewerin[40] an die jeweiligen Äußerungen spontan angepasst werden. Der Vorteil dieser Methode liegt darin, dass durch den Leitfaden konkrete Fragen gestellt werden können, das Gespräch sich aber trotzdem in gewissem Rahmen in Richtungen entwickeln kann oder Aspekte vertiefen kann, die den Teilnehmern des Interviews wichtig sind. Dem Interviewenden obliegt die Steuerung des gesamten Prozesses mit der Freiheit, die Reihenfolge der Fragen zu verändern oder vertiefende Nachfragen zu stellen. So sollen auch in der vorliegenden Studie zusätzliche Hilfen gegeben werden, wenn die Interviewten diese benötigen.

7.4 Mathematisch Argumentieren beim Randomisierungstest

Wie sich aus den Darstellungen in Kapitel 5 und 6.3 erkennen lässt, stellt der Randomisierungstest keinen formalen mathematischen Beweis im klassischen Sinn dar. „Unter einem mathematischen Beweis versteht man die deduktive Herleitung eines mathemati-

[40] Die Interviewerin ist in der vorliegenden Arbeit gleichzeitig die Autorin.

schen Satzes aus Axiomen und zuvor bereits bewiesenen Sätzen nach spezifizierten Schlussregeln.“ (Jahnke & Ufer 2015, 331). In der Inferenzstatistik werden jedoch keine Sätze bewiesen, sondern die am Ende stehende Beurteilung einer (Null-)Hypothese geschieht unter Restunsicherheit, es gibt also keine Sicherheit darüber, die „wahre“ Hypothese gefunden zu haben. Dennoch spielt ähnlich wie beim streng logischen Beweisverfahren der Argumentationsprozess eine wichtige Rolle. Ausgehend von einer Behauptung, in der Inferenzstatistik von einer Nullhypothese ausgeht, wird eine Argumentation aufgebaut, die zu einer Beurteilung der Nullhypothese führt. Wenn diese Argumentation logisch aufgebaut und dokumentiert ist, so kann sie „intersubjektiv Geltung beanspruchen“ (Brunner 2014, 7), von anderen nachvollzogen und geprüft werden. Eine solche Argumentation in der Inferenzstatistik ist eine induktive Argumentation, da im Allgemeinen vom Speziellen auf die Gesamtheit geschlossen werden soll. Dass dies bei Randomisierungstests nicht möglich ist, wurde bereits hinlänglich geklärt, aus diesem Grund wird eine solche Argumentation nicht als „gültig“, sondern als korrekt bezeichnet (vgl. Bayer 2007 oder Brunner 2014). Der Argumentationsprozess spielt jedoch bei einem solchen statistischen Schluss eine wichtige Rolle.

Im Sinne der Bildungsstandards, aber auch der NCTM-Standards ist mathematisch Argumentieren eine wichtige prozessbezogene Kompetenz bei der Auseinandersetzung mit mathematischen Inhalten. Allerdings wird der Argumentationsbegriff unterschiedlich verwendet und ist nach Schwarz, Hershkowitz und Prusak (2010, 116) ein „multifaceted term with different meanings“. Eine verbreitete internationale Definition ist die folgende:

> „Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a ‚rational judge‘. (Eemeren et al. 1996, 5)

Bei dieser Definition steht die soziale Dimension im Vordergrund in Verbindung mit einer Kontroverse zwischen den Beteiligten. In mathematischen Argumentationen steht jedoch selten eine Kontroverse im Mittelpunkt, wichtiger ist der dialogische Charakter einer Argumentation. Diesem Aspekt wird in der folgenden Definition stärker Rechnung getragen, in der das interaktionistische Element einer Argumentation stärker herausgestellt wird.

> „Unter einer Argumentation wird ein *zwischenmenschlicher Prozess* verstanden, der folgendermaßen gekennzeichnet wird: Zum einen wird öffentlich ein Begründungsbedarf angezeigt und zum anderen wird versucht, diesen Begründungsbedarf zu befriedigen.“ (Schwarzkopf 2001, 255)

Mathematisch Argumentieren wird als eine zentrale Kompetenz angesehen, die sich sowohl in internationalen (Common Core State Standards Initiative 2012, NCTM 2000) als auch in nationalen Bildungsstandards (Blum, Drüke-Noe, Hartung & Köller 2006,

Kultusministerkonferenz 2003) wiederfindet. Nach den deutschen Bildungsstandards gehört zu „mathematisch argumentieren“:

- „Fragen stellen, die für die Mathematik charakteristische sind (‚Gibt es…?‘, ‚Wie verändert sich…?‘, ‚Ist das immer so…?‘) und Vermutungen begründet äußern,
- Mathematische Argumentationen entwickeln (wie Erläuterungen, Begründungen, Beweise),
- Lösungswege beschreiben und begründen.“ (Kultusministerkonferenz 2003, 8)

Hierbei wird jedoch nicht ausdifferenziert, welche Unterschiede zwischen Erläuterungen, Begründungen und Beweisen liegen. All dies wird benötigt, um den Argumentationsprozess während eines Randomisierungstests zu durchlaufen. Hier muss vor allem im Anforderungsbereich III: *Verallgemeinern und Reflektieren* gearbeitet werden, der in den Bildungsstandards beschrieben wird als „das Bearbeiten komplexer Gegebenheiten u. a. mit dem Ziel, zu eigenen Problemformulierungen, Lösungen, Begründungen, Folgerungen, Interpretationen oder Wertungen zu gelangen“ (Kultusministerkonferenz 2003, 13). Im Gegensatz zu anderen Gebieten der Mathematik handelt es sich bei inferentiellen Schlussfolgerungen nicht um (mathematische) Beweise, da solche bei inferentiellen Schlussfolgerungen gerade nicht möglich sind. Vielmehr wird eine Argumentation aufgebaut, unter der eine bestimmte Annahme begründet abgelehnt werden soll. Es muss also im Laufe des Argumentationsprozesses begründet werden, weshalb eine bestimmte Schlussfolgerung gezogen wird. In Anlehnung an Meyer und Prediger (2009) wird hier Begründen als ein im Gegensatz zu Beweisen breiterer Begriff gesehen, da sich Beweisen häufig eng auf axiomatisch-deduktives Schließen bezieht. Mit dem Argumentieren erhält das Begründen eine zusätzliche kommunikative Dimension, da die Begründungen logisch aufgebaut und versprachlicht werden müssen (Meyer & Prediger 2009; Brunner 2014). In der Literatur wird die didaktische Diskussion bezüglich des Argumentierens meist im Hinblick auf Beweise geführt (Brunner 2014, Jahnke & Ufer 2015, Krummheuer 2003). Im Zusammenhang mit inferentiellen Schlussfolgerungen wird eine aus mehreren Schritten bestehende Argumentation benötigt, um die Schlussfolgerung zu stützen.

Was genau ist nun ein Argument? Hierzu definiert Bayer (2007, 18, Hervorhebung im Original):

> „Ein Argument besteht immer aus mehreren Sätzen: der **Konklusion**, dem Satz, den wir begründen wollen, und aus einer oder mehreren **Prämissen**, welche die Konklusion stützen sollen. (…) Argumente sind also Mengen von Sätzen; **Argumentationen** sind dagegen sprachliche Handlungen, bei deren Vollzug wir ein Argument oder auch mehrere miteinander verknüpfte Argumente äußern. Wir argumentieren, um Behauptungen zu begründen oder Entscheidungen zu rechtfertigen.“

In der Interviewstudie dieser Arbeit sollen während des Randomisierungstests Argumentationen im Dialog zwischen zwei Studierenden entwickelt werden, so dass zusätzlich noch der Kompetenz Kommunizieren eine Bedeutung zukommt. In der obigen Darstellung ist bereits ein interaktionistisches Element in Argumentationen enthalten, so

beispielsweise in der Definition von Schwarzkopf (2001), die explizit einen „zwischenmenschlichen Prozess“ zugrunde legt. Laut den Bildungsstandards gehört zur Kompetenz Kommunizieren

- „Überlegungen, Lösungswege bzw. Ergebnisse dokumentieren, verständlich darstellen und präsentieren, auch unter Nutzung geeigneter Medien,
- die Fachsprache adressatengerecht verwenden,
- Äußerungen von anderen und Texte zu mathematischen Inhalten verstehen und überprüfen.“ (Kultusministerkonferenz 2003, 9)

Gerade in einem Interaktionsprozess des gemeinsamen Aufgabenlösens wird diese Kompetenz benötigt.

7.5 Anlage der Studie

Der Fokus dieses Teils der vorliegenden Arbeit soll auf der Analyse der Bearbeitungsprozesse der Teilnehmer der Studie liegen. Im Sinne des Design-Based Research wird aufbauend auf Literatur zum Einsatz von Randomisierungstests (Kapitel 5 und 6) zunächst eine hypothetische Lerntrajektorie entwickelt „preparing for a design experiment“, die in Kapitel 8 beschrieben wird. Diese Lerneinheit wurde im Sommersemester 2017 in bestehende Lehrveranstaltungen implementiert und durchgeführt „conducting a design experiment“. Eine direkte retrospektive Analyse dieser Einheit geschieht nur kurz in Abschnitt 8.3 „conducting retrospective analyses“, denn hierfür soll vor allem die Fallstudie genutzt werden. Für die Fallstudie wird eine Aufgabe aus der Literatur adaptiert und eine a priori Analyse durchgeführt, um einen möglichen Lösungsweg durch Studierende aufzuzeigen, die an der Lerntrajektorie teilgenommen haben. Gleichzeitig sollen in der Aufgabenanalyse mögliche Schwierigkeiten aufgedeckt werden. Die Analysen der Kommunikations- und Bearbeitungsprozesse der Teilnehmer sollen einerseits genutzt werden, um Faktoren für gelingende kollaborative Bearbeitungsprozesse zu identifizieren und andererseits, um retrospektiv Erkenntnisse für das Redesign der Lerntrajektorie zu gewinnen.

8 Die Lerneinheit „Inferenzstatistik mit Randomisierungstests“

Im fünften Kapitel wurden Grundlagen der Inferenzstatistik mit Randomisierungstests erläutert und der Einsatz von Randomisierungstests motiviert. Im sechsten Kapitel wurde der Forschungsstand dazu aufgearbeitet. Es wurden Schwierigkeiten und Fehlvorstellungen von Lernenden berichtet, Designideen zu Lernumgebungen vorgestellt und Konzepte extrahiert, die beim Durchführen eines Randomisierungstests adressiert werden sollten. Auf dieser Grundlage soll nun eine Lerneinheit „Inferenzstatistik mit Randomisierungstests unter Verwendung der Software TinkerPlots“ für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung konzipiert werden, die in eine bestehende Veranstaltung implementiert werden kann.

Als Forschungsfrage hierzu wird die folgende formuliert:

1. Wie kann eine konkrete Lerneinheit zur Inferenzstatistik mit Randomisierungstests für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung konzipiert und gestaltet werden, welche die instrumentelle Genese der Lernenden zur Benutzung der Software TinkerPlots für Randomisierungstests anleitet und welche die in Kapitel 6 genannten Konzepte und Ideen umsetzt?

Konkretisiert wird dies durch zwei Unterforschungsfragen.

- Welche Inhalte soll eine solche Einheit umfassen?
- Welche Maßnahmen lassen sich für eine Lerneinheit zur Inferenzstatistik erstellen, um den Lernprozess zu unterstützen?

Dabei liegt der Fokus dieses Kapitels darauf, die im Sinne des Design Based Research Ansatzes (Kapitel 7.2) designte Lerneinheit zu präsentieren. Aufschluss über die Wirksamkeit und Nachhaltigkeit der Lerneinheit, auch im Sinne einer retrospektiven Analyse, gibt die Fallstudie, die im Anschluss an die Lerneinheit durchgeführt wird (Kapitel 9 bis 11). Die hierzu formulierten Forschungsfragen sind in Kapitel 7.1 zu finden und werden in Kapitel 10.3 durch Unterforschungsfragen präzisiert.

Als passender Rahmen zur Erprobung der neu designten Lerneinheit „Inferenzstatistik mit Randomisierungstests“ wird die Veranstaltungssequenz „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ ausgewählt, so dass die Lerneinheit für die dortigen Bedingungen konzipiert wird.

8.1 Rahmenbedingung: Die Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ im Sommersemester 2017

Die Veranstaltung „Modellieren, Größen, Daten und Zufall I“ als Vorlesung in Verbindung mit dem Seminar „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“, in welchem Inhalte der Vorlesung vertieft und didaktisch reflektiert werden, existiert seit 2013 als Einheit an der Universität Paderborn und wird seitdem jedes Jahr kooperativ von Rolf Biehler, Daniel Frischemeier und der Autorin dieser Arbeit im Sommersemester durchgeführt und kontinuierlich weiter entwickelt. Inhaltlich bestanden die Veranstaltungen von 2013 bis 2016 aus drei Bausteinen, die in unterschiedlichem Umfang thematisiert wurden. Die drei Bausteine Datenanalyse, Kombinatorik und Wahrscheinlichkeitsrechnung sollten im Sommersemester 2017 erstmalig zum Ende der Veranstaltung durch einen vierten Baustein zur Inferenzstatistik ergänzt werden.

Die Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ (Vorlesung) und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ (Seminar) können als Einheit betrachtet und mit dem Konstrukt einer Vorlesung mit Übung verglichen werden, da von Beginn an eine starke Verzahnung der Inhalte stattfindet. Die Veranstaltungen sind laut Studienverlaufsplan der „Prüfungsordnung für den Bachelorstudiengang Lehramt an Grundschulen mit dem Lernbereich Mathematische Grundbildung an der Universität Paderborn[41]“ für das vierte Bachelorsemester von Grundschullehramtsstudierenden vorgesehen. Die Veranstaltung „Modellieren, Größen, Daten und Zufall I“ ist Teil des Basismoduls Ba3 „Modellieren, Größen, Daten und Zufall“, das im jeweils nachfolgenden Wintersemester mit der Veranstaltung „Modellieren, Größen, Daten und Zufall II“ abschließt. Für die Vorlesung im Sommersemester ist der Inhaltsbereich „Daten und Zufall“ vorgesehen und für die Vorlesung im Wintersemester der Inhaltsbereich „Modellieren und Größen“. Da die Vorlesung „Modellieren, Größen, Daten und Zufall II“ inhaltlich losgelöst von der Veranstaltung „Modellieren, Größen, Daten und Zufall I“ ist, wird im Rahmen dieser Arbeit nicht weiter darauf eingegangen.

Im Folgenden werden Inhalte der Vorlesung und des Seminars dargestellt, wie sie im Sommersemester 2017 an der Universität Paderborn stattgefunden haben.

Allgemeines Ziel der Veranstaltungen ist, dass angehende Primarstufenlehrer statistische Aktivitäten im Sinne der Leitidee „Daten, Häufigkeit und Wahrscheinlichkeit“ von Hasemann und Mirwald (2012), die solche Aktivitäten für die Grundschule fordern, fachlich verstehen und didaktisch reflektieren. Dabei sollen die Veranstaltungen sowohl

[41] Die Prüfungsordnung ist im Archiv der Universitätsbibliothek abrufbar: https://digital.ub.uni-paderborn.de/ihd/content/pageview/1066429 (abgerufen am 29.06.2018).

nationalen (z. B. Kurtzmann & Sill 2012, Sill 2018) als auch internationalen (z. B. Batanero, Burrill & Reading 2011) Forderungen der Stochastikdidaktik genügen, wonach sowohl die fachliche als auch die fachdidaktische stochastische Ausbildung von Lehramtsstudierenden des Grundschullehramts gefordert wird. Der kontinuierliche Einsatz der Software TinkerPlots erst zur Datenanalyse und später zur stochastischen Simulation und damit der Aufbau der instrumentellen Genese (Rabardel 2002) kommt einer dieser Forderungen nach und bildet einen Schwerpunkt der Veranstaltungen. Die Wahl der Software TinkerPlots stammt aus den umfangreichen Analysen, die im Rahmen der vorliegenden Arbeit zum Unterstützungspotential von TinkerPlots zur stochastischen Simulation durchgeführt wurden (Kapitel 3 und 4) und weiteren, von Frischemeier (2017) durchgeführten Analysen im Bereich der Datenanalyse. Diese Analysen und in früheren Lehrveranstaltungen gesammelte Erfahrungen (Podworny 2013, Podworny 2018, Podworny & Biehler 2014) führten dazu, dass die Software TinkerPlots von Beginn der Veranstaltungen aktiv eingesetzt wird. Dabei bekommen die Studierenden sowohl die Gelegenheit, die Software aus Lernerdensicht zu erfahren, um die individuelle instrumentelle Genese zu durchlaufen, als auch aus Lehrersicht zu reflektieren und dadurch Hinweise zur möglichen instrumentellen Orchestrierung (Guin & Trouche 2002) zu erhalten. Somit wird für die stochastische Simulation die instrumentelle Genese der Lernenden im Sinne des ersten Teils dieser Arbeit in den Veranstaltungen angeleitet.

Im Sinne von Wassong und Biehler (2010) sollen durch die Veranstaltungen sowohl das „content knowledge“, das „technological content knowledge“, das „pedagogical content knowledge“ sowie das „technological pedagogical content knowledge“ der Teilnehmer entwickelt werden. Kurz dargestellt ist *content knowldge* das Wissen, das Definitionen, mathematische Konzepte und deren Eigenschaften und Verbindungen dazwischen umfasst. Das *pedagogical content knowledge* umfasst das fachdidaktische Wissen, durch das auch verschiedene Beispiele in unterschiedlichen Schwierigkeitsstufen reflektiert und verwendet werden sollen. Dazu gehören ebenso didaktische Analysen der Beispiele, das Gestalten von Unterrichtsplänen und das Aufdecken typischer Lernschwierigkeiten. Unter *technological content knowledge* wird das Wissen verstanden, das erlaubt, eine Technologie, hier die Software TinkerPlots, zunächst für einen konkreten Inhalt auszuwählen um diesen in Verbindung mit dem Wissen, welche Möglichkeiten und Beschränkungen durch die bestimmte Technologie gegeben sind, zu verwenden. Diese Möglichkeiten und Beschränkungen wurden für die Software TinkerPlots im Kapitel 3 und 4 dieser Arbeit analysiert und die Erkenntnisse sind in das Design der Lehrveranstaltungen eingeflossen, um das *technological content knowledge* der Teilnehmer über TinkerPlots im Sinne einer instrumentellen Genese aufzubauen. Das *technological pedagogical content knowledge* schließlich umfasst das fachdidaktische Wissen zum Softwareeinsatz.

In den Veranstaltungen werden die fachlichen Inhalte vor allem in der Vorlesung besprochen und die didaktischen Komponenten sind vor allem Teil der Seminarsitzungen. Das technologische Wissen wird in der Vorlesung vorbereitet und im Seminar durch viele eigene Aktivitäten der Lernenden zur Verwendung der Software TinkerPlots als Instrument im Sinne der instrumentellen Genese vertieft. Sowohl Vorlesung als auch Seminar haben einen Umfang von zwei Semesterwochenstunden, wobei das Seminar in fünf Gruppen zu je etwa 25-40 Teilnehmern gehalten wird.

Die folgende Tabelle 18 gibt eine Übersicht über die Inhalte der Veranstaltungen im Sommersemester 2017. Im Anschluss wird das Design des neuen Bausteins zur Inferenzstatistik beschrieben. Weitere Informationen zu den Veranstaltungen lassen sich in Podworny, Frischemeier und Biehler (2017) nachlesen.

Tabelle 18: Übersicht über die Inhalte der Veranstaltung

Datum	Baustein	Inhalt
20.04.2017	Datenanalyse	Vorlesung 1 Einführung in Datenanalyse in der Grundschule, Statistische Erhebungen planen, Grundbegriffe der beschreibenden Statistik, erstes Organisieren von Daten, Darstellungen von kategorialen Merkmalen, PPDAC-Cycle
25.04.2017	Datenanalyse	Seminar 1 Einführung in TinkerPlots zur Datenanalyse anhand von eigenständigen Explorationen am selbst erhobenen Datensatz „StudierendeMGDZ_2016“
27.04.2017	Datenanalyse	Vorlesung2 Darstellungen und Verteilungen numerischer Merkmale, Kennzahlen, insbesondere Median und arithmetisches Mittel
02.05.2017	Datenanalyse	Seminar2 Besprechung der Explorationen vom 25.04.2017 mit besonderem Fokus auf dem Erlernen der Bedienung von TinkerPlots
04.05.2017	Datenanalyse	Vorlesung 3 Boxplots, Streumaße und erste Verteilungsvergleiche
09.05.2017	Datenanalyse	Seminar 3 Didaktische Analyse einer TinkerPlots Aufgabe („Katzen“)
11.05.2017	Datenanalyse	Vorlesung 4 Relative Häufigkeiten/natürliche Häufigkeiten im Zusammenhang mit Gesundheitstests, Baumdiagramme, Verteilungsvergleiche numerischer Variable
16.05.2017	Datenanalyse	Seminar 4 Bearbeitung und Diskussion von zwei Artikeln über Datenanalyse in der Grundschule
18.05.2017	Datenanalyse/ Kombinatorik	Vorlesung 5 Abschluss Verteilungsvergleiche und Beginn Kombinatorik
23.05.2017	Datenanalyse	Seminar 5 Didaktische Analyse einer Grundschulbuchaufgabe zur Datenanalyse
25.05.2017	-	Vorlesung 6 - (entfallen wegen Feiertag)
30.05.2017	Datenanalyse	Seminar 6 Unterrichtsprojekt „Weitspringen der Papierfrösche“ aus Lernen-

		densicht durchführen
01.06.2017	-	Vorlesung 7 - (entfallen wegen Unifest)
06.06.2017	-	Seminar 7 Entfallen wegen Pfingstwoche (Lesewoche)
08.06.2017	Kombinatorik	Vorlesung 8 Produktregel der Kombinatorik, Fakultät, Grundmodelle des „Urnenziehens“
13.06.2017	Datenanalyse	Seminar 8 Didaktische Analyse des Unterrichtsprojekts „Weitspringen der Papierfrösche“
15.06.2017	-	Vorlesung 9 - (entfallen wegen Feiertag)
20.06.2017	Kombinatorik	Seminar 9 Didaktische Analyse von Kombinatorikaufgaben in der Grundschule
22.06.2017	Wahrscheinlichkeitsrechnung	Vorlesung 10 Einführung in Wahrscheinlichkeitsrechnung in der Grundschule, Grundbegriffe und –definitionen, Laplacescher und Frequentistischer Wahrscheinlichkeitsbegriff, Simulation von Zufallsexperimenten als informeller Zugang zur Wahrscheinlichkeitsrechnung
27.06.2017	Wahrscheinlichkeitsrechung	Seminar 10 Einführung in die Simulation mit TinkerPlots
29.06.2017	Wahrscheinlichkeitsrechnung	Vorlesung 11 Baumdiagramme als Lösestrategie, Zufallsgrößen, Empirisches Gesetz der großen Zahlen, $1/\sqrt{n}$ – Gesetz, Genauigkeit von Simulationen
04.07.2017	Wahrscheinlichkeitsrechnung	Seminar 11 Didaktische Analyse einer Schulbuchaufgabe zur Wahrscheinlichkeitsrechnung
06.07.2017	Wahrscheinlichkeitsrechnung	Vorlesung 12 Galtonbrett, Binomialkoeffizient, Hinführung zur Binomialverteilung, Eigenschaften der Binomialverteilung
11.07.2017	-	Seminar 12 - (entfallen wegen Tagung)
13.07.2017	-	Vorlesung 13 - (entfallen wegen Tagung)
18.07.2017	Wahrscheinlichkeitsrechnung	Seminar 13 Fachliche Vertiefung von Aufgaben zur Wahrscheinlichkeitsrechnung und deren Einsatz im Unterricht der Grundschule, Analyse einer Grundschulbuchaufgabe
20.07.2017	Inferenzstatistik	Vorlesung 14 Stichprobenschwankungen, Einführung in Denkweise der Inferenzstatistik an einem durchgeführten Randomisierungstestbeispiel, Einführung des Dreischritts (1-Nachdenken über Daten 2-Randomisieren und Wiederholen, 3-Schlussfolgerungen ziehen)
25.07.2017	Inferenzstatistik	Seminar 14 Eigenständiges Durchführen eines Randomisierungstests mit Hands-on Aktivität und anschließender Simulation, Nutzen des „Schemas zur Durchführung eines Randomisierungstests“, Fokus auf den Schritt 2-Randomisieren und Wiederholen
27.07.2017	Inferenzstatistik	Vorlesung 15 Vertiefendes Beispiel und Schematisierung eines Randomisierungstests, Fallstricke dazu, Fokus auf den Voraussetzungen zu einem

	Randomisierungstest und möglichen und nicht möglichen Schlussfolgerungen

Das Seminar wird im Sommersemester 2017 in fünf Gruppen durchgeführt, die alle am gleichen Tag stattfinden. Zwei dieser Seminargruppen werden durch die Autorin dieser Arbeit geleitet, die drei weiteren werden von Daniel Frischemeier gehalten. In den Seminaren werden zum einen Hausaufgaben besprochen, welche die fachlichen Inhalte der Vorlesung vertiefen, zum anderen werden sogenannte Präsenzaufgaben durchgeführt, in denen sich die Studierenden meistens didaktisch mit den jeweiligen Inhalten auseinandersetzen (vgl. Frischemeier & Podworny 2017 und Podworny et al. 2017). Hier werden Artikel aus Grundschulzeitschriften zur Stochastik diskutiert, typische Schulbuchaufgaben didaktisch analysiert oder statistische Projekte durchgeführt und didaktisch reflektiert. Die Präsenzaufgaben werden in verschiedenen, kooperativen Lernformen für das Arbeiten in Kleingruppen durchgeführt, so zum Beispiel Gruppenpuzzle oder Arbeiten nach dem Ich-Du-Wir Prinzip (Barzel 2006). An geeigneten Stellen werden Peerfeedback (Boxtel, Linden & Kanselaar 2000) und Expertenfeedback eingebaut. Der Einsatz der Software TinkerPlots geschieht durch verschiedene instrumentelle Orchestrierungen (vgl. Kapitel 2.2.2).

Der größte Teil der Veranstaltungen entfällt auf den Baustein Datenanalyse. Den PPDAC-Cycle nach Wild und Pfannkuch (1999) zu durchlaufen ist eine wichtige Komponente in den ersten Wochen. In der Vorlesung lernen die Studierenden die explorative Datenanalyse kennen und in den Seminaren diese didaktisch zu analysieren und für den späteren Unterricht zu reflektieren. Ein besonderer Fokus liegt auf dem Erstellen und Interpretieren von Darstellungen uni- und bivariater Daten. Dies geschieht von Anfang an sowohl mit als auch ohne Software zum Ausbau des *content knowledge* und des *technological content knowledge*. Das *pedagogical content knowledge* sowie das *pedagogical technological knowledge* der Lernenden werden in den Seminaren unter Einsatz verschiedener kooperativer Lernmethoden entwickelt. Zum Abschluss dieses Bausteins werden Verteilungsvergleiche ausführlich behandelt. Der Einsatz von TinkerPlots zur explorativen Datenanalyse, aber auch zur Reflektion über Einsatzmöglichkeiten in der Grundschule zieht sich kontinuierlich durch alle einzelnen Sitzungen dieses Bausteins durch verschiedene Orchestrierungen. Nach diesem Baustein sind die Studierenden in der Lage, uni- und bivariate Daten mit und ohne Software zu visualisieren, Lage- und Streumaße zu berechnen und Gruppenvergleiche sowohl kategorialer als auch numerischer Merkmale durchzuführen.

Der zweite Baustein zur Kombinatorik ist vergleichsweise kurz. Die Produktregel der Kombinatorik und vier kombinatorische Figuren, die auf dem Ziehen aus einer Urne basieren (Ziehen mit oder ohne Zurücklegen und mit oder ohne Beachtung der Reihenfolge) werden vorgestellt. Es werden typische Themen und Aufgaben mit Einsatzmöglichkeit in der Grundschule thematisiert und wiederum in den Seminaren didaktisch

analysiert. Lernen auf verschiedenen Darstellungsebenen nach dem E-I-S Prinzip nach Bruner (1971) wird hier ebenfalls thematisiert, händisch durchgeführt und didaktisch anhand von Fachartikeln reflektiert. Der Einsatz von TinkerPlots spielt während dieses kurzen Bausteins keine Rolle.

Im dritten Baustein wird eine Einführung in den Wahrscheinlichkeitsbegriff gegeben. Grundbegriffe für den Anfangsunterricht (sicher, möglich, unmöglich) und grundlegende Definitionen (Ergebnismenge, Ereignis, Wahrscheinlichkeit, etc.) werden gegeben und der Laplacesche und der frequentistische Wahrscheinlichkeitsbegriff werden vermittelt. In diesem Baustein werden die händische und die computerbasierte Simulation als Alternative zu rechnerischen Lösungen eingeführt. Von Beginn an spielt der Einsatz von TinkerPlots eine wichtige Rolle und die Studierenden erhalten viel Gelegenheit, selbst Simulationen in TinkerPlots zu erstellen, um die instrumentelle Genese ausführlich herbeizuführen. Hieran anknüpfend wird das Gesetz der großen Zahlen thematisiert. Dies geschieht stochastische Annäherung der relativen Häufigkeiten für ein Ereignis A an die tatsächliche Wahrscheinlichkeit P(A) für dieses Ereignis bei wachsender Wiederholungszahl. Ansätze, wie dies in den Grundschulunterricht integriert werden kann, werden diskutiert. Als Abschätzungsmöglichkeit für die Genauigkeit von Simulationen wird das $1/\sqrt{n}$ - Gesetz eingeführt (Biehler & Prömmel 2013, Riemer 1991). Zwei Anwendungen werden hierzu thematisiert. Einerseits wird das $1/\sqrt{n}$ – Gesetz als 95%-Prognoseintervall verwendet, um das Intervall vorherzusagen, in welchem mit 95% Sicherheit die relative Häufigkeit für ein Ereignis fällt, wenn die Wahrscheinlichkeit P für dieses Ereignis bekannt ist. Andererseits wird das $1/\sqrt{n}$ – Gesetz als Schätzung einer Wahrscheinlichkeit in Abhängigkeit der Wiederholungsanzahl n verwendet, um eine Schätzung über die unbekannte Wahrscheinlichkeit P für ein Ereignis zu geben, wenn eine relative Häufigkeit – zum Beispiel per Simulation – für dieses Ereignis beobachtet wurde. Am Ende dieses Bausteins wird die Binomialverteilung über das Galtonbrett eingeführt und Eigenschaften der Binomialverteilung werden thematisiert. Im Zentrum steht am Ende vor allem das *technological content knowledge* und das *pedagogical content knowledge*. Die Studierenden lernen während dieses Bausteins die Fabrikmetapher kennen (vgl. Kapitel 3.5) und sie lernen, TinkerPlots als Instrument für stochastische Simulationen einzusetzen und didaktisch zu reflektieren, um TinkerPlots beispielsweise als Demonstrationsmedium im späteren Unterricht einzusetzen.

Organisatorisches zu den Veranstaltungen

Alle Vorlesungen werden mittels PowerPoint-Präsentationen gehalten und den Studierenden spätestens am Vortag der Vorlesung über die Onlineplattform Moodle der Universität Paderborn zur Verfügung gestellt. Desgleichen werden TinkerPlots-Lernumgebungen über diese Plattform zur Verfügung gestellt bzw. in den Veranstaltungen erstellte Dateien nachträglich zur Verfügung gestellt.

Jede Woche werden Hausaufgaben gegeben, die eine Auseinandersetzung mit den fachlichen Inhalten der Vorlesung erfordern. Die Besprechung dieser Aufgaben findet in den Seminaren statt.

Für die Prüfungsleistung der Seminare ist das Führen eines Portfolios erforderlich (Stratmann, Preussler & Kerres 2009). In diesem werden erstens alle Hausaufgaben und alle Präsenzaufgaben gesammelt, und zweitens soll hierin regelmäßig der eigene Lernfortschritt reflektiert werden. So ist nach jedem Baustein eine mindestens zweiseitige Reflexion über den eigenen Wissensstand zu erstellen.

Zum Erlangen der Studienleistung für die Vorlesung ist das Bestehen eines zweistündigen Tests am Ende des Semesters nötig, in welchem aus allen Bausteinen der Vorlesung Aufgaben gestellt werden.

8.2 Der Baustein Inferenzstatistik

Der Baustein Inferenzstatistik wird zur Implmentation in die in Abschnitt 8.1 beschriebene Veranstaltungssequenz designt, so dass auf dem dort erworbenen Vorwissen zur Datenanalyse, zur Wahrscheinlichkeitsrechnung und zur aufgebauten instrumentellen Genese für das Instrument TinkerPlots aufgebaut werden kann.

Als Zeitrahmen für die Lerneinheit stehen drei 90 minütige Sitzungen am Ende des Sommersemesters 2017 zur Verfügung. Dazu gehören die Vorlesung am 20.07.2017, die Seminarsitzung am 25.07.2017 und die letzte Vorlesung des Semesters am 27.07.2017. Der Baustein wird im Wesentlichen theoriegeleitet aufgrund von Forderungen, Erkenntnissen und Erfahrungen aus G. Cobb (2007), Rossman et al. (2008) und Pfannkuch et al. (2015) designt, es fließen aber auch Erkenntnisse aus anderen, in Kapitel 6 diskutierten Studien ein. Außerdem werden eigene Erfahrungen zum Thema Inferenzstatistik aus zwei Seminaren im Sommersemester 2012 und im Wintersemester 2012/2013 genutzt (Podworny 2013, Podworny & Biehler 2014). In diesem Baustein soll thematisiert werden, wie sich Ergebnisse von Verteilungsvergleichen, wie sie im ersten Baustein der Veranstaltungen durchgeführt werden, auf bestimmte Erklärungen zurückführen lassen. Somit soll an das begonnene Thema Verteilungsvergleiche wie beispielsweise von Watson und Moritz (1999) oder von Makar und Confrey (2002) vorgeschlagen, angeschlossen werden. Vor allem die mit zweimal sechzig Minuten wesentlich kürzere Einheit in Pfannkuch et al. (2015) inspirierte zur Realisierung einer Einführung in die beurteilende Statistik als noch fehlendem Teilgebiet der Stochastik in einem recht kurzen Baustein am Ende der beiden Veranstaltungen.

8.2.1 Ziele des Bausteins „Inferenzstatistik“

Das Hauptziel dieses Bausteins ist es, Studierenden des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung, die bereits Erfahrung mit explorativer

Datenanalyse und stochastischer Simulation von Zufallsexperimenten mit TinkerPlots in den Veranstaltungen erworben haben, einen ersten Berührungspunkt mit inferentiellem Schlussfolgern zu bieten und ihnen so ein umfassenderes Bild der Stochastik als in den vergangenen Durchführungen der Veranstaltungen zu bieten. Es sollen ausschließlich Randomisierungstests als eine Methode der Inferenzstatistik thematisiert werden, um den Studierenden Gelegenheit zu geben, einen Test exemplarisch zu verstehen und nicht durch viele verschiedene verwirrt zu werden. Die Studierenden sollen am Ende des Bausteins in der Lage sein, den Argumentationsprozess mit den in Kapitel 6 identifizierten Konzepten beim Durchführen eines Randomisierungstests nachzuvollziehen und diesen selbst zu durchlaufen. Das Thema Unsicherheit soll dabei immer wieder aufgegriffen werden (im Sinn von Pfannkuch et al. 2015) und verdeutlicht werden, dass keine „sicheren“ Schlüsse möglich sind. Die Teilnehmer sollen in der Lage sein, ein Studiendesign zu hinterfragen, Unterschiede in beobachteten Daten zu identifizieren, Null- und Forschungshypothese aufzustellen, eine Referenzverteilung zu erzeugen und einen P-Wert zu ermitteln und daraus adäquate Schlussfolgerungen zu ziehen. Dazu sollen die Teilnehmer zur Benutzung von TinkerPlots zur Durchführung eines Randomisierungstests per stochastischer Simulation instrumentell angeleitet werden. In diesem letzten Baustein der Lehrveranstaltung findet eine Konzentrierung auf *content knowledge*, *technological knowledge* und auf *technological content knowledge* statt. Das p*edagogical knowledge* soll in Form von typischen Fehlvorstellungen und Schwierigkeiten in der letzten Vorlesung aufgegriffen werden, im Vergleich zu den vorangegangenen Bausteinen jedoch etwas in den Hintergrund treten. Dies zeigt die Darstellung in Abbildung 165, in der die Wissensbereiche grau markiert sind, die in diesem Baustein hauptsächlich adressiert werden sollen.

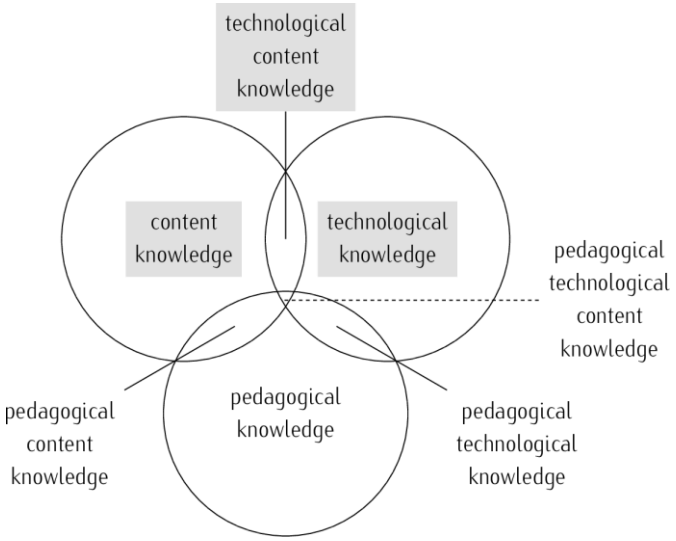


Abbildung 165: Wissensbereiche, die im Baustein Inferenzstatistik hauptsächlich vermittelt werden (eigene Darstellung, angelehnt an Mishra & Koehler 2006, 1025)

8.2.2 Die Sitzungen im Detail

Die zur Verfügung stehenden drei Sitzungen sollen so aufgeteilt werden, dass in der ersten Vorlesung ein einführendes Beispiel vorgestellt wird, um die Denkweise der beurteilenden Statistik („the core logic of inference“ nach G. Cobb 2007) am Beispiel eines Randomisierungstests mit einem signifikanten P-Wert (gemäß Forderung von Holcomb, Chance, Rossman, Tietjen, et al. 2010) zu demonstrieren. In der zweiten Sitzung, dem Seminar, sollen die Teilnehmer durch verschiedene instrumentelle Orchestrierungen (Kapitel 2.2.2) an einem von ihnen selbst durchgeführten Experiment den Prozess des Randomisierens händisch mit Zetteln erleben (nach Gould et al. 2010; Budgett et al. 2012; Tintle et al. 2014) und anschließend eine Simulation in TinkerPlots dazu instrumentell angeleitet durchführen, um für beobachtete Unterschiede zwischen zwei Gruppen einen kausalen Schluss zu ziehen. In der dritten Sitzung, der letzten Vorlesung, sollen die einzelnen Konzepte (Kapitel 6.4) des Randomisierungstests noch einmal detailliert vorgestellt und auf mögliche Schwierigkeiten und Fallstricke hingewiesen werden.

Diese drei Sitzungen werden nun vorgestellt, wie sie für die Implementation in den Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen für Modellieren, Größen, Daten und Zufall“ im Sommersemester 2017 als hypothetische Lerntrajektorie designt wurden.

8.2.2.1 Vorlesung am 20.07.2017

In dieser Einführungsvorlesung sollen die Studierenden als Vorbereitung zunächst das Konzept von Stichprobenschwankungen kennen lernen und anschließend die Logik der Inferenzstatistik an einem ersten Beispiel eines Randomisierungstests. Die Vorlesung wird hauptsächlich per Frontalunterricht (mittels der Orchestrierungen *explain the screen* und *link screen board*) durchgeführt, in welchen aktivierende Phasen (orchestriert durch *discuss the screen*) integriert werden, um die Lernenden kognitiv zu beteiligen. Der Frontalunterricht wird mit Hilfe einer PowerPointpräsentation durchgeführt, für die aktivierenden Phasen werden TinkerPlots, die Tafel und ein Overheadprojektor eingesetzt. Über die Plattform Moodle können die Studierenden die Präsentation im Vorfeld der Vorlesung herunterladen.

8.2.2.1.1 Lernziele der Vorlesung am 20.07.2017

Als Lernziele zu dieser Sitzung lassen sich die folgenden festhalten.

Die Teilnehmer …

- lernen Fragestellungen kennen, die sich mit Hilfe von Randomisierungstests untersuchen lassen,
- lernen ein komplettes Beispiel der beurteilenden Statistik anhand des Verfahrens des Randomisierungstests kennen,

- lernen, dass die zufällige Zuordnung von Versuchspersonen zu Gruppen ein wichtiges Konzept ist,
- lernen den Prozess des Randomisierens anhand der Symbolisierung durch Neuzuordnung von Datenkarten auf zwei neue Gruppen kennen (vertiefende Aktivität dazu im Seminar),
- lernen die Software TinkerPlots als nützliches Instrument kennen, um Randomisierungstests durchzuführen,
- lernen mögliche und nicht mögliche Schlussfolgerungen aus einem Randomisierungstest kennen; in Bezug auf Ursache (kausaler Schluss) und Verallgemeinerung (Schluss auf Population).

8.2.2.1.2 Geplanter Ablauf der Vorlesung am 20.07.2017

Zu Beginn (nach der Begrüßung und Einordnung des Themas Inferenzstatistik in den Vorlesungsverlauf) werden als Motivation mögliche Fragestellungen vorgestellt, auf die durch die Inferenzstatistik eine Antwort gegeben werden können soll, wie zum Beispiel, ob Gähnen ansteckend ist oder ob Schlafentzug zu schlechterer Leistung führt. Als Grundlage, wie zum Beispiel von Wild et al. (2011) ausgeführt, wird die Idee von Stichprobenschwankungen vorgestellt, um die Idee von „first building an appreciation of sample versus population, and description *versus* inference“ (Wild et al. 2011, 252) vorzustellen. Dies geschieht anhand eines Datensatzes, den die Studierenden zu Beginn des Semesters über sich selbst erstellt haben. Aus den vorhandenen Daten wird zum Merkmal Körpergröße das arithmetische Mittel berechnet und anschließend Stichproben im Umfang von n=30 (dies entspricht etwa einer Seminargruppe) gezogen und der jeweilige Stichprobenmittelwert des Merkmals Körpergröße berechnet. Verschiedene solche Stichproben werden gezogen und jeweils das arithmetische Mittel der Körpergröße berechnet, um zu demonstrieren, dass Schwankungen einer Statistik (hier des arithmetischen Mittels) in verschiedenen Stichproben aus einer Grundgesamtheit auftreten können. Die Begriffe Grundgesamtheit und Stichprobe werden hieran (erneut) thematisiert.

Um das Lernziel zu erreichen, den Studierenden ein komplettes Beispiel der beurteilenden Statistik vorzustellen, wird ein Beispiel aus Holcomb, Chance, Rossman, Tietjen, et al. (2010) adaptiert, das auf realen Daten basiert und nah an der Erlebniswelt der Studierenden sein soll. Es dient als roter Faden durch die Vorlesung hindurch. In diesem Beispiel ist die zufällige Zuordnung der Studienteilnehmer zu den Gruppen explizit beschrieben. Der Text aus Abbildung 166 wird vorgestellt.

Das Lernziel, dass die zufällige Zuordnung von Versuchspersonen zu Gruppen eine wichtige Voraussetzung ist, wird direkt mit der Aufgabe verknüpft und das Experimentdesign näher besprochen. Es wird erläutert, wie eine zufällige Zuordnung von Versuchspersonen auf zwei Gruppen geschehen kann und die Notwendigkeit dieses zufälli-

gen Einteilens wird herausgestellt. Als erste Möglichkeit wird vorgestellt, wie durch einen Münzwurf Personen zufällig einer von zwei Gruppen zugeordnet werden können, mit der Problematik, dass sich ggf. unterschiedlich große Gruppen ergeben. Als Alternative dazu wird genannt, dass ein Kartenspiel mit gleichvielen roten und schwarzen Karten in Höhe der jeweiligen Gruppenanzahl verwendet werden kann, um die Versuchspersonen zufällig einer von zwei Gruppen zuzuordnen. Auf die zufällige Zuordnung wird im weiteren Verlauf immer wieder Bezug genommen, um die Wichtigkeit dieser Voraussetzung herauszustellen.

Forscher haben festgestellt, dass Schlafentzug eine schädliche Wirkung auf das visuelle Lernen hat. Aber hält dieser Effekt mehrere Tage an, oder kann eine Person den Schlaf „nachholen“, indem sie in den folgenden Nächten ausreichend schläft?

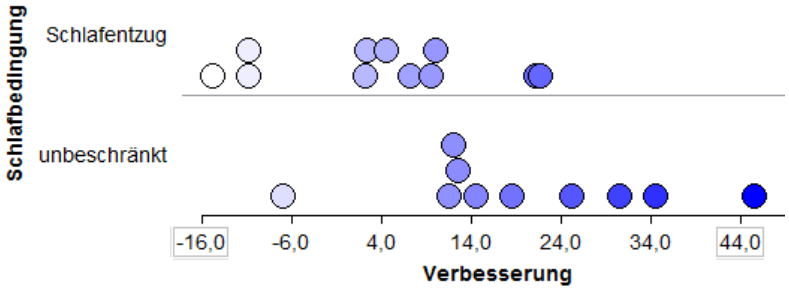
Eine neuere Studie (Stickgold, James & Hobson 2000) untersuchte diese Frage durch zufällige Einteilung von 20 Probanden (Freiwillige im Alter zwischen 18 und 25 Jahren) zu einer von zwei Gruppen: Am ersten Tag wurde eine Aufgabe zum visuellen Urteilsvermögen durchgeführt. Eine Gruppe wurde in der Nacht nach dem Training zeitweise wachgehalten, und der anderen Gruppe wurde erlaubt, uneingeschränkt zu schlafen in der ersten Nacht. Beiden Gruppen wurde danach so viel Schlaf erlaubt, wie sie es in den folgenden zwei Nächten wollten. Alle Teilnehmer wurden am dritten Tag erneut getestet.

Die Leistung der Probanden im Test wurde gemessen als die minimale Zeit (in Millisekunden) zwischen auf einem Computerbildschirm auftretenden Reizen, für die sie genau berichten konnten, was sie auf dem Bildschirm gesehen hatten.

In der geordneten Tabelle und im Punktdiagramm finden sich die Verbesserungen der Zeiten zwischen dem Vor- und dem Nachtest (ein negativer Wert zeigt eine Abnahme der Leistung) in Bezug auf die Teilnehmer in beiden Gruppen:

Schlafentzug (n=11): -14,7; -10,7; -10,7; 2,2; 2,4; 4,5; 7,2, 9,6; 10,0; 21,3; 21,8

Unbeschränkter Schlaf (n=10): -7,0; 11,6; 12,1; 12,6; 14,5; 18,6; 25,2; 30,5; 34,5; 45,6



Frage: Sieht es so aus, als ob Teilnehmer, die in der ersten Nacht unbeschränkt schlafen durften dazu tendieren, eine größere Verbesserung in der Leistung zu haben, als Teilnehmer, denen Schlaf entzogen wurde in der ersten Nacht?

Abbildung 166: Einführungsbeispiel für den 20.07.2017

An dem Beispiel werden nun die einzelnen Konzepte der Argumentation eines Randomisierungstests (vgl. Kapitel 6.4) demonstriert. Zunächst werden die vorliegenden Daten in TinkerPlots gezeigt und Unterschiede zwischen beiden Gruppen herausgearbeitet wie in Abbildung 167.

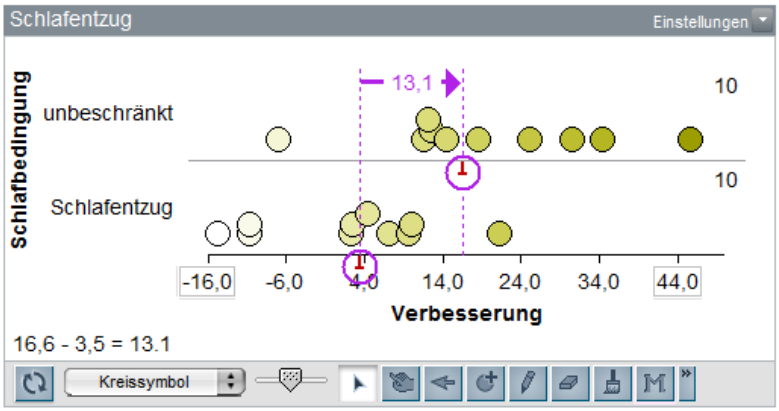


Abbildung 167: Unterschied zwischen den beiden Gruppen des Einführungsbeispiels

Der Unterschied der Mediane wird notiert als $\tilde{x}_{unbeschränkt} - \tilde{x}_{Schlafentzug} = 13{,}1$. Im Folgenden wird das Ziel der Forscher rekonstruiert als *Das Ziel des Experiments war zu entscheiden, ob die Beschränkung des Schlafs eine Verringerung der durchschnittlichen Leistung bewirkt* und der Frage *Ist es möglich, dass es keinen negativen Effekt von Schlafentzug gibt und dass nur die zufällige Zuordnung der Teilnehmer zu den Gruppen die beobachteten Unterschiede zwischen den beiden Gruppen verursacht hat?* gegenübergestellt. Um die Begrifflichkeiten eines Randomisierungstests einzuführen, wird das Ziel der Forscher in die Frage überführt, ob es möglich ist, *dass es keinen negativen Effekt von Schlafentzug gibt und dass nur die zufällige Zuordnung der Teilnehmer zu den Gruppen die beobachteten Unterschiede zwischen den beiden Gruppen verursacht hat?* und dies als Nullmodell bezeichnet. Dazu wird erläutert, dass dieses soeben eingeführte Nullmodell beispielhaft für die These steht, dass es keinen (Null) Unterschied zwischen der Behandlung der beiden Gruppen gibt und dass beobachtete Unterschiede am Zufall, also an der zufälligen Gruppeneinteilung, liegen. Dabei wird immer wieder die Formulierung „zufällige Einteilung der Versuchspersonen zu den Gruppen“ oder Variationen davon verwendet, um wie bei Budgett et al. (2013) angeregt von Anfang an den Fokus auf dieses Szenario zu legen und eine schwerer greifbare „es liegt am Zufall“ Formulierung zu vermeiden.

In einem weiteren Schritt sollen die Fachbegriffe Nullhypothese und Alternativhypothese eingeführt werden. Um nun an den bisher verwendeten Sprachgebrauch anzuknüpfen, wird entschieden, die Alternativhypothese als „Forschungshypothese“ zu bezeichnen, um deutlich herauszustellen, dass diese mit dem zuvor formulierten Ziel der Forscher verknüpft ist. Der Begriff Nullhypothese wird mit

dem zuvor besprochenen Nullmodell in Zusammenhang gebracht und herausgestellt, dass diese zwei Modelle (Nullmodell und Ziel der Forscher) nun in Hypothesen übertragen werden. Hier wird bewusst von der Norm abgewichen, die Alternative zur Nullhypothese mit Alternativhypothese zu bezeichnen, um zu verdeutlichen, dass die Forschungshypothese mit dem Ziel der Forscher in Verbindung steht und so einer möglichen Verwechslung der beiden Hypothesen (wie beispielsweise von Vallecillos 1999 berichtet) entgegenzuwirken. Die dem Ziel der Forscher gegenübergestellte Frage (s. o.) wird als Nullmodell bezeichnet und so eine sprachliche Verbindung zur Nullhypothese aufgebaut.

Als Forschungshypothese wird schriftlich *Die Beschränkung des Schlafs bewirkt eine Verringerung der Leistung* notiert. Die Formulierung der Nullhypothese lautet *Schlafentzug hat keinen negativen Effekt, sondern die zufällige Zuordnung der Teilnehmer zu den Gruppen führt zu dem beobachteten Ergebnis*.

Da davon ausgegangen wird, dass der Begriff „Randomisierung“ nicht allen Studierenden geläufig ist, wird von Anfang an eine sprachliche Alternative mit „zufällige Zuordnung“ verwendet. Um den Begriff zu festigen, aber auch um das Konzept als fest verbunden mit der Nullhypothese zu verdeutlichen, wird diese Versprachlichung in die Nullhypothese aufgenommen. Es wird der Vorschlag von Pfannkuch et al. (2015) aufgegriffen, direkt in die Nullhypothese die Formulierung der zufälligen Zuordnung aufzunehmen: *Kann die zufällige Zuordnung der Teilnehmer zu den beiden Gruppen die Ursache für den beobachteten Unterschied sein?*. Zusätzlich zum Begriff der Nullhypothese wird hier *Frage, durch die das Nullmodell ausgedrückt wird* verwendet. Da dies die Grundlage für die Modellannahme beim Randomisierungstest ist, soll dies auch in der Bezeichnung zum Ausdruck kommen, um potentiellen Verwirrungen der Lernenden bezüglich der zu simulierenden Hypothese entgegen zu wirken.

Dadurch, dass die Forschungshypothese mit dem Ziel der Forscher verknüpft wird, wird als Ziel des durchzuführenden Tests erklärt, dass die Nullhypothese zugunsten der Forschungshypothese abgelehnt werden soll. Dieses Ziel wird sowohl schriftlich auf den Folien der Präsentation als auch mündlich deutlich herausgestellt, um den nun folgenden Randomisierungstest zu motivieren.

Für das Lernziel „Die Teilnehmer lernen den Prozess des Randomisierens anhand der Symbolisierung durch Neuzuordnung von Datenkarten auf zwei neue Gruppen kennen“, wird dieser Prozess zunächst real anhand einer händischen Aktivität durchgeführt. Vorbereitete Zettel als Datenkarten mit den Daten des Schlafentzugexperiments wie in Abbildung 168 werden gezeigt, erklärt und gemischt. An der Tafel werden zwei neue, fiktive Gruppen „neue Gruppe 1“ und „neue Gruppe 2“ erstellt. Immer ein Zettel wird zufällig gezogen und die ersten zehn so gezogenen Zettel der neuen Gruppe 1 zugeordnet und die zweiten zehn Zettel der neuen Gruppe 2. Als Terminologie hierfür

wird „neue Gruppe 1“ und „neue Gruppe 2“ gewählt, um zu verdeutlichen, dass diese neuen Gruppen unabhängig davon sind, ob den Personen vorher der Schlaf beschränkt wurde oder nicht (vgl. Rossman et al. 2008). Nach der erfolgten händischen Neuzuordnung der Werte auf diese zwei neuen Gruppen werden die Mediane dieser beiden neuen Gruppen bestimmt und deren Unterschied notiert und als Datenpunkt in eine Verteilung an der Tafel eingetragen.

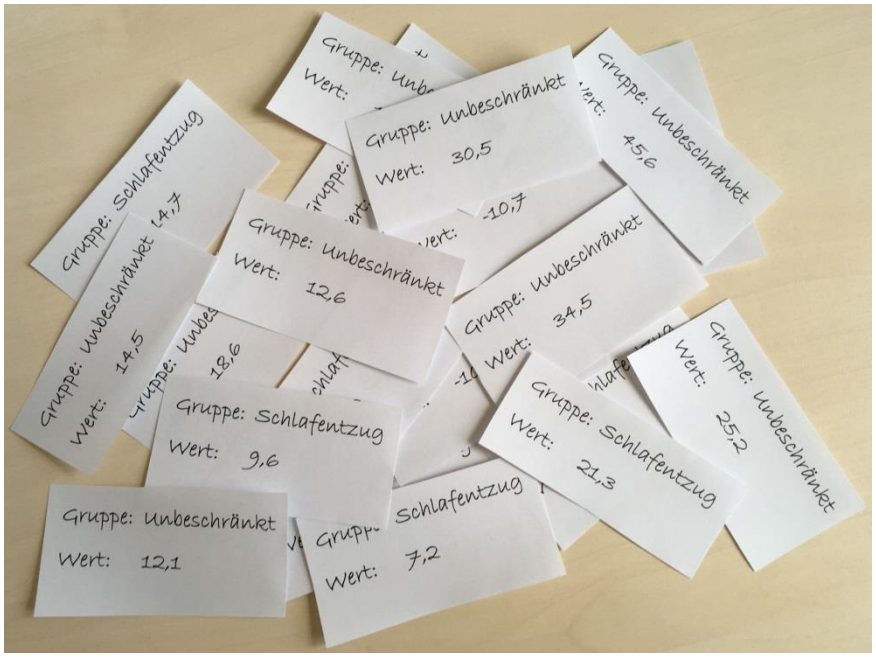


Abbildung 168: Vorbereitete Zettel für die Aktivität des händischen Randomisierens in der Vorlesung am 20.07.2017

Um zu erfahren, welche Differenz zwischen den Medianen eine weitere Randomisierung der Werte ergibt, wird dieser Prozess ein zweites und ein drittes Mal komplett durchgeführt und die erhaltene Differenz in die Verteilung eingetragen. Mit der Begründung, dass eine solche händische Randomisierung der Werte sehr zeitaufwändig sei, wenn ein großer Stichprobenumfang angestrebt wird, wird daraufhin in die Software TinkerPlots gewechselt. Hiermit wird das Lernziel „Die Studierenden lernen Software als nützliches Instrument kennen, um Randomisierungstest häufig durchzuführen“ angestrebt. Die Daten werden in eine Zufallsmaschine kopiert und eine Randomisierung wird durchgeführt, für die der Unterschied der Mediane der beiden neuen Gruppen ermittelt wird (Abbildung 169).

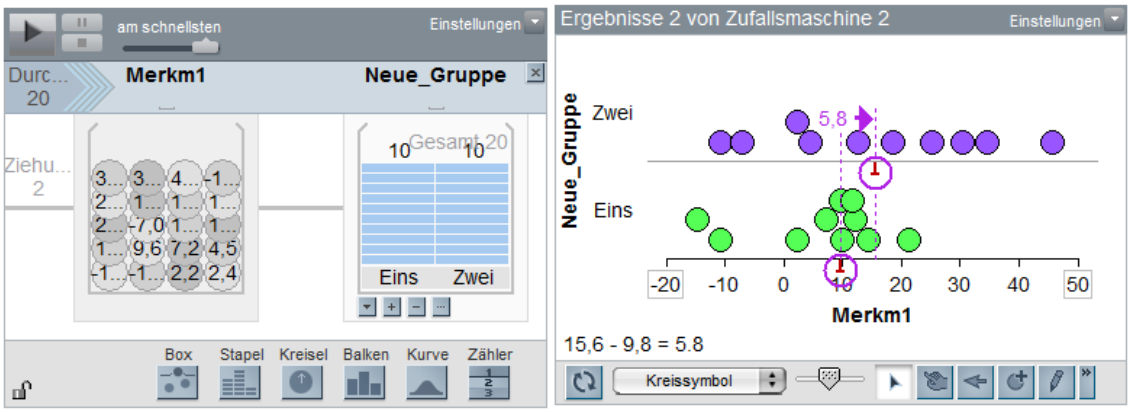


Abbildung 169: Zufallsmaschine zum Einführungsbeispiel und ermittelter Unterschied der Mediane der beiden neuen Gruppen

Alle notwendigen Schritte werden langsam in TinkerPlots vorgeführt und immer wieder mit den zuvor gezeigten händischen Aktionen verbal verknüpft. Um darzustellen, wie sich die Differenzen unter der Annahme der zufälligen Zuordnung der beobachteten Werte zu zwei neuen Gruppen verteilen, wird der Unterschied der Mediane als Messgröße definiert und eine Referenzverteilung aufgebaut. Zu der Referenzverteilung wird als Ziel genannt, den zuvor beobachteten Wert mit Hilfe der Referenzverteilung einschätzen bzw. beurteilen zu können.

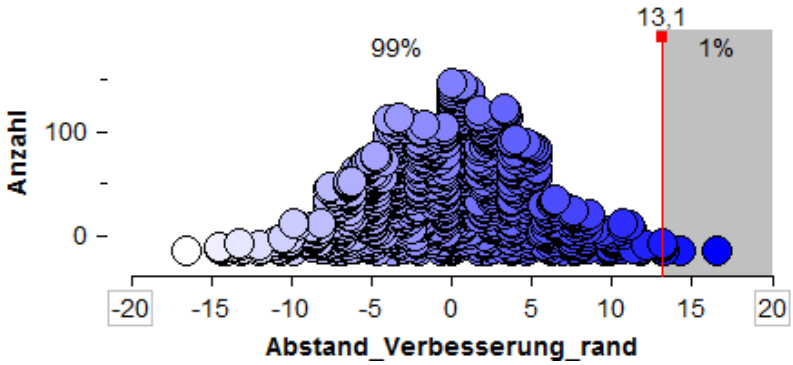


Abbildung 170: Referenzverteilung mit eingezeichnetem P-Wert für die Vorlesung am 20.07.2017

Zunächst wird zu der Referenzverteilung die Frage an die Teilnehmer gestellt, was ein solcher Datenpunkt der Referenzverteilung darstellt. Dies soll in zwei Minuten von den Teilnehmern mit ihren Sitznachbarn diskutiert und anschließend kurz im Plenum besprochen werden. Daraufhin wird in die Referenzverteilung der in dem ursprünglich durchgeführten Experiment beobachtete Unterschied der Mediane mit einer Hilfslinie eingetragen (Abbildung 170) und erläutert, dass allen Datenpunkten der Referenzverteilung die Nullhypothese und somit die zufällige Zuteilung von Versuchspersonen zu den Gruppen zugrunde liegt, um diesen wichtigen Zusammenhang noch einmal deutlich herauszustellen. Es wird festgehalten, dass der beobachtete Wert

von 13,1 weit am Rand der Verteilung liegt. Das Konzept des P-Werts wird mit der Frage *Wie wahrscheinlich ist ein solches Ergebnis wie das beobachtete, unter der Annahme, dass die zufällige Zuordnung der Teilnehmer zu den zwei Gruppen die Ursache ist (und nicht die Beschränkung des Schlafs)?* motiviert. Daraufhin wird mit Hilfe des Einteilers in TinkerPlots der Anteil der Fälle bestimmt, die genauso groß wie oder sogar noch extremer als der beobachtete Wert sind. Zudem wird beschrieben, dass *der P-Wert ‚misst‘, wie extrem ein beobachtetes (d. h. in einem Experiment gefundenes) Ergebnis ist* dann formalisiert als *Wahrscheinlichkeit, ein Ergebnis wie das beobachtete oder ein noch extremeres zu erhalten, wenn die Nullhypothese gilt, heißt P-Wert* mit weiteren sprachlichen Umschreibungen. Es soll deutlich gemacht werden, dass sich der P-Wert als Zahl exakt berechnen lässt, dass dies allerdings eine sehr große Anzahl an Permutationen erfordert und deshalb von uns die Simulation verwendet wird. Aus der Simulation lässt sich (wie in den vorherigen Vorlesungen zum Wahrscheinlichkeitsbegriff bereits thematisiert) jedoch nur eine Schätzung für eine Wahrscheinlichkeit, in diesem Fall für den P-Wert, angeben. Eine Sequenz von möglichen Schlussfolgerungen aus diesem geschätzten P-Wert von 1% ähnlich zu Rossman et al. (2008) wird vorgestellt wie folgt:

- Wenn wirklich keine Unterschiede zwischen den beiden Gruppen wären (wie im Nullmodell angenommen), dann würde ein Ergebnis wie das des Experiments oder ein noch extremeres in ca. 1% der Fälle auftreten.
- Dies (beobachteter Unterschied von 13,1 zwischen den Medianen) wäre ein recht extremes Ereignis, wenn das Nullmodell richtig wäre (aber möglich!).
- Wenn der P-Wert kleiner als 10% ist, dann spricht man in der Statistik davon, dass der Test ein signifikantes Ergebnis zeigt, was dazu führt, dass die Nullhypothese abgelehnt werden kann.
- Deshalb lehnen wir hier das Nullmodell ab und schließen, dass Schlafentzug zu schlechterer Leistung führt.
- Mit einer Wahrscheinlichkeit von ca. 1% ist die zufällige Zuordnung tatsächlich die wahre Erklärung (die wir dann fälschlicherweise abgelehnt hätten).

Als didaktische Reduktion wird eine scharfe Grenze bei $\alpha \leq 10\%$ eingeführt, ab der ein P-Wert als signifikant anzusehen ist und demzufolge die Nullhypothese verworfen werden kann. Hier wird bewusst ein Hybridansatz der Testtheorie von Fisher und Neyman-Pearson verwendet (vgl. Rossman 2008), um sowohl die „Evidenz gegen die Nullhypothese“ von Fisher mit dem aus der Simulation geschätzten P-Wert quantifizieren zu können, aber dem Wunsch der Lernenden nach einer eindeutigen Entscheidung (vgl. Garfield et al. 2012) über die Nullhypothese nachkommen zu

können. Die verbleibende Restunsicherheit des Fehlers erster Art, bei Ablehnung der Nullhypothese eine falsche Entscheidung zu treffen, wird diskutiert, ohne dies mit dem Terminus „Fehler erster Art“ zu bezeichnen.

Am Ende dieser Vorlesung wird das letzte Lernziel „Die Teilnehmer lernen mögliche und nicht mögliche Schlussfolgerungen aus einem Randomisierungstest kennen; in Bezug auf Ursache (kausaler Schluss) und Verallgemeinerung (Schluss auf Population)“ thematisiert. Mögliche Schlussfolgerungen bezüglich der Ursache und nicht mögliche Schlussfolgerungen bezüglich einer Grundgesamtheit werden als Abschluss des Beispiels diskutiert. Dazu sollen zwei Fragen gestellt werden. Erstens, ob sich nun schlussfolgern lasse, dass der Schlafentzug die Ursache für die geringere Verbesserung der Leistung sei? (kausaler Schluss?) und zweitens, wie weit sich diese Ergebnisse nun verallgemeinern lassen? (Schluss auf Population). Wiederum gibt es Gelegenheit für die Teilnehmer, sich kurz mit ihren Nachbarn auszutauschen. Bezüglich der ersten Frage nach dem kausalen Schluss werden die folgenden Punkte besprochen:

- Ja, weil der Test mit einem P-Wert von 1% ein signifikantes Ergebnis zeigt.
- Es handelt sich bei der Studie um ein randomisiertes Experiment. Die zufällige Zuordnung der Teilnehmer in der Studie zu den Gruppen durch die Forscher sorgt dafür, dass andere mögliche Einflussfaktoren ausgeglichen werden.
- Wenn Losglück bei der Gruppenzuordnung abgelehnt wird, dann bleibt als Erklärung nur, dass Schlafentzug tatsächlich negative Effekte hat.
- Es bleibt eine Unsicherheit von ca. 1% dafür, die falsche Schlussfolgerung getroffen zu haben.

Auch dass sich kein Schluss auf die Population (zweite Frage) ziehen lässt, wird besprochen: *Die Teilnehmer an der Studie waren Freiwillige zwischen 18 und 25 Jahren einer speziellen geographischen Gegend. Es handelt sich nicht um eine Zufallsstichprobe aus der Gesamtpopulation, deshalb sollten die Ergebnisse dieser Studie nicht verallgemeinert werden.* In einer weiteren Folie am Ende der Präsentation werden die Bedeutung der zufällligen Zuordnung der Versuchspersonen zu den Gruppen diskutiert als notwendige Voraussetzung, um den kausalen Schluss ziehen zu können. Das Auftreten systematischer Fehler bei Zuordnung der Versuchspersonen nach bestimmten Kriterien zu den Gruppen wird als Abschluss thematisiert.

8.2.2.2 Seminar am 25.07.2017

Im Mittelpunkt der sich an die erste Vorlesung anschließenden Seminarsitzung soll der Prozess des Randomisierens – sowohl händisch als auch per Computersimulation mit

TinkerPlots stehen. Hier soll vor allem auch die Grundlage der instrumentellen Genese zur Verwendung der Software TinkerPlots für Randomisierungstests gelegt werden.

8.2.2.2.1 Lernziele des Seminars am 25.07.2017

Das, was in der vorherigen Vorlesung thematisiert und demonstriert wurde, soll in dieser Sitzung von den Teilnehmern durchgeführt werden. Als Lernziele lassen sich die folgenden formulieren.

Die Teilnehmer …

- lernen das „Schema zur Durchführung eines Randomisierungstests“ kennen und anwenden,
- führen den Vorgang des zufälligen Zuordnens durch das händische Neuzuordnen der Werte zu zwei neuen Gruppen selbst durch,
- übertragen diesen Vorgang in eine TinkerPlots-Simulation,
- führen eine Simulation zur Randomisierung in TinkerPlots durch,
- ziehen mögliche Schlussfolgerungen.

In der nachfolgenden Vorlesung soll der komplette Prozess des Argumentierens in der beurteilenden Statistik als Prozess in drei Schritten vorgestellt werden (1-Nachdenken über die Daten, 2-Randomisieren und wiederholen, 3-Schlussfolgerungen ziehen), von dem in dieser Seminarsitzung der zweite Schritt *Randomisieren und wiederholen* ausführlich thematisiert wird. Das geschieht in dieser Seminarsitzung an den Daten eines Experiments, die die Studierenden zuvor selbst erhoben haben.

8.2.2.2.2 Geplanter Verlauf des Seminars am 25.07.2017

Als Vorbereitung zu der Seminarsitzung wurde den Studierenden acht Wochen zuvor folgende Hausaufgabe gestellt:

Hausaufgabe: Gedächtnisleistung (24.05.2017)

Suchen Sie sich 20 Testpersonen und werfen Sie eine Münze, um zu entscheiden, welche Sequenz sich eine Versuchsperson merken soll (Kopf=Sequenz1, Zahl=Sequenz2). Geben Sie diese eine der unten stehenden Sequenzen an die Versuchsperson. Diese hat 20 Sekunden Zeit, sich die Buchstabenfolge einzuprägen. Anschließend lassen Sie sich die Sequenz aushändigen und die Versuchsperson soll auf einem weiteren Zettel so viele Buchstaben auflisten, wie sie sich merken konnte (nur die richtige Reihenfolge zählt).
Notieren Sie Punkte: Die Anzahl der richtig genannten Buchstaben vor dem ersten Fehler sind die Punkte (nicht die Buchstabenpäckchen, sondern jeder einzelne Buchstabe zählt).

Suchen Sie sich mindestens 20 Testpersonen, gerne mehr! Notieren Sie die erreichten Punkte mit Bezug zur geübten Sequenz (!). Sie haben für diese Aufgabe bis zum 06.07.17 Zeit.
Übertragen Sie die erreichten Punkte in die TinkerPlots-Vorlage.

Hier abschneiden!

Aufgabe für die Versuchsperson:
Merken Sie sich so viele Buchstaben (in der vorgegebenen Reihenfolge!) wie sie können. Sie haben dazu 20 Sekunden Zeit. Es sind 30 Buchstaben vorgegeben.

Sequenz 1

KM — ADAC — GPS — HNO — HSV — FAQ — DIN — DE — IMT — NASA

Hier abschneiden!

Aufgabe für die Versuchsperson:
Merken Sie sich so viele Buchstaben (in der vorgegebenen Reihenfolge!) wie sie können. Sie haben dazu 20 Sekunden Zeit. Es sind 30 Buchstaben vorgegeben.

Sequenz 2

MK — DCAA — PGS — NOH — VSH — QFA — DNI — ED — TMI — ASNA

Abbildung 171: Vorbereitende Hausaufgabe für die Seminarsitzung am 25.07.2017

Der Ablauf der Sitzung ist geplant wie in Tabelle 19 zu sehen. Als Strukturierung werden Fragen formuliert, die den Studierende auf dem Präsenzübungszettel 9 (Abbildung 172) präsentiert werden. Ein Auftrag an die Studierenden ist es, das „Schema zur Durchführung eines Randomisierungstets“ (Abbildung 174) auszufüllen.

Tabelle 19: Geplanter Ablauf des Seminars am 25.07.2017

Zeit	Phase	Inhalt	Medium	Orchestrierung
0	1	Anknüpfung an die Aufgabe Gedächtnisleistung (Abbildung 171) Erarbeitung der Forschungsfrage zu dem Experiment und Diskussion zur Datenerhebung	Schema Randomisierungstest	Unterrichtsgespräch *Discuss-the-screen*
10	2	Analyse der erhobenen Daten	TinkerPlots	Partnerarbeit *Spot-and-show*
20	3	Formulieren der Null- und Forschungshypothese	Schema Randomisierungstest	Partnerarbeit/ Unterrichtsgespräch
25	4	Demonstration händische Randomisierung und händische Randomisierung der eigenen Daten in Partnerarbeit	Overhead/Tafel Zettel	Dozentenvortrag Partnerarbeit
50	5	Randomisieren per Simulation in TinkerPlots, Demonstration zum Nacharbeiten mit den eigenen Daten	TinkerPlots	*Technical-demo* und *explain-the-screen*, gefolgt von *work-and-walk-by*
70	6	Bestimmen des P-Werts	TinkerPlots	*Technical-demo*

		Schlussfolgerungen ziehen aus kleinem P-Wert (1%) und großem P-Wert (15%)		und d*iscuss-the-screen*
85	7	Übertragen der Ergebnisse in das Schema	Schema	Einzelarbeit

Phase 1: Anknüpfen an das Experiment Gedächtnisleistung, Forschungsfrage, Datenerhebung

Zu Beginn der Sitzung wird den Seminarteilnehmern das Ziel der Sitzung vorgestellt, einen Randomisierungstest selbständig erst händisch und dann per TinkerPlots-Simulation anhand der von ihnen erhobenen Daten im Rahmen der Hausaufgabe zur Gedächtnisleistung durchzuführen. Dazu bekommt jeder Teilnehmer seine eigenen Daten, sprich die von ihm erhobenen Daten zum Experiment der Gedächtnisleistung. Zunächst wird mit den Seminarteilnehmern das Studiendesign diskutiert und die zugehörige Forschungsfrage aufgestellt, wobei auch gefragt wird, wie sie ihre Daten erhoben haben. Es soll gefragt werden, welche Testpersonen (beispielsweise Familienmitglieder, Freunde, Verwandte oder Kommilitonen) ausgewählt wurden. Anschließend soll diskutiert werden, wie die Sequenzen den Testpersonen zugeordnet wurden, um sicher zu stellen, dass diese zufällig zugelost wurden und andere Fälle zu diskutieren. Für die folgenden Demonstrationen in TinkerPlots wird ein von mir selbst erhobener Datensatz verwendet, zu dem im Vorfeld 20 Personen an dem Experiment teilgenommen hatten und zufällig Sequenz 1 oder 2 zugeteilt bekamen.

Nach dieser Einführung wird den Studierenden ein Präsenzübungsblatt (Abbildung 172) ausgehändigt. Es wird deutlich gemacht, dass die Antworten auf die Fragen im gemeinsamen Gespräch erarbeitet werden und als Strukturierung dienen sollen, aber nicht schriftlich beantwortet werden müssen.

Präsenzübung 9

Gedächtnisstudie - Einen Randomisierungstest durchführen

Im Folgenden finden Sie einen „Plan“ für die Durchführung eines Randomisierungstests. Auf der Rückseite ist dieser zusammengefasst in einem Schema. Dies sollen Sie anhand Ihrer eigenen Daten (eigene Umfrage + Simulation) ausfüllen.

a) Welche Frage soll mit der Studie beantwortet werden?

b) Wie habe ich meine Daten erhoben?

c) Welche Unterschiede lassen sich in meinen Daten erkennen zwischen Personen, die sich Sequenz Eins gemerkt haben und denen, die sich Sequenz Zwei gemerkt haben?

d) Wie lautet die Forschungshypothese? Wie lautet die Nullhypothese?

e) Welches Modell liegt der Nullhypothese zugrunde? (Nullmodell)

f) Welcher Wert soll überprüft werden? (=Teststatistik)

g) Welchen Wert nimmt die Teststatistik in meinen eigenen Daten (siehe c) an?

h) Randomisieren: Berechne den Wert der Teststatistik für eine Randomisierung der Daten. (Im Plenum und in Gruppen: von Hand)

i) Referenzverteilung erzeugen: Für jede Randomisierung muss der Wert der Teststatistik berechnet werden. Alle diese Teststatistiken führen zur Referenzverteilung.

j) P-Wert ermitteln: Wie groß ist die Wahrscheinlichkeit, einen Wert wie den beobachteten in meinen Daten (g) oder einen noch extremeren zu erhalten unter der Voraussetzung, dass die Nullhypothese (dass die Zuordnung zu den beiden Sequenzen keine Rolle spielt) gilt?

k) Ablehnen? Kann ich aufgrund der in j) ermittelten Wahrscheinlichkeit die Nullhypothese als glaubhafte Erklärung ablehnen?

l) Welche Schlussfolgerungen kann ich nun ziehen? Gibt es einen kausalen (ursächlichen) Zusammenhang? Gilt dies für alle Personen?

Abbildung 172: Präsenszübungsblatt für die Seminarsitzung am 25.05.2017

Als Hilfestellung befindet sich auf der Rückseite des Zettels eine Zusammenfassung zum P-Wert wie in Abbildung 173.

Zusammenfassung P-Wert

Definition P-Wert

Der P-Wert ist die Wahrscheinlichkeit, einen Wert wie den beobachteten oder einen noch extremeren zu erhalten **unter der Voraussetzung, dass die Nullhypothese gilt.**

Überblick Evidenz

Ein P-Wert …	**zeigt…**
≥ 10%	kein signifikantes Ergebnis. Die Nullhypothese kann nicht abgelehnt werden. Es ist allerdings auch nicht gezeigt, dass sie wahr ist.
≤ 10%	ein signifikantes Ergebnis auf dem Niveau des P-Werts. Die Nullhypothese kann abgelehnt werden und die Forschungshypothese kann angenommen werden. Es bleibt eine Restunsicherheit, die falsche Beurteilung getroffen zu haben.

„Je kleiner der P-Wert ist, desto stärker spricht der experimentelle Befund gegen die Gültigkeit der Nullhypothese.“

Abbildung 173: Hilfestellung P-Wert in der Seminarsitzung am 25.05.2017

Um das Lernziel „Die Teilnehmer lernen das „Schema zur Durchführung eines Randomisierungstests“ kennen und anwenden“ zu erreichen, wird als letzte Maßnahme noch das Schema zur Durchführung eines Randomisierungstests wie in Abbildung 174 ausgeteilt. Die Studierenden werden aufgefordert, dieses Schema bis zum Ende der Sitzung auszufüllen und später im Rahmen ihres Portfolios abzugeben.

Schema für die Durchführung eines Randomisierungstests

0. Welche Frage soll beantwortet werden?	
1. Beobachtete Daten Zufallsstichprobe?/Randomisierung? Welcher Unterschied lässt sich feststellen?	
2. Nullmodell Wie lautet die Forschungshypothese?	

Wie lautet die Nullhypothese?	
3. Simulation Wie wird randomisiert?	
4. Teststatistik Welcher Wert soll überprüft werden? Wie sieht die Referenzverteilung aus?	
5. P-Wert Wie groß ist der P-Wert? Welche Evidenz zeigt der P-Wert?	
6. Schlussfolgerung Erläuterung der Ergebnisse Ablehnen? Kausal? Population?	

Abbildung 174: Ausgeteiltes Schema zur Durchführung eines Randomisierungstests am 25.07.2017

Für die Sitzung wird zwischen Plenumsdiskussionen und Partnerarbeitsphasen abgewechselt und verschiedene instrumentelle Orchestrierungen genutzt.

Phase 2: Analyse der erhobenen Daten

Nach der einführenden Diskussion im Plenum über die Datenerhebung sollen sich die Studierenden in Zweier- oder Dreiergruppen zusammenfinden, um während der Arbeitsphasen gemeinsam an ihren Daten arbeiten zu können. Als nächstes wird Tinker-Plots genutzt, um die erhobenen Daten in Kleingruppen zu analysieren und Unterschiede zwischen Personen der Gruppe, die sich Sequenz 1 gemerkt hat und der Gruppe der Personen, die sich Sequenz 2 gemerkt hat, herauszuarbeiten. Im Demonstrationsdatensatz wird in einer *discuss-the-screen* Orchestrierung der Unterschied zwischen den Medianen der beiden Gruppen herausgearbeitet, wie in Abbildung 175 zu sehen. Im Median werden 7,5 Punkte mehr bei Sequenz 1 erreicht als bei Sequenz 2.

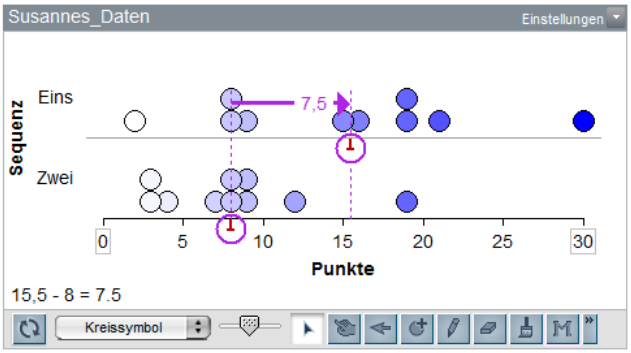


Abbildung 175: Unterschied der Mediane im Demonstrationsdatensatz am 25.05.2017

Phase 3: Nullmodell

Nachdem alle Seminarteilnehmer den Unterschied in ihren Daten zwischen den Medianen in ähnlicher Weise herausfinden und im Schema zur Durchführung eines Randomisierungstests (Abbildung 174) notieren, wird kurz der Frage nachgegangen, welche Erklärungen es für diese beobachteten Unterschiede gibt. Hierzu sollen Inputs durch die Studierenden gemacht werden. Wenn nicht die beiden erwarteten Erklärungen (*Sequenz 1 lässt sich besser merken als Sequenz 2* und *die gemerkte Sequenz spielt keine Rolle, sondern die zufällige Zuordnung der Teilnehmer zu den Sequenzen hat zu den beobachteten Unterschieden geführt*) durch die Teilnehmer genannt werden, sollen sie durch die Dozentin vorgegeben werden. Es soll deutlich herausgestellt werden, dass es bei der zweiten Erklärung keine Rolle spielt, welche Sequenz sich eine Testperson gemerkt hat, da davon ausgegangen wird, dass die erreichten Punkte unabhängig von der zuvor gemerkten Sequenz sind. Hieraus werden die Forschungs- und die Nullhypothese generiert, die die Studierenden in eigenen Worten formulieren und in das Schema eintragen sollen. Zur Ergebnissicherung werden einige dieser Hypothesen im Plenum vorgestellt und ggf. umformuliert. Als mögliche Nullhypothese kann thematisiert werden, dass als Erklärung die zufällige Zuordnung der Testpersonen zu den Gruppen für die beobachteten Unterschiede angenommen wird.

Phase 4: Händische Randomisierung

Hieraus wird die Absicht formuliert zu prüfen, welche Unterschiede auftreten, wenn die Werte zufällig auf zwei neue Gruppen aufgeteilt werden. Mittels kleiner Overheadzettel, auf denen die erreichten Punkte des Demonstrationsdatensatzes stehen, wird eine händische Randomisierung zu zwei neuen Gruppen A und B[42] durchgeführt. Anschließend werden die Werte der neuen Gruppen geordnet, an der Tafel notiert, der jeweilige Median ermittelt und der Unterschied zwischen den Medianen bestimmt. Diesen Prozess sollen die Studierenden anschließend in Partnerarbeit nachvollziehen, indem sie ihre eigenen Werte auf kleine Zettel schreiben und händisch randomisieren und die Unterschiede der Mediane ermitteln und in eine selbstgezeichnete Graphik eintragen. Hier wird der Median statt des arithmetischen Mittels vorgeschlagen, um die Ermittlung des Werts zu erleichtern.

Phase 5: Simulation in TinkerPlots

Mit Bezug darauf, diesen Prozess nun sehr häufig durchführen zu wollen, wird TinkerPlots verwendet, um den Randomisierungstest zu simulieren. In einer langsamen Demonstration, die die Studierenden an ihren Laptops simultan nachvollziehen sollen, wird gezeigt, wie TinkerPlots verwendet werden kann, um den Prozess des händischen

[42] A und B als neue Gruppennamen, weil 1 und 2 schon für die Sequenzen belegt sind.

Simulierens in der Software abzubilden (Orchestierung *technical-demo* zusammen mit *explain-the-screen*).

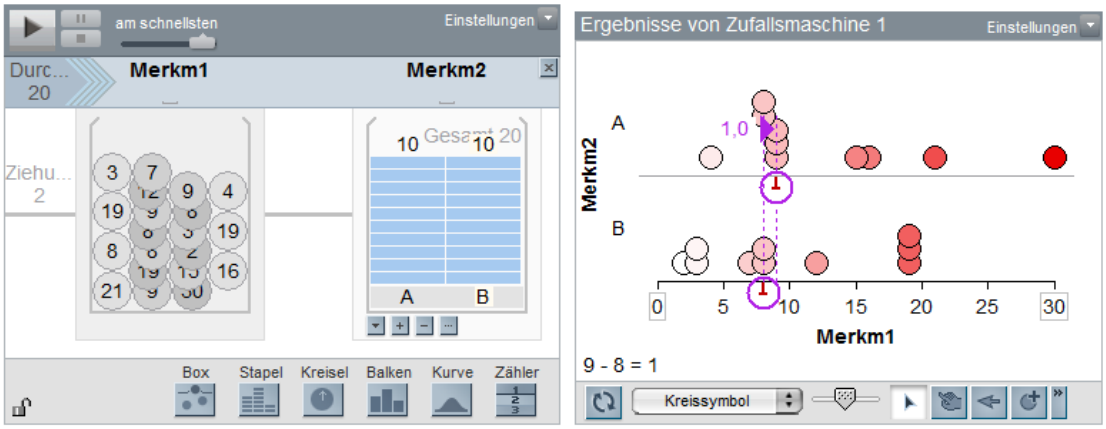


Abbildung 176: Im Seminar erstellte Zufallsmaschine (links) mit dargestellten Ergebnissen und gemessenen Unterschieden der Mediane (rechts)

Dazu wird eine Zufallsmaschine wie in Abbildung 176 erstellt. Die simulierten Daten werden ausgewertet und der Unterschied der Mediane als Messgröße gesammelt und in einer Referenzverteilung (Abbildung 177 rechts) dargestellt. Dabei werden die nötigen Feinheiten der Übertragung in die Zufallsmaschine besprochen, wie dem Ziehen ohne Zurücklegen, aber auch technische Details, wie dem Ausstellen der Animation zur Beschleunigung des Messgrößensammelvorgangs (vgl. Kapitel 3.5.3). Der Bezug zwischen den Aktionen in TinkerPlots und den händisch durchgeführten Aktivitäten wird immer wieder hergestellt.

Im Anschluss haben die Studierenden Gelegenheit, die Simulation in Kleingruppen durchzuführen. Währenddessen können einzelne Fragen zur Simulation in den Gruppen von der Dozentin geklärt werden (*work-and-walk-by* Orchestrierung), bis alle Teilnehmer die fertige Simulation bis zur Referenzverteilung erstellt haben.

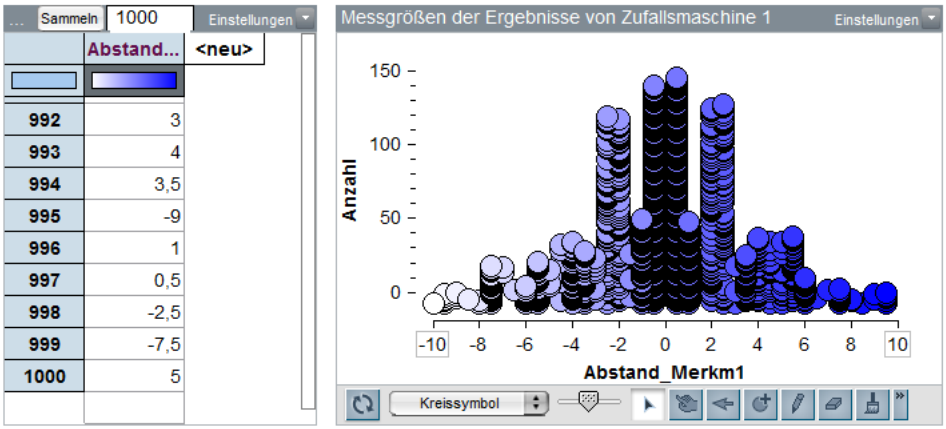


Abbildung 177: Referenzverteilung zur Gedächtnisleistung

Phase 6: P-Wert bestimmen und Schlussfolgerungen ziehen

Hiernach wird der P-Wert für den Demonstrationsdatensatz in TinkerPlots verbalisiert als *geschätzte Wahrscheinlichkeit, einen Wert von 7,5 wie den beobachteten oder einen noch extremeren zu erhalten, vorausgesetzt, dass die Nullhypothese wahr ist* und ermittelt als ca. 1% wie in Abbildung 178.

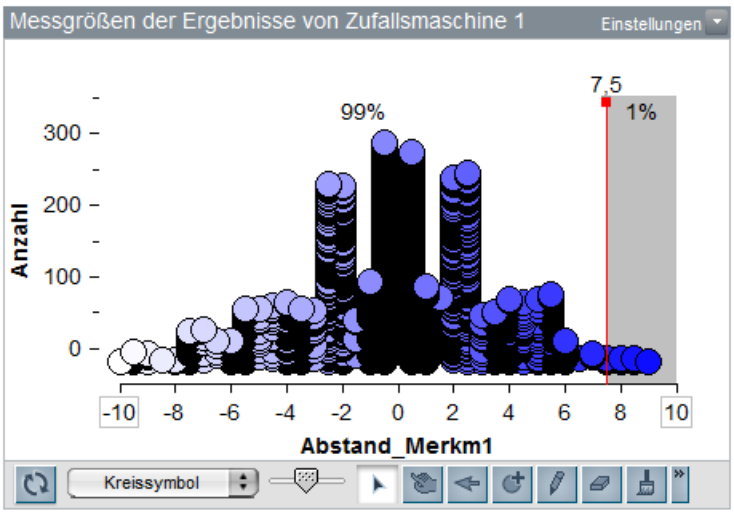


Abbildung 178: Aus der Simulation geschätzter P-Wert

Es wird vorgetragen, dass dieser P-Wert ein signifikantes Ergebnis darstellt, weil er kleiner als 10% ist und somit die Nullhypothese abgelehnt werden kann. Gleichermaßen wird darauf hingewiesen, dass dies eine Entscheidung unter Unsicherheit ist, denn im Beispiel verbleibt eine einprozentige Wahrscheinlichkeit dafür, ein Ergebnis wie das beobachtete oder extremer zu erhalten, wenn die zufällige Zuordnung der Testpersonen zu den Gruppen die richtige Erklärung ist. Ebenfalls wird besprochen, dass somit (mit geringer Restunsicherheit) geschlossen werden kann, dass sich Sequenz 1 besser merken lässt als Sequenz 2, allerdings nur bezogen auf die Testpersonen, die am Experiment teilgenommen haben und keine Verallgemeinerung möglich ist, weil keiner der Anwesenden zuvor eine Zufallsstichprobe von Testpersonen ausgewählt hat. Zum Abschluss können Schlussfolgerungen bezüglich einiger Studierendenergebnisse besprochen werden, die aufgrund abweichender Ausgangsdaten möglicherweise unterschiedlich große P-Werte haben. Im Fall großer P-Werte (>10%) soll diskutiert werden, dass sich in diesem Fall die Nullhypothese nicht ablehnen lässt, sie deshalb aber auch nicht richtig sein muss. Am Ende der Sitzung haben die Studierenden noch Gelegenheit, die einzelnen Schritte im Schema zur Durchführung eines Randomisierungstests zu dokumentieren.

8.2.2.3 Vorlesung am 27.07.2017

In der dritten Veranstaltung des Bausteins sollen die Studierenden ein vertieftes Verständnis zu den einzelnen Konzepten beim Randomisierungstest erhalten. Entlang des Beispiels, das die Studierenden in der vorangegangen Seminarsitzung ausführlich kennen gelernt haben (Gedächtnisleistung), werden Feinheiten und wichtige Aspekte zu den einzelnen Konzepten in der Vorlesung aufgezeigt. In der Seminarsitzung wird ausführlich der Prozess des Randomisierens thematisiert, um hierdurch innerhalb einer Seminarsitzung ein möglichst tiefes Verständnis für den Prozess zu erzeugen. Mit diesem Vorwissen sollen nun die Feinheiten und einige typische Fehlvorstellungen bezüglich des Randomsierungstests erklärt werden. Wie die Vorlesung zuvor wird diese Sitzung im Wesentlichen per Frontalunterricht anhand einer PowerPoint-Präsentation durchgeführt, in die einige aktivierende Phasen eingebaut werden. Die PowerPoint-Präsentation kann im Vorfeld von den Studierenden über die Plattform Moodle heruntergeladen werden.

8.2.2.3.1 Lernziele der Vorlsung am 27.07.2017

Als Lernziele zu dieser Sitzung lassen sich die folgenden festhalten.

Die Teilnehmer …

- lernen die Logik des Randomisierungstestens zusammengefasst als Dreischritt kennen,
- lernen eine Verbindung zwischen der Randomisierung beim Datenerheben und möglichen Schlussfolgerungen zu ziehen (Konzept der zufälligen Zuordnung),
- lernen explizit zwei mögliche Erklärungen für beobachtete Unterschiede kennen. Zum einen den Zufall (hier als die zufällige Zuordnung zu Gruppen) und zum anderen die „Wirkung“ oder den „Effekt“ einer Behandlung/Maßnahme (Konzept der möglichen Erklärungen),
- vertiefen ihr Wissen über den Randomisierungstest als schematisches Verfahren mit den weiteren Konzepten *Forschungsfrage formulieren, Beobachtete Daten analysieren, Nullmodell aufstellen, Simulation durchführen, Teststatistik berechnen, Referenzverteilung erstellen und P-Wert ermitteln* und *Schlussfolgerungen ziehen*,
- lernen verschieden große P-Werte zu interpretieren,
- lernen, dass unabhängig von der Größe eines P-Werts kein sicherer Schluss (für oder gegen die Nullhypothese) möglich ist,
- lernen, welche Schlussfolgerungen möglich sind in Bezug auf Ursache und Population.

8.2.2.3.2 Geplanter Verlauf der Vorlesung am 27.07.2017

Als durchgängiges Beispiel werden die von den Studierenden selbst erhobenen Daten zur Gedächtnisleistung verwendet (Abbildung 171), die bereits in der vorangegangenen Seminarsitzung behandelt wurden. Hierdurch soll gewährleistet werden, dass der Prozess des Randomisierens bereits händisch und durch Simulation durchdrungen ist. Somit kann der Fokus stärker auf die Konzepte und Argumentationen im gesamten Prozess gelegt werden, also gerade die Logik des Schließens stärker in den Vordergrund gerückt werden.

Die Logik des Randomisierungstestens anhand eines Dreischritts zu verwenden wird u. a. von G. Cobb (2007), Tintle et al. (2009), Zieffler und Catalysts for Change (2013) und Pfannkuch et al. (2015) vorgeschlagen. Wie in Kapitel 6.3 herausgearbeitet, soll in dieser Vorlesung der Dreischritt *1-Nachdenken über die Daten, 2-Randomisieren und Wiederholen, 3-Schlussfolgerungen ziehen* als Grundgerüst für die Erreichung des übergeordneten Lernziels (vgl. 8.2.1), die Logik des Randomisierungstestens zu demonstrieren, genutzt werden. Diese drei Schritte werden zu Beginn der Vorlesung mit den Konzepten des Schemas zur Durchführung eines Randomisierungstests (Abbildung 174), welches die Studierenden in der Seminarsitzung kennengelernt haben, verknüpft.

Von Rossman et al. (2008) wird explizit gefordert, eine Verbindung zu ziehen zwischen der anfänglichen Randomisierung der Versuchseinheiten auf die Gruppen in einem Experiment und möglichen Schlussfolgerungen. Bei G. Cobb (2007) findet sich ein ähnlicher Ansatz. Dies ist ein weiteres Lernziel dieser Vorlesung und wird ausführlich thematisiert. Gestartet wird mit der Erklärung des randomisierten Studiendesigns und erläutert, wieso eine zufällige Zuordnung von Versuchseinheiten auf die Gruppen notwendig ist und wie dies mit den späteren Schlussfolgerungen zusammenhängt. Dies wird inhaltlich entlang des Schemas von Ramsey und Shafer (2013) (vgl. Kapitel 5.4) erläutert. In diesem Zug wird vor allem darauf eingegangen, dass Faktoren wie beispielsweise Alter, Vorwissen, etc. möglicherweise ebenso wie die durchgeführte Maßnahme Einfluss auf die Ergebnisse haben können. Um einen kausalen Schluss bezüglich der Maßnahme ziehen zu können, müssen diese Faktoren jedoch kontrolliert werden, was durch zufällige Aufteilung der Versuchseinheiten auf die Gruppen, also durch Randomisierung, weitgehend geschieht. Somit werden weitere Variablen thematisiert und durch Randomisierung als nicht plausible Erklärung ausgeschlossen.

Im Folgenden wird das Lernziel verfolgt, den Randomisierungstest als schematisches Verfahren zu festigen, um wie von Budgett et al. (2013) gefordert den *cognitive load* zu reduzieren. Am Beispiel der Gedächtnisleistung werden die einzelnen Schritte des Schemas besprochen, beginnend mit der Forschungsfrage, zu der explizit der Fehlvorstellung entgegengearbeitet wird, dass diese mit der Nullhypothese in Verbindung steht. Im Vergleich zur vorherigen Vorlesung wird hier verfeinert, dass zwei mögliche Erklä-

rungen in Frage kommen. Diese Erarbeitung soll nicht wie bei Pfannkuch et al. (2015) den Studierenden überlassen werden, sondern dieses Wissen soll transparent dargestellt werden. Als erste mögliche Erklärung wird die zufällige Aufteilung von Versuchseinheiten zu Gruppen genannt, als zweite mögliche Erklärung der Effekt einer durchgeführten Maßnahme, immer mit Bezug zum Beispiel der Gedächtnisleistung. Um eine weitere Anwendung zu thematisieren, wird bei der Analyse der Daten statt des Unterschieds zwischen den Medianen der beiden Gruppen nun der Unterschied zwischen den arithmetischen Mitteln verwendet und auf andere Statistiken, wie zum Beispiel Quartile oder andere Anteile hingewiesen. Es wird demonstriert, dass bei der Beschreibung der Daten ein beobachteter Unterschied herausgegriffen und notiert werden soll, da dieser in Bezug zur Teststatistik gesetzt werden muss.

Das Konzept der Nullhypothese und die damit verbundene Logik des Tests werden vertiefend thematisiert. Zur Logik des Testens wird eine Verbindung zum indirekten Beweis hergestellt, dass das Gegenteil der Forschungshypothese abgelehnt werden soll, um diese annehmen zu können. Gleichzeitig wird deutlich herausgestellt, dass es sich bei einem Randomisierungstest um keinen „Beweis“ handelt. Weiter wird die Nullhypothese als Versprachlichung der Randomisierung erklärt und betont, dass diese für den Test als wahr vorausgesetzt wird. Der Logik des indirekten Beweises folgend soll die Nullhypothese getestet werden, um sie gegebenenfalls zugunsten der Forschungshypothese ablehnen zu können. Die mögliche Verwechslung von Null- und Forschungshypothese (Vallecillos 1999) wird hier ebenfalls thematisiert als potentielle Schwierigkeit. Ebenso wird die Unmöglichkeit thematisiert, die Forschungshypothese direkt testen zu können.

Wie bei Rossman et al. (2008) vorgeschlagen, wird sich dem Aufstellen bzw. Versprachlichen des Nullmodells ausführlich gewidmet. Die Verbindung von Nullhypothese und Nullmodell wird thematisiert und die erneute Randomisierung der Werte beschrieben. Dabei wird das Modell von Rossman et al. (2008) abgewandelt. Um Verwirrung zu vermeiden, werden nach der Randomisierung der Werte diese zwei *neuen Gruppen A und B* zugeordnet. Besprochen wird hier auch, dass in der Theorie alle möglichen Permutationen der Werte zu den Gruppen gebildet werden, wobei die Grundvoraussetzung ist, dass sowohl für die theoretische Berechnung als auch für die Simulation alle möglichen Permutationen gleichwahrscheinlich sind. Das Nullmodell wird hier beschrieben als Situation, in der sich alle Daten als Karten in einer Urne befinden. Aus dieser wird nun zufällig gezogen und der Wert einer von zwei neuen Gruppen A und B zugeordnet, wobei alle möglichen Zuordnungen gleichwahrscheinlich sind.

Beim Übertragen des Nullmodells in die Software TinkerPlots wird Bezug genommen zu den händischen Aktionen, die im Seminar durchgeführt wurden. Gleiches gilt für das Berechnen der Teststatistik der Randomisierungen und dem Erzeugen der Referenzver-

teilung. Die Ermittlung des P-Werts, seine Definition und Bedeutung werden ausführlich thematisiert. Es wird die sprachliche Norm zur Interpretation des P-Werts gesetzt: *Bei einem P-Wert $\alpha \leq 10\%$ liefert der Test ein signifikantes Ergebnis. Die Nullhypothese kann abgelehnt und die Forschungshypothese kann angenommen werden*. Um typischen Fehlvorstellungen entgegenzuwirken, wird dabei auch darauf eingegangen, dass somit die Nullhypothese nicht als wahr bewiesen ist und der P-Wert keine Wahrscheinlichkeit für oder gegen die Nullhypothese oder die Forschungshypothese ist (Krauss & Wassner 2001, Liu & Thompson 2009). Ebenso viel Zeit soll danach auf die möglichen Schlussfolgerungen verwendet werden, die aus einem P-Wert gezogen werden können. Hier wird darauf eingegangen, dass bei einem signifikanten Ergebnis die Nullhypothese abgelehnt werden kann, da man etwas Seltenes beobachtet hätte falls die Nullhypothese richtig wäre. Somit liefert die Forschungshypothese eine plausiblere Erklärung für den beobachteten Unterschied. Weiter wird thematisiert, dass für die Beurteilung, ob ein P-Wert signifikant ist, auch andere Grenzen gängig sind, so beispielsweise $\alpha \leq 5\%$ oder $\alpha \leq 1\%$, je nach Forschungsgebiet, wodurch die Wahrscheinlichkeit verringert würde, die Nullhypothese abzulehnen, obwohl sie in Wahrheit die richtige Erklärung ist. Gleichzeitig wird herausgestellt, dass zwar bei einem kleinen P-Wert die Nullhypothese abgelehnt wird, sie aber trotzdem wahr sein kann. Die Gültigkeit der Forschungshypothese lässt sich nicht beweisen, lediglich kann sie als die plausiblere Erklärung beurteilt werden. Mit Bezug auf den Kontext wird geschlussfolgert, dass aufgrund der anfänglichen zufälligen Zuordnung der Versuchsteilnehmer zu den Gruppen ein kausaler Schluss möglich ist, dass also geschlossen werden kann, dass sich die erste Sequenz an Buchstaben besser merken lässt als die zweite Sequenz. Aufgrund der nichtzufälligen Auswahl der Versuchspersonen ist jedoch kein Schluss auf die Allgemeinheit möglich und somit gilt diese Schlussfolgerung nur für die Teilnehmer des Experiments. Weiter soll erarbeitet werden, dass bei jeder Schlussfolgerung eine Restunsicherheit bleibt. Betont wird ebenfalls, dass beide möglichen Erklärungen (zufällige Zuordnung bzw. bessere Merkbarkeit der ersten Sequenz) möglich bleiben.

8.3 Reflexion zum Baustein Inferenzstatistik

Eine retrospektive Analyse der Lerneinheit geschieht vor allem durch die Auswertung der im Anschluss an die Lerneinheit durchgeführten Studie (Kapitel 11).

Auch wenn hier keine detailreiche retrospektive Analyse im Sinne des Design-based Research Ansatzes (P. Cobb et al. 2003) stattfinden soll, so soll doch eine kurze Reflexion über die drei Sitzungen des Bausteins Inferenzstatistik stattfinden. Als Grundlage dazu dienen neben der persönlichen Erinnerung und neben im Anschluss an die Sitzungen gemachten knappen Notizen Audioaufnahmen aller Sitzungen.

Insgesamt lässt sich festhalten, dass das Design der drei Sitzungen des Bausteins Inferenzstatistik von der Zeitplanung her gut funktioniert hat. Die Sitzungen konnten entlang der Planung realisiert werden. Die Rückmeldungen der Studierenden mittels einer informellen Moodle-Umfrage ergab, dass es in diesem Baustein für sie besonders wertvoll war, die Vorlesungsinhalte durch die PowerPointfolien vorab erhalten zu haben, da sie dadurch während der Vorlesungen nicht mit Mitschreiben beschäftigt waren, sondern sich besser auf die Inhalte fokussieren konnten[43]. In beiden Vorlesungen waren die Studierenden sehr ruhig, was ungewöhnlich im Licht der vorangegangenen Vorlesungen der anderen Bausteine war.

Es wird nicht davon ausgegangen, dass alle angeführten Aspekte direkt nach der ersten Einführung am 20.07.2017 von den Studierenden verstanden und sogar verinnerlicht wurden. Ziel der ersten Sitzung war es, in die Denkweise der Inferenzstatistik mit Randomisierungstests einzuführen, um hierauf in den beiden folgenden Sitzungen (Seminar und Vorlesung) aufzubauen. In der zweiten Sitzung, dem Seminar am 25.07.2017, haben die meisten Studierenden ebenfalls im Vergleich zu vorangegangenen Seminarsitzungen sehr gut und konzentriert mitgearbeitet. Retrospektiv lässt sich sagen, dass es wünschenswert gewesen wäre, einige Aspekte im Seminar stärker vertiefen zu können.

Bezüglich möglicher Erklärungen für die beobachteten Unterschiede wurde die erste Erklärung, dass es daran liegt, dass sich Sequenz 1 besser merken lässt als Sequenz 2, von den Studierenden genannt. Die zweite Erklärung, dass die gemerkte Sequenz keine Rolle spielt, sondern die zufällige Zuordnung der Teilnehmer zu den Sequenzen zu den beobachteten Unterschieden geführt hat, wurde nicht von den Studierenden genannt, sondern musste von dem Dozenten vorgegeben werden. Es wäre wünschenswert, in der Seminarsitzung mehr Rückmeldungen zu den Bearbeitungen geben zu können, beispielsweise zu den aufgestellten Null- und Forschungshypothesen. Dies konnte nur für sehr wenige Meldungen aufgrund der knappen Zeitplanung gegeben werden.

Die händischen Randomisierungen dauerten länger als vorab geplant. Die Kleingruppen schafften in den geplanten zwanzig Minuten zwischen drei und neun händischer Randomisierungen. Wünschenswert wäre eine Anzahl von mindestens zehn händischen Randomisierungen, um besser zu sehen, wie sich die Referenzverteilung aufbaut und ein Gefühl für die Schwankung der Testgröße zu bekommen. Auch für das eigenständige Erstellen der Simulation durch die Studierenden wäre mehr Zeit wünschenswert. Durch die Begrenzung dieses Schritts auf etwa 20 Minuten wurde die Simulation durch Nachahmen des Demonstrierten vollzogen. Hier konnte nicht sichergestellt werden, inwieweit durch die Demonstration in der ersten Vorlesung und dem Nachvollziehen in der Seminarsitzung die instrumentelle Genese der Lernenden ausgeprägt werden konnte.

[43] Natürlich wurden auch für die anderen Vorlesungen die Folien im Vorfeld bereitgestellt, dies wurde jedoch weniger durch die Studierenden herausgestellt.

Ebenso wäre eine vertiefende Diskussion über die Analogie zwischen der händischen Simulation und der Abbildung dieses Prozesses in TinkerPlots vorteilhaft gewesen, um die Ebenen Kontext, Statistik und Software stärker mit einander in Bezug zu setzen wie in Kapitel 6 erläutert.

Ein weiterer Aspekt, der gegen Ende kurz gekommen ist, war die ausführliche Interpretation der verschiedenen Ergebnisse der Studierenden. Da die Studierenden mit ihren eigenen Daten arbeiten durften, ergaben sich durch den Randomisierungstest sehr unterschiedlich große P-Werte von nahezu Null bis hin zu knapp 0,3. Dies hätte gut als Gelegenheit genutzt werden können, verschieden große P-Werte zu interpetieren. Leider ließ die Zeit am Ende der Seminarsitzung eine ausführliche Diskussion nicht mehr zu, es wurde nur auf kleine P-Werte von etwa 0,05 und große P-Werte, etwa 0,3 eingegangen.

Trotzdem wird die Tatsache, dass ein kompletter Randomisierungstest inklusiver händischer Aktivität und Simulation in einer 90minütigen Einheit durchgeführt wurde, retrospektiv positiv bewertet. Es wurden alle geplanten Aspekte im Seminar umgesetzt und obwohl es ein sehr straffer Zeitplan war, gab es am Ende doch einige Nachfragen von Seiten der Studierenden, die zeigten, dass sie den Inhalten aufmerksam gefolgt waren. Gleichzeitig gab es jedoch auch Rückmeldungen, dass sich die Studierenden mehr Zeit gewünscht hätten, um vor allem die eigenständige Simulation und die Interpretation vertiefen zu können.

In der letzten Vorlesung wurde noch einmal ein Randomisierungstest anhand des Schemas zur Durchführung eines Randomisierungstests vorgeführt und in Einzelheiten den Studierenden vorgestellt. Hier wurde besonderer Wert auf die Definition und Ermittlung des P-Werts und den daraus möglichen und nicht möglichen Schlussfolgerungen gelegt. Auch in dieser Sitzung waren die Studierenden sehr ruhig, und trotz der ausführlichen Informationen auf den Folien haben viele Studierende zusätzliche Notizen gemacht. Insgesamt entsteht nach dieser subjektiven Reflexion der Eindruck, dass der Baustein Inferenzstatistik gelungen ist.

Die Verwendung des Schemas zur Durchführung eines Randomisierungstests hat sich als Strukturierungshilfe in diesem Baustein bewährt.

8.4 Fazit zur Lerneinheit „Inferenzstatistik mit Randomisierungstests“

Mit der Darstellung in Abschnitt 8.3 wurde gezeigt, wie eine Lerneinheit zur Inferenzstatistik mit Randomisierungstests aussehen kann, um die eingangs aufgestellte Forschungsfrage *Wie kann eine konkrete Lerneinheit zur Inferenzstatistik mit Randomisierungstests für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung konzipiert und gestaltet werden, welche die instrumentelle Genese der Lernenden zur Benutzung der Software TinkerPlots für Randomisierungstests*

anleitet und welche die in Kapitel 6 genannten Konzepte und Ideen umsetzt? zu beantworten.

Für Unterforschungsfrage *Welche Inhalte soll eine solche Einheit umfassen?* wurden auf zwei Ebenen Entscheidungen getroffen. Auf einer strukturellen Ebene wurde aufgrund des aufgearbeiteten Forschungsstands in Kapitel 6 entschieden, einen Dreischritt einzuführen, der die „Kernlogik der Inferenzstatistik“ im Sinne von G. Cobb (2007) zeigt. Dieser Dreischritt wurde als

1-Nachdenken über die Daten

2-Randomisieren und Wiederholen

3-Schlussfolgerungen ziehen

verwirklicht und vor allem in der dritten Sitzung als Struktur eingeführt.

Auf der inhaltlichen Ebene wurde entschieden, die in Kapitel 6.4 identifizierten Konzepte, die zur Argumentation eines Randomisierungstests gehören, nach und nach einzuführen, an zwei unterschiedlichen Beispielen zu diskutieren und auf typische Schwierigkeiten hinzuweisen.

Als Beispiele wurde zur Einführung ein Beispiel von Holcomb, Chance, Rossman, Tietjen, et al. (2010) adaptiert, das einen signifikanten P-Wert beinhaltet. An diesem konnte die Logik des Randomisierungstestens demonstriert werden. Als weiteres Beispiel wurde für eine höhere intrinsische Motivation ein Beispiel genommen, dass von Zieffler und Catalysts for Change (2013) adaptiert wurde, bei welchem die Studierenden zunächst selbst Daten erheben konnten und diese dann mittels Randomisierungstest auswerten konnten. So gelang es, mit echten empirischen Daten zu arbeiten, so wie es empfohlen wird (Pfannkuch et al. 2011), und auf natürliche Weise verschieden große P-Werte und die aus ihnen möglichen Schlussfolgerungen thematisieren zu können.

Für die zweite Unterforschungsfrage *Welche Maßnahmen lassen sich für eine Lerneinheit zur Inferenzstatistik erstellen, um den Lernprozess zu unterstützen?* wurden im Wesentlichen drei Maßnahmen entwickelt.

Erstens wurde das Schema zur Durchführung eines Randomisierungstests (Abbildung 174) entwickelt und eingesetzt, um eine Struktur vorzugeben und den *cognitive load* (Chandler & Sweller 1991) zu verringern.

Zweitens wurde entschieden, den Studierenden Gelegenheit zu geben, eine händische Simulation selbst durchzuführen, auch wenn die Studienergebnisse von Holcomb, Chance, Rossman und Cobb (2010) keine Aufschluss über eine bessere Wirksamkeit gegenüber der alleinigen computergestützten Simulation zulassen. Hiervon wurde sich jedoch erhofft, dass der Prozess des Randomisierungstests im Sinne des operativen Handelns (Wittmann 1985) besser verstanden und durchdrungen wird.

Als drittes wurde auf der bereits im Laufe der gesamten Veranstaltungen erfolgten instrumentellen Genese hinsichtlich der stochastischen Simulation mit TinkerPlots aufgebaut und für die Simulation in den Vorlesungen und zum eigenständigen Durchführen im Seminar das immer gleiche Vorgehen genutzt und erklärt. Dadurch konnten die *instrumented action schemes* (Rabardel 2002) für einen Randomisierungstest ausgeprägt werden.

Auf Grundlage der im Sommersemester 2017 durchgeführten Lerneinheit werden folgende Hypothesen generiert:

- Die in Kapitel 6.4 identifizierten Konzepte stellen wichtige Komponenten eines Randomisierungstests dar, die von Lernenden adressiert werden müssen, um einen Randomisierungstest selbständig durchzuführen.
- Auf Basis der Eigenschaften der Software TinkerPlots wurde ein instrumentelles Design für die Lerneinheit entwickelt, das in der so entwickelten Form gelingt und die instrumentelle Genese der Lernenden ausreichend anleitet, um TinkerPlots in der Durchführung eines Randomisierungstests gewinnbringend einzusetzen.
- Durch eine im Schema zur Durchführung eines Randomisierungstests eingangs aufzustellende Forschungsfrage und den expliziten Bezug in der Bezeichnung dazu wird die Verwechslungsgefahr von Null- und Forschungshypothese deutlich verringert.
- Das eingeführte Schema zur Durchführung eines Randomisierungstests (Abbildung 181) strukturiert das Vorgehen der Lernenden bei der Durchführung eines Randomisierungstests.

Die Wirksamkeit der designten und durchgeführten Lerneinheit wird nun durch die in den Kapitel 9 bis 11 beschriebene Fallstudie überprüft.

9 Design der explorativen Fallstudie „Randomisierungstests mit TinkerPlots“

Bis zum Entstehen dieser Arbeit wurde noch nicht untersucht, wie der komplette Bearbeitungsprozess, Kommunikation der Teilnehmenden, inhaltliche Argumentationen hinsichtlich der benötigten Konzepte und die Verwendung der Software TinkerPlots im Prozess von Lernenden beim Durchführen eines Randomisierungstests mit TinkerPlots aussehen. Mit Hilfe einer explorativen Fallstudie sollen hier die kognitiven Prozesse von Lernenden beim Argumentieren während des Bearbeitens eines Randomisierungstests untersucht werden und mit antizipierten Schwierigkeiten aus einer a priori Analyse verglichen werden. Während Biehler et al. (2015) u. a. ein Schema für Randomisierungstests entwickelt und untersucht haben, welche statistischen und auf TinkerPlots bezogene Schritte eines Randomisierungstests Lernende erfolgreich (im Sinne einer ja/nein Kodierung) durchführen können, haben Noll und Kirin (2017) dieses Schema angewendet, um vertiefend zu untersuchen, wie eine Zufallsmaschine in TinkerPlots von Lernenden genutzt wird, um die Nullhypothese zu einem Randomisierungstest zu modellieren. Pfannkuch et al. (2015) haben untersucht, wie Studierende spezielle Fragen über einen Randomisierungstest beantworten (z. B. bei der Frage nach möglichen Erklärungen für beobachtete Unterschiede zwischen zwei Gruppen), wobei sich diese Autoren hauptsächlich auf die Auswertung eines schriftlichen Vor-/Nachtests zu einer Einheit „Randomisierungstests“ mit angeschlossener Bearbeitung in der von ihnen entwickelten VIT-Software (vgl. Kapitel 6.1) konzentrierten. Wie bereits am Anfang dieser Arbeit beschrieben, soll mit der vorliegenden explorativen Fallstudie untersucht werden, welche Schwierigkeiten beim Durchführen eines Randomisierungstests mit TinkerPlots durch Paare von Studierenden auftreten und wie Studierende im Hinblick auf die verschiedenen Konzepte argumentieren, die zur Durchführung eines Randomisierungstests in Kapitel 6.4 identifiziert wurden. Hieraus sollen Faktoren für gelingende, kollaborative Bearbeitungsprozesse abgeleitet werden. Hierzu wurden bisher in Deutschland keine Untersuchungen durchgeführt.

In diesem Kapitel wird das Design der Studie beschrieben. Zu Beginn werden die Forschungsfragen zur Studie vorgestellt. Die im vorigen Kapitel 8.2.2 dargestellte Lerneinheit „Inferenzstatistik mit Randomisierungstests“ bildet das Hintergrundwissen der Teilnehmer. Im Anschluss wird die Aufgabe der Studie vorgestellt und eine a priori Analyse dazu durchgeführt, um Lösungswege und mögliche Schwierigkeiten aufzudecken. Es folgt eine Beschreibung des Designs und des Ablaufs der Studie wie sie im Sommersemester 2017 stattfand.

9.1 Forschungsfragen zur Studie

In dieser Studie sollen Einblicke in die Bearbeitungsprozesse von Studierenden beim Durchführen eines Randomisierungstests mit der Software TinkerPlots gegeben werden. Die geäußerten Argumentationen sollen mit Mitteln der Interaktionsanalyse analysiert werden. Dies soll unter zusammenfassenden und vergleichenden Gesichtspunkten geschehen. Da es sich hierbei um eine explorative Studie handelt, sind somit auch spezifische Unterfragen der jeweiligen Forschungsfragen von Interesse. Die Forschungsfragen an die Studie lauten im Hinblick auf mögliche Problemstellen (Forschungsfrage 2), auf Faktoren für gelingende, kollaborative Bearbeitungsprozesse (Forschungsfrage 3) und Erkenntnisse für das Redesign der Lerneinheit (Forschungsfrage 4):

2. Inwieweit treten die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen auf? Welche unterwarteten Problemstellen können identifiziert werden?

Die a priori Analyse der zu bearbeitenden Aufgabe sieht einen „idealen“ Lösungsweg vor, wie er von Studierenden, die an der Lerneinheit Inferenzstatistik teilgenommen haben, begangen werden könnte. Über kollaborative Bearbeitungsprozesse weiß man nach der a priori Analyse noch nichts, deshalb steht auch im Forschungsinteresse, wie Studierende zu zweit eine Randomisierungstestaufgabe bearbeiten. Hieraus resultiert die dritte Forschungsfrage.

3. Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots können festgestellt werden?

Im Hinblick auf die retrospektive Analyse und ein mögliches Redesign der Lerneinheit wird die vierte Forschungsfrage gestellt.

4. Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?

9.2 Die Interviewstudie

Im Sinn von Beck und Maier (1993) und Döring und Bortz (2016) soll ein Leitfadeninterview durchgeführt werden, allerdings in einem angepassten Format. Nach einer kurzen Einführung sollen die Teilnehmer in Zweierteams einen Randomisierungstest selbständig durchführen. Wie von Beck und Maier (1993) beschrieben, soll die Aufgabe selbständig bearbeitet und dabei die Gedanken möglichst ausgesprochen werden. Anschließend werden anhand eines Fragenkatalogs zu den verschiedenen Konzepten eines Randomisierungstests (Kapitel 6.4) Interviewfragen gestellt, um die Studierenden anzu-

regen, ihre zugehörigen Argumentationen zu verbalisieren, damit diese später analysiert werden können. Da im Vorfeld während der Lehrveranstaltung, an der die Studierenden teilgenommen und das Verfahren des Randomisierungstests kennen gelernt haben, die Teilnehmer nur einen einzigen Randomisierungstest selbst durchgeführt haben, sollen während der selbständigen Aufgabenbearbeitung Nachfragen bezüglich der Simulation erlaubt sein und entsprechende ‚minimale Hilfen‘ (Leiss 2007) gegeben werden dürfen. Die Studierenden sollen aufgefordert werden, ihre einzelnen Schritte zu verbalisieren und zu erklären. Sollte dies nicht eigenständig geschehen, so gibt ein Fragenkatalog den Rahmen vor, um die Verbalisierung der entsprechenden Argumentationen hervorzurufen.

Um die Argumentationen der Studierenden zu verstehen, sind Einzelfallanalysen notwendig, da nur durch diese eine detailreiche Analyse möglich ist (vgl. Häder 2015). Das Leitfadeninterview erlaubt durch direktes Nachfragen genaue Argumentationsstrukturen hervorzuarbeiten. Es geht beim Leitfadeninterview darum, die subjektiven Argumentationen der Teilnehmer zu erfassen, dies deckt sich mit dem Forschungsinteresse dieser Arbeit.

Im nächsten Abschnitt wird die Aufgabe der Studie vorgestellt und analysiert. Daran anschließend findet sich der Fragenkatalog für das Leitfadeninterview.

9.2.1 Die Aufgabe der Studie

Als Aufgabe für die Studie soll eine Aufgabe dienen, der reale Daten zugrunde liegen und die in die Lebenswelt der Studierenden passt (wie zum Beispiel von Garfield & Ben-Zvi 2008b gefordert). Es sollen numerische Daten zu zwei Gruppen vorliegen, da dies zu den Beispielen der Vorlesung passt (vgl. Kapitel 8.2). Dabei soll der Kontext der Aufgabe für die Studierenden ein aktuelles Thema aufweisen ohne dabei stark zu polarisieren. Ein möglicher Bereich, in welchem auch Randomisierungstests häufig Anwendung finden, ist der medizinische, sofern ein nicht zu spezielles Thema gewählt wird. Die Gestaltung der Aufgabe soll ähnlich zu Aufgaben aus der Vorlesung sein, um die Studierenden nicht mit neuen Anforderungen zu überfordern. Es soll die Möglichkeit gegeben sein, die Randomisierung sowohl händisch mit Datenkarten als auch mit Softwareunterstützung durchzuführen. Um den Lösungsprozess zu strukturieren, soll das in der Veranstaltung eingeführte Schema zur Durchführung eines Randomisierungstests als Hilfsmittel zur Verfügung gestellt werden.

Ein Kontext, der diese Anforderungen erfüllt, ist im Pre-/Posttest Szenario von Pfannkuch et al. (2015) durch die „Fish Oil and Blood Pressure Study“ gegeben.[44] Dies

[44] Das von Pfannkuch et al. (2015) verwendete Szenario basiert auf einem real durchgeführten Experiment von Knapp und Fitzgerald (1989), in welchem an 32 Personen, die zufällig auf mehrere Gruppen aufgeteilt wurden, untersucht wurde, ob sich durch die regelmäßige Einnahme von Fischöl in Tablet-

wird, wie von Beck und Maier (1993) vorgeschlagen, als Aufgabe für die vorliegende Studie adaptiert. Es wird vermutet, dass das medizinische Phänomen von Bluthochdruck den Teilnehmern der explorativen Studie rudimentär bekannt ist und zusätzlich interessant ist, ob die Einnahme eines einfachen Präparats wie Fischöl zu einer Verbesserung, d. h. Senkung des Blutdrucks führen kann.

Die Formulierung des Aufgabenblatts, das den Teilnehmern der Studie ausgehändigt werden soll, kann der folgenden Abbildung 180 entnommen werden.

Aufgabe: Fischöl und Blutdruck – eine Studie
(Basierend auf einer Studie von Knapp und Fitzgerald (1989). The antihypertensive effects of fish oil: A controlled study of polyunsaturated fatty acid supplements in essential hypertension. *N. Engl. J. Med 320*(16), S. 1037-1043.)

Menschen mit Bluthochdruck haben eine höhere Wahrscheinlichkeit, einen Herzinfarkt zu erleiden. Deshalb suchen Forscher nach einem Weg, den Blutdruck von Bluthochdruckpatienten zu senken. Dazu wurde diese Studie durchgeführt.
Vor der Durchführung dieser Studie vermuteten Forscher, dass Personen, die an einer Fischöl-Diät teilnehmen, dazu neigen, dass ihr Blutdruck stärker gesenkt wird, als bei denjenigen, die eine normale Diät halten. Für die Studie wählten die Forscher 14 männliche Freiwillige mit hohem Blockdruck aus und teilten sie zufällig in zwei Gruppen ein. Die erste Gruppe machte vier Wochen lang eine „Fischöl-Diät“ und die zweite Gruppe machte eine „normale Öl-Diät“ für vier Wochen. Die Behandlungsgruppe ist also die Fischöl-Diät Gruppe und die Kontrollgruppe ist die normale Öl-Diät Gruppe.
Von jedem Teilnehmer wurde am Anfang und am Ende der Studie der Blutdruck gemessen. Die daraus resultierenden Blutdruckreduktionen(-senkungen) in Millimeter Quecksilber waren:

Fischöl-Diät	8	12	10	14	2	0	0
Normales Öl-Diät	-6	0	1	2	-3	-4	2

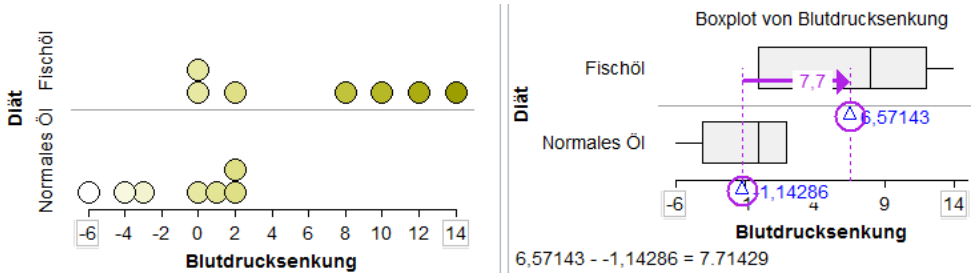


Abbildung 179: Punktdiagramm und Boxplot mit Differenz der Mittelwerte zur Fischöl Studie

Die beobachteten Daten sind in Abbildung 1 dargestellt und zeigen, dass die Senkungen der Blutdruckwerte in der „Fischöl“ Gruppe tendenziell größer sind als die der „normale Öl“ Gruppe.

tenform in verschiedenen Dosierungen der systolische Blutdruck von Bluthochdruckpatienten innerhalb von vier Wochen senken lässt. Pfannkuch et al. (2015) konzentrieren sich in ihrem Szenario ohne die Angabe von Gründen auf die Daten von zwei Gruppen; auf diejenigen, die normales Öl eingenommen hatten (als Kontrollgruppe) im Vergleich mit den Daten einer zweiten Gruppe, die Fischöl in einer hohen Dosierung eingenommen hatte. Eine solche Reduzierung ist auch für die vorliegende Arbeit sinnvoll, da in der Veranstaltung Randomisierungstests nur für zwei Gruppen thematisiert wurden.

Was lässt sich hier nun schließen? **1 - Nachdenken über die Daten** **2 - Randomisieren und Wiederholen** **3 - Schlussfolgerungen ziehen**

Abbildung 180: Das Aufgabenblatt "Fischöl und Blutdruck" zur Fallstudie

Diese Aufgabe soll eingebettet werden am Anfang des Fragenkatalogs, so dass die Teilnehmer die Gelegenheit bekommen, die Daten dieses Experiments zu durchdenken und einen Randomisierungstest durchzuführen. Als Hilfsmittel werden Stifte, ein leeres Blatt, leere kleine Zettel, ein Randomisierungstestschema (Abbildung 181) und ein Laptop mit den in TinkerPlots eingegebenen Daten des Experiments (Abbildung 182) zur Verfügung gestellt. Dabei ist das leere Blatt für Notizen gedacht, die leeren kleinen Zettel können, müssen aber nicht, benutzt werden, um eine Randomisierung der Daten händisch durchzuführen wie in der Lerneinheit gelernt. Für die Simulation des Randomisierungstests soll TinkerPlots genutzt werden, und das Schema ist als Strukturierungs- und Dokumentierungshilfe gedacht.

0. Welche Frage soll beantwortet werden?	
1. Beobachtete Daten Randomisierung? Welcher Unterschied lässt sich feststellen?	
2. Nullmodell Wie lautet die Forschungshypothese? Wie lautet die Nullhypothese?	
3. Simulation Wie wird randomisiert?	
4. Teststatistik Welcher Wert soll überprüft werden? Wie sieht die Referenzverteilung aus?	
5. P-Wert Wie groß ist der P-Wert? Signifikanz?	
6. Schlussfolgerung Erläuterung der Ergebnisse Ablehnen? Kausal? Population?	

Abbildung 181: Schema für die Durchführung eines Randomisierungstests

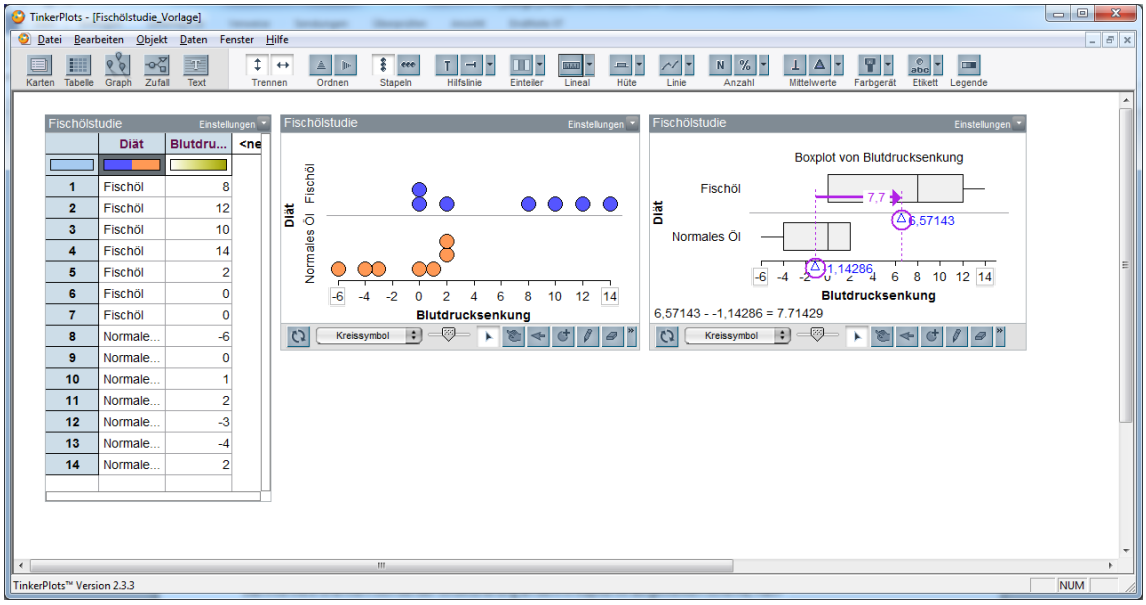


Abbildung 182: Vorbereiteter TinkerPlots-Bildschirm zu Beginn des Interviews

9.2.2 A priori Analyse der Aufgabe

Im Folgenden wird die Aufgabe aus Abbildung 180 analysiert auf Grundlage der Inhalte der Lerneinheit „Inferenzstatistik mit Randomisierungstests“ (Kapitel 8.2). Diese a priori Analyse (wie u. a. bei Hadas & Hershkowitz 2002) soll mögliche Lösungswege der Teilnehmer aufzeigen und mögliche Lösungsalternativen. Es wird erwartet, dass die Teilnehmer nach Besuch der vorgestellten Veranstaltungen in der Lage sind, mit Hilfe des Schemas zur Durchführung eines Randomisierungstests einen Randomisierungstest mit TinkerPlots durchzuführen. Deshalb wird die Struktur des Schemas verwendet, um die a priori Analyse durchzuführen. Die Analyse wird auf dem Niveau durchgeführt, das den Teilnehmern in der Lernsequenz am Ende der Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall I“ im Sommersemester 2017 an der Universität Paderborn vermittelt wurde. Zum einen werden die Schritte des Schemas zur Durchführung eines Randomisierungstests mit einer möglichen Lösung ausgefüllt, zum anderen wird jedoch auf weitere Argumentationen bezüglich der einzelnen Schritte eingegangen. Außerdem wird auf mögliche Schwierigkeiten bei der Bearbeitung der einzelnen Schritte eingegangen. An einigen Stellen im Argumentationsprozess muss statistisch argumentiert werden, an einigen Stellen muss dies in Software umgesetzt werden und an anderen Stellen müssen Softwareergebnisse im Kontext interpretiert werden. Diese Ebenen werden bei der folgenden Analyse miteinbezogen.

Verstehen der Aufgabe

Bevor die Aufgabe von den Teilnehmern bearbeitet werden kann, muss sie zunächst ge-

lesen und verstanden werden. Zum einen muss das Design des Experiments aus dem Text erschlossen werden und zum anderen müssen die beobachteten Daten, die in Tabellenform, als gestapeltes Punktdiagramm und als Boxplots (vgl. Abbildung 182) vorliegen, interpretiert werden.

Die Informationen, die im Kontext eines durchzuführenden Randomisierungstests relevant sind, müssen von den Teilnehmern aus dem Text herausgefiltert werden. Dies kann sowohl im Vorfeld der Aufgabe geschehen, als auch während der einzelnen Schritte des Randomisierungstests. Zum Design des Experiments lässt sich im Aufgabentext finden, dass

- es um Menschen mit Bluthochdruck geht,
- ausschließlich Männer betrachtet werden, die sich zudem freiwillig gemeldet haben,
- insgesamt 14 Personen teilgenommen haben,
- das Experiment 4 Wochen dauerte, in denen eine Gruppe von Personen eine „Fischöl“-Diät machte und die andere Gruppe eine „normale Öl“-Diät,
- die Versuchspersonen zufällig auf die beiden Gruppen aufgeteilt wurden,
- die Blutdruckwerte zu Beginn und zum Ende des Experiments gemessen und die daraus resultierende Differenz protokolliert wurde.

Auf statistischer Ebene kann festgehalten werden, dass eine nichtzufällige Stichprobe von $n = 14$ männlichen Personen vorliegt, die zufällig auf zwei Gruppen der Größen $n_1 = n_2 = 7$ aufgeteilt wurden.

Da es sich bei den Teilnehmern der Interviewstudie um Studierende handelt, wird davon ausgegangen, dass das Textverständnis keine Schwierigkeiten bereitet.

An dieser Stelle ist es zumindest im Rahmen dieser Analyse erwähnenswert, welche Informationen nicht im Text enthalten sind. So finden sich zum Beispiel keinerlei Angaben zu sonstigen möglichen Krankheitsbildern der Teilnehmer oder über weitere Merkmale, wie zum Beispiel Sportverhalten oder Essverhalten. Dies ist insofern relevant, weil gängig angenommen werden kann, dass Bluthochdruck nicht allein vom eingenommenen Öl beeinflusst wird. Es wird jedoch nicht davon ausgegangen, dass die Studierenden detailliertes medizinisches Wissen über die Einflussfaktoren auf den Blutdruck haben. Da es sich um echte Daten einer wissenschaftlichen Studie handelt, sollte davon ausgegangen werden, dass die Studie unter kontrollierten Bedingungen durchgeführt wurde.[45]

[45] In der Tat lässt sich in der Originalstudie nachlesen, dass der Lebensstil der Versuchspersonen kontrolliert und teilweise geregelt wurde. So waren beispielsweise die Bedingungen, unter denen zu Beginn und am Ende des Experiments der Blutdruck gemessen wurde, genau geregelt. Der Orginialtext

Weiter müssen die beobachteten Daten, die in der Tabelle gegeben und in den beiden verschiedenen Graphiken dargestellt sind, im Kontext verstanden werden. Hier ist die Information im Text entscheidend, dass es sich um Blutdruckreduktionen handelt, die, und das muss nun erschlossen werden, aus der Differenz des am Ende des Experiments gemessenen Blutdrucks und des am Beginn des Experiments gemessenen Blutdrucks resultieren. Darüber hinaus muss man sich klarmachen, dass ein positiver Wert eine Verbesserung darstellt, denn je größer die Reduktion ist, desto stärker ist der positive Unterschied vom Messen zu Beginn zum Messen am Ende des Experiments. Reduktionen von Null bedeuten, dass der Blutdruckwert zu den beiden Messzeitpunkten gleich war, es also keine Verbesserung gegeben hat und negative Werte bedeuten, dass im Vergleich zum Beginn des Experiments der Blutdruck sich erhöht hat. Hier wird ebenfalls davon ausgegangen, dass dies von den Teilnehmern der Interviewstudie verstanden wird, da es sich um ein einfaches Verständnis der einzelnen Datenwerte handelt.

Am Ende des Aufgabenblatts findet sich der Hinweis, dass „die Senkungen der Blutdruckwerte in der ‚Fischöl‘ Gruppe tendenziell größer sind als die der ‚normale Öl‘ Gruppe mit der Frage, was sich hieraus schließen lässt. Die Frage ist bewusst so offengehalten, um die Teilnehmer nicht zu lenken und ihnen die Gelegenheit zu geben, die Aufgabe nach eigener Herangehensweise zu bearbeiten.

Im Folgenden wird nun ein Randomisierungstest hierzu vorgestellt, für den die Schritte des Schemas als Strukturierung verwendet werden. Darüber hinausgehende Fragen werden im Anschluss durch die Fragen des Leitfadens für das Interview geklärt.

Welche Frage soll beantwortet werden?

Zu Beginn der Bearbeitung muss geklärt und motiviert werden, welches die Absicht des durchgeführten Experiments war. Durch die bewusst offen gehaltene Frage am Ende des Arbeitsblatt „Was lässt sich hier nun schließen?“ soll angeregt werden, dass sich durch die Formulierung des zu untersuchenden Sachverhalts in Form einer Forschungsfrage eine tiefere Auseinandersetzung mit dem Inhalt des Experiments ergibt. Zu Beginn des Texts sind die Vermutungen der Forscher notiert, aus denen auf ihr Ziel geschlossen werden kann. Nach Lesen des Texts ergibt sich der Eindruck, dass das Experiment durchgeführt wurde, um herauszufinden, ob sich durch eine Diät mit Fischöl der Blutdruck von Bluthochdruckpatienten bei den 14 freiwilligen männlichen Probanden senken lässt. In der Forschungsfrage wird das Problem (im Sinne des PPDAC Zyklusses von Wild & Pfannkuch 1999) benannt bzw. das Forschungsinteresse erklärt. Da dies ein elementarer Bestandteil einer Forschung ist, soll dies hier nicht vorgegeben werden,

wurde den Studierenden jedoch nicht zur Verfügung gestellt, diesbezügliche Nachfragen hätten jedoch von der Interviewerin beantwortet werden können.

sondern von den Studierenden erschlossen werden, unter welcher Fragestellung das Experiment vermutlich durchgeführt wurde.

Im medizinischen Kontext ist allgemein die Frage nach der Wirksamkeit einer durchgeführten Behandlung aufzufinden. Eine mögliche Forschungsfrage der vorliegenden Studie kann mit Bezug auf den Kontext wie in Abbildung 183 formuliert werden als *Trägt eine Diät mit Fischöl zur Senkung des Blutdrucks bei?* Hierdurch wird die „Forscherperspektive“ eingenommen, aus deren Sicht die Aufgabe formuliert ist und es wird eine Erwartungshaltung aufgebaut, ob nun tatsächlich die Einnahme von Fischöl zu einer Senkung des Blutdrucks führt.

0. Welche Frage soll beantwortet werden?	Trägt eine Diät mit Fischöl zur Senkung des Blutdrucks bei den Versuchspersonen bei?

Abbildung 183: Mögliche Lösung des nullten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Mögliche Schwierigkeiten, die hier auftreten können, sind ungenaue Formulierungen der Frage. So könnte es beispielsweise sein, dass die Frage nur allgemein formuliert wird ohne Bezug zum Kontext, wodurch die Frage jedoch nicht unbedingt falsch, sondern nur abstrakt würde. Da auch in der Veranstaltung immer aus der „Forscherperspektive“ heraus argumentiert wurde, wird erwartet, dass die Studierenden dies hier auch tun. Es wird nicht erwartet, dass es den Studierenden nicht gelingt, eine Frage zu formulieren.

Beobachtete Daten

Im nächsten Schritt ist es nötig, sich mit dem Design und den Daten des Experiments zu beschäftigen. Welches Design liegt dem Experiment zugrunde, welche Werte wurden beobachtet und welche Unterschiede lassen sich herausarbeiten? Sofern die Datenerhebung nicht selbst geplant wurde, müssen diese Informationen rekonstruiert werden. Es wird erwartet, dass die Teilnehmer der Studie nicht alle diese Aspekte präsent haben. Jedoch sieht das Schema zur Durchführung eines Randomisierungstests (Abbildung 181) kurze Hinweise vor, um daran zu erinnern, sich mit dem Design und den Daten auseinanderzusetzen. Der erste Hinweis ist das Wort „Randomisierung?“, um daran zu erinnern, sich damit zu beschäftigen, wie das Experiment designt wurde und um zu prüfen, ob eine zufällige Aufteilung der Versuchspersonen auf die Gruppen stattgefunden hat. Der zweite Hinweis ist die Frage „Welcher Unterschied lässt sich feststellen?“, um bewusst darauf hinzuarbeiten, von möglichen Unterschieden, die sich zwischen den beiden Gruppen finden lassen, einen auszuwählen, der die Teststatistik festlegt.

Damit ein Randomisierungstest durchgeführt werden kann, müssen die Versuchspersonen im Vorfeld des Experiments zufällig auf die Gruppen aufgeteilt worden sein. Das

ist die einzig nötige Voraussetzung (Ernst 2004). Wenn das Experiment nicht selbst geplant und durchgeführt wurde (wie das für die Studierenden hier der Fall ist), so müssen die Informationen aus der Beschreibung geschlossen werden. Denn nur im Fall einer zufälligen Gruppenzuordnung lassen sich hinterher Schlussfolgerungen bezüglich der Wirksamkeit von Fischöl ziehen (vgl. Ramsey & Shafer 2013). Es wird erwartet, dass die Studierenden herausfinden, dass die Versuchspersonen zufällig auf die beiden Gruppen aufgeteilt wurden. Diese Information ist direkt im Text gegeben und sollte somit keine Schwierigkeit darstellen. An dieser Stelle sind genaue Argumentationen jedoch interessant, ob die Studierenden darauf eingehen, aus welchem Grund die Versuchspersonen zufällig auf die Gruppen eingeteilt wurden. Der Grund ist, dass durch die zufällige Aufteilung der Versuchspersonen auf die beiden Gruppen der Einfluss weiterer Merkmale, wie beispielsweise Alter, Essverhalten oder Sportverhalten, kontrolliert werden soll. Hiermit sollen nicht kontrollierte bzw. nicht kontrollierbare Faktoren der Versuchspersonen kontrolliert werden in dem Sinne, dass sie die unabhängige Variable, also die Wirksamkeit des Fischöls, nicht beeinflussen. Dies ist ein zentraler Aspekt an einem Randomisierungstest. Bei der Beschäftigung mit dem Design des Experiments sollte weiterhin auffallen, dass keine Zufallsstichprobe von Personen gezogen wurde, weil erstens nur Männer ausgewählt wurden und diese sich zweitens freiwillig gemeldet hatten. Dies ist ein wichtiger Aspekt für die am Ende erfolgende Reichweite der Schlussfolgerungen.

Anschließend müssen die Daten analysiert werden und ein Vergleich zwischen der Fischöl Gruppe und der „normale Öl“ Gruppe muss durchgeführt werden. Allerdings soll auf dem durchzuführenden Gruppenvergleich kein Schwerpunkt liegen. Aus diesem Grund sind sowohl auf dem Aufgabenblatt als auch in der TinkerPlotsdatei bereits zwei Graphiken gegeben, die den Gruppenvergleich ermöglichen sollen, ohne dass weitere Darstellungen durch die Studierenden notwendig sind. In der Graphik mit dem Boxplot (Abbildung 180, rechts) sind zusätzlich die Mittelwerte eingeblendet und mit Hilfe der Linealfunktion von TinkerPlots ist die Differenz zwischen den Mittelwerten berechnet. Dies zeigt auch der TinkerPlotsbildschirm, der zur Verfügung gestellt wird (Abbildung 182). Auch eine Beschreibung wird mit „die Senkungen der Blutdruckwerte in der ‚Fischöl‘ Gruppe [sind] tendenziell größer als die der ‚normale Öl‘ Gruppe“ bereits auf dem Aufgabenblatt geliefert. Hier lässt sich offensichtlich herausarbeiten, dass der Mittelwert der Reduktionen in der Fischöl Gruppe bei 6,57143 mmHg liegt und derjenige der „normale Öl“ Gruppe bei -1,14266 mmHg, die Differenz zwischen den beiden Mittelwerten also bei ca. 7,7 mmHg liegt. Da dies auf dem Arbeitsblatt bereits so dargestellt ist, wird erwartet, dass die Studierenden sich auf diesen Vergleich zwischen den beiden Gruppen konzentrieren. Weitere mögliche Unterschiede, die sich herausarbeiten lassen, sind diejenigen zwischen den Medianen oder auch zwischen den Maxima oder den Minima. Mithilfe von TinkerPlots lassen sich auch diese Unterschiede leicht ermit-

teln. Der Unterschied zwischen den Medianen beträgt 8 mmHg, zwischen den Minima findet sich ein Unterschied von 6 mmHg und zwischen den Maxima ein Unterschied von 12 mmHg, wobei der größere Wert immer in der Fischöl-Gruppe entspringt. Von all diesen möglichen Vergleichen wurden in der Vorlesung nur die Unterschiede zwischen den arithmetischen Mitteln oder den Medianen verwendet, um einen Randomisierungstest durchzuführen, insofern wird erwartet, dass die Studierenden eventuell zwar weitere Unterschiede herausarbeiten, sich jedoch beim Aufschreiben auf einen dieser beiden konzentrieren.

Es wird im Schema zur Durchführung eines Randomisierungstests explizit nach einem Unterschied gefragt („Welcher Unterschied lässt sich feststellen?“).

1. Beobachtete Daten Randomisierung? Welcher Unterschied lässt sich feststellen?	Die Versuchspersonen wurden zufällig auf zwei Gruppen zu je 7 Personen aufgeteilt. Eine Gruppe erhielt eine Fischöl-Diät, die andere Gruppe erhielt eine normale Öl Diät. Die Differenz der arithmetischen Mittel zwischen der Fischöl-Gruppe und der normale Öl Gruppe beträgt 7,7.

Abbildung 184: Mögliche Lösung des ersten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Schwierigkeiten hierzu werden kaum erwartet, da die Studierenden im Laufe der Veranstaltung viele Gruppenvergleiche kennengelernt und durchgeführt haben und sich hier auf einen Aspekt, nämlich den Unterschied zwischen Mittelwerten, konzentrieren können. Durch die Beschäftigung mit den Daten lassen sich jedoch noch weitere Erkenntnisse gewinnen, durch die weitere Vermutungen angeregt werden könnten. So sieht man zum Beispiel in den Darstellungen, dass die Werte der „normale Öl“ Gruppe zwischen -6 und 2 streuen, die der Fischöl Gruppe zwischen 0 und 14. Auch lässt sich mit einem lokalen Blick, der häufig von Lernenden eingenommen wird (Bakker & Gravemeijer 2004) sehen, dass bei drei Versuchspersonen keine Veränderung stattgefunden hat (also ein Unterschied von 0 beobachtet wurde), davon kommen zwei Teilnehmer aus der Fischöl Gruppe. Hierauf wird ggf. eine Erwartungshaltung aufgebaut. Durch einen solchen lokalen Blick auf einzelne Werte sieht man, dass drei Personen der Fischöl Gruppe im Bereich des oberen Quartils der „normale Öl“ Gruppe liegen, woraus sich manche Studierende eventuell bestärkt fühlen würden, die Unterschiede zwischen den beiden Gruppen als geringer einzuordnen, als sie tatsächlich sind, weil der globale Blick auf die Verteilungen fehlt.

Mögliche Erklärungen für den beobachteten Unterschied

Hat man sich auf einen Unterschied festgelegt und diesen notiert, zum Beispiel den Unterschied von 7,7 mmHg zwischen den arithmetischen Mitteln der beiden Gruppen wie durch die Boxplot Darstellung angeregt, so ist damit implizit die Teststatistik definiert und formal der beobachtete Wert notiert. Es liegt die Frage nahe, woran dieser beobachtete Unterschied liegen könnte. Diese Frage wird im Schema zur Durchführung eines Randomisierungstests jedoch nicht explizit gestellt und deshalb als Teil des Leitfadeninterviews aufgenommen (vgl. Abschnitt 9.4).

Eine mögliche Erklärung für den beobachteten Unterschied ist, dass Fischöl tatsächlich wirksam ist, um Blutdruckreduktionen hervorzurufen. Eine andere mögliche Erklärung liegt in der zufälligen Aufteilung der Versuchspersonen auf die Gruppen und nicht in der Wirksamkeit von Fischöl. Hierbei wird davon ausgegangen, dass jede Versuchsperson unabhängig vom eingenommenen Öl eine bestimmte Blutdruckreduktion nach den vier Wochen aufweisen würde und diese Unterschiede nur zufällig in den beiden Gruppen unterschiedlich hoch sind. Eine dritte mögliche Erklärung sind Störfaktoren, die das Ergebnis beeinflussen. Durch die zufällige Gruppeneinteilung sind diese Störfaktoren jedoch weitgehend kontrolliert.

Mögliche Schwierigkeiten:

Bezüglich der Argumentationen dieser drei möglichen Erklärungen für die beobachteten Unterschiede werden durchaus Schwierigkeiten auf Seiten der Teilnehmer erwartet. Hierbei sind genaue Argumentationen nötig, die zu formulieren eine Herausforderung darstellen kann. Dieser Aspekt wurde während des Bausteins Inferenzstatistik nur an zwei Stellen kurz thematisiert und aus der Studie von Pfannkuch et al. (2015) ist bekannt, dass Lernende gerade Schwierigkeiten haben, die zufällige Aufteilung der Versuchspersonen auf die Gruppen als mögliche Ursache zu benennen.

Nullmodell

Im nächsten Schritt muss ein Nullmodell aufgestellt werden, das getestet werden kann. Dazu kann zunächst die aus der Forschungsfrage abgeleitete Forschungshypothese aufgestellt werden. Eine mögliche Formulierung folgt für die

Forschungshypothese: Fischöl ist wirksam, um Blutdruckreduktionen bei den Versuchspersonen zu verursachen.

Im Gegensatz dazu kann eine Nullhypothese formuliert werden. Es wird erwartet, dass die Studierenden in der Nullhypothese den Sachverhalt ausdrücken, den sie anschließend testen wollen und von dem klar ist, dass es derjenige ist, der bei genügend kleinem P-Wert zugunsten der Forschungshypothese verworfen werden kann. In Verbindung mit

den zuvor besprochenen möglichen Ursachen für die beobachteten Unterschiede lautet eine mögliche Formulierung der Nullhypothese

H_0: Die beobachteten Unterschiede liegen an der zufälligen Zuordnung der Versuchsteilnehmer zu den Gruppen.

Zu dieser Formulierung sind alternative Formulierungen möglich, beispielsweise kann auch formuliert werden, dass die beobachteten Unterschiede unabhängig von der Gruppenzuteilung sind oder dass die beobachteten Unterschiede am Zufall liegen oder eine Kombination aus diesen Formulierungen.

Dadurch dass im Schema zuerst nach der Forschungshypothese und anschließend nach der Nullhypothese gefragt wird, wird erwartet, dass durch die Verbindung zur Forschungsfrage, die zu Beginn aufgestellt werden soll deutlich wird, dass in der Forschungshypothese die „Absicht der Forscher“ ausgedrückt wird und dadurch eine Verwechslung von Forschungs- und Nullhypothese vermieden wird.

2. Nullmodell Wie lautet die Forschungshypothese? Wie lautet die Nullhypothese?	Forschungshypothese: Fischöl ist wirksam, um Blutdruckreduktionen bei den Versuchspersonen zu verursachen. Nullhypothese: Die beobachteten Unterschiede liegen an der zufälligen Zuordnung der Versuchsteilnehmer zu den Gruppen.

Abbildung 185: Mögliche Lösung des zweiten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Aus der Literatur sind Schwierigkeiten beim Formulieren der Nullhypothese bekannt (z. B. Castro Sotos et al. 2007, Liu & Thompson 2009, Vallecillos 1999). Durch die zuvor herausgearbeiteten möglichen Erklärungen wird der Prozess des Aufstellens einer mit den möglichen Erklärungen verknüpften, geeigneten Forschungs- und einer geeigneten Nullhypothese erleichtert und sollte deshalb weniger Schwierigkeiten bereiten. Eine typische mögliche Schwierigkeit wäre es beispielsweise, Null- und Forschungshypothese zu verwechseln (Vallecillos 1999) und zu versuchen, die Forschungshypothese zu testen, also dem Nullmodell nicht die zufällige Zuordnung zugrunde zu legen, sondern das Szenario, in welchem die Wirksamkeit von Fischöl vorausgesetzt wird. Eine weitere Schwierigkeit liegt darin, den Zusammenhang zwischen der Nullhypothese und der im nächsten Schritt auszuführenden Simulation zu sehen, so dass eventuell die richtige Situation simuliert wird, allerdings die Nullhypothese falsch formuliert wird und dies erst nachträglich korrigiert wird, so wie beispielsweise in Biehler et al. (2015) berichtet.

Simulation

Der in der Nullhypothese ausgedrückte Sachverhalt muss nun modelliert werden. Dieser Sachverhalt (Nullmodell) besteht darin, dass die erreichten Blutdruckreduktionen unabhängig von der Einnahme von Fischöl oder von „normalem Öl“ sind. Es kann also davon ausgegangen werden, dass jede Versuchsperson „ihren“ Wert erreicht hat, unabhängig davon, in welcher Gruppe sie war. Somit ist die zufällige Aufteilung der Versuchspersonen auf die beiden Gruppen die Erklärung für die beobachteten Unterschiede. Um zu testen, wie die Unterschiede zwischen den arithmetischen Mitteln verteilt sind, wenn die beobachteten Werte zufällig auf zwei Gruppen aufgeteilt werden, wird diese Situation nun simuliert.

Das bedeutet, dass die Daten des Experiments zufällig neu auf zwei Gruppen aufgeteilt werden müssen. Um einen Zugang hierzu zu bekommen, kann die Situation entweder direkt durch eine Simulation in TinkerPlots ausgedrückt werden, oder es wird zuvor eine händische Randomisierung vorgenommen, wie in der Veranstaltung gelernt. Um sich den Prozess des Randomisierens der Werte ins Gedächtnis zu rufen, ist der Ansatz über eine händische Randomisierung mit Hilfe von Zetteln nützlich. In diesem Fall kann anschließend die Übertragung in die Software geschehen. Es wird zunächst der händische Prozess als Lösungsweg charakterisiert. Diesbezüglich gibt es keine Erwartungen, ob die Studierenden den Weg einer händischen Randomisierung wählen oder nicht. Es wurden als Verdeutlichung in der Veranstaltung solche hands-on Aktivitäten durchgeführt. Ob die Studierenden nun allerdings den Zugang zu einem Randomisierungstest über die händische Aktivität nehmen oder direkt mit der Software arbeiten, soll unter anderem in der vorliegenden Studie untersucht werden.

Übertragung der Nullhypothese in eine händische Simulation

Für die händische Simulation liegen kleine, leere Zettel bereit. Auf diese können die erreichten Blutdruckreduktionen notiert werden. Jeder Zettel stellt nun eine Person dar, die einen bestimmten Wert erreicht hat, unabhängig davon, ob sie Fischöl eingenommen hat oder „normales Öl“.

Alle diese Zettel werden gemischt und zufällig auf zwei neue Gruppen aufgeteilt. Dabei müssen die originalen Gruppengrößen von jeweils sieben wieder erreicht werden.

Mögliche Schwierigkeiten

Bei dem händischen Randomisieren wird deutlich, dass es sich um einen Vorgang des Ziehens ohne Zurücklegen handelt. Dies zu erkennen, wurde von Noll und Kirin (2017) als Schwierigkeit berichtet.

Bezüglich der händischen Simulation kann mit der Schwierigkeit gerechnet werden, dass die Studierenden nicht alle erreichten Werte auf Zetteln notieren, also eventuell ei-

nige Werte doppelt notieren oder einzelne Werte vergessen. Eventuell wissen sie auch nicht mehr, dass ein komplettes Durchmischen der Werte nötig ist und dass die ursprünglichen Gruppengrößen wieder erreicht werden müssen. In diesem kurzen Prozess offenbart sich eine Menge Verständnis bezüglich des Randomisierungstests, da hier die Verbindung zwischen der Nullhypothese und der durchzuführenden Simulation gelegt wird. Das Aufteilen der Werte (Personen) auf zwei neue Gruppen ist die Visualisierung der Nullhypothese, da hierdurch zum Ausdruck kommt, dass zwischen dem eingenommenen Öl und der beobachteten Blutdruckreduktion Unabhängigkeit herrscht.

Übertragung der Nullhypothese in eine TinkerPlots Simulation

An dieser Stelle ist zum ersten Mal ein Arbeiten mit Software nötig. Die Modellierung der Nullhypothese in TinkerPlots verläuft ähnlich zu der wahlweise vorher durchgeführten händischen Simulation. Alle beobachteten Blutdruckwerte werden in ein Bauteil der Zufallsmaschine gelegt (z. B. durch copy&paste), dies entspricht dem Notieren der Werte auf Zetteln. Nun wird noch ein zweites Bauteil benötigt, durch welches eine neue, zufällige Gruppenzuordnung vorgenommen wird. Damit zwei neue Gruppen der Größe sieben entstehen, muss die Einstellung „ohne Zurücklegen“ bei beiden Bauteilen ausgewählt werden und insgesamt müssen 14 Durchgänge gewählt werden. Die Ziehungsanzahl wird von TinkerPlots automatisch auf 2 gesetzt, da bei mehreren Bauteilen automatisch aus jedem einmal gezogen wird. Eine mögliche Modellierung ist in Abbildung 186 zu sehen.

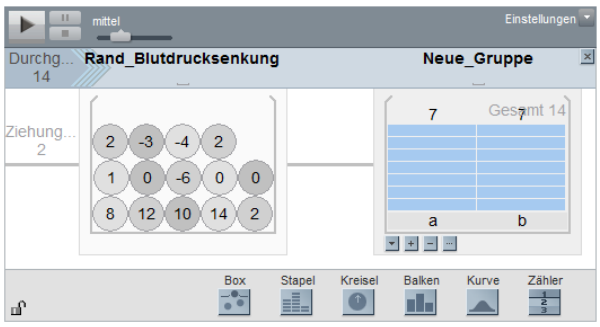


Abbildung 186: Mögliche Übertragung der Nullhypothese in die Zufallsmaschine von TinkerPlots

Bei der Erstellung der Zufallsmaschine gibt es mehrere Alternativen. Zum Beispiel können verschiedene Bauteile gewählt werden. Als Bauteile sind nur solche geeignet, die ein Ziehen ohne Zurücklegen erlauben. Dies sind die Bauteile Box oder Stapel. Alternativ kann für eins der beiden Bauteile auch der Zähler verwendet werden (Ausführungen hierzu finden sich in Kapitel 4). Die Reihenfolge der beiden Bauteile ist beliebig. Es spielt gedanklich, aber nicht für die Simulation eine Rolle, ob zunächst ein Wert der Blutdrucksenkung gezogen wird und anschließend die Gruppe zugeteilt wird oder ob zunächst die neue Gruppe gezogen wird und anschließend ein Wert zugeteilt wird (vgl. hierzu Rossman 2008). In der möglichen Lösung in Abbildung 186 wurden die

beiden Merkmale umbenannt, die Benennung kann variieren, ggf. bis dahin, dass die automatisch vergebenen Namen „Merkm1“ und „Merkm2“ beibehalten werden (vgl. Noll und Kirin (2017) zur Schwierigkeit bei der Nichtvergabe von eigenen Merkmalsnamen). Auch die Etikettierung der Gruppen kann eine Variation beinhalten, ob hier „Eins“ und „Zwei“ oder die automatisch erzeugten „a“ und „b“ verwendet werden, ist unwichtig, solange zwei Gruppenetikettierungen unterschieden werden und ausreichend Werte vorhanden sind.

3. Simulation Wie wird randomisiert?	Die vorhandenen Blutdrucksenkungen werden unabhängig vom eingenommenen Öl zufällig auf zwei neue Gruppen aufgeteilt. In TinkerPlots wird dies durch zwei hintereinandergeschaltete Bauteile realisiert, aus denen jeweils 14 Mal ohne Zurücklegen gezogen wird.

Abbildung 187: Mögliche Lösung des dritten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Im Gegensatz zu den vorherigen Schritten sind hier vielfältige Schwierigkeiten auf Seiten der Teilnehmer denkbar. Dies stützt sich vor allem auf die durchgeführte Werkzeuganalyse in den Kapiteln 3 und 4. Die Modellierung der Zufallsmaschine ist ein komplexer Prozess, bei dem viele Einstellungen vorgenommen werden müssen und viele statistische Aspekte in die Software übertragen werden müssen. Mögliche Fehlerquellen sind die Auswahl falscher Bauteile, also solcher, die die Einstellung „ohne Zurücklegen“ nicht erlauben, wie zum Beispiel das Glücksrad oder das Vergessen dieser Einstellung. Denkbar ist auch, dass es Schwierigkeiten gibt, die ursprünglichen Werte des Experiments in ein Bauteil zu übertragen. Dies kann entweder über einzelnes Eintippen der 14 Werte oder über Kopieren der Werte aus der Tabelle und Einfügen in ein zuvor geleertes Bauteil mit STRG+C und STRG+V geschehen. Eine weitere mögliche Fehlvorstellung liegt darin, die Durchgangsanzahl mit der Gesamtwiederholung des Tests zu verwechseln und hier zum Beispiel 1000 einzutragen statt 14. Die gesamte Erstellung der Zufallsmaschine zeigt in gewissem Sinne das Verständnis über den Randomisierungsprozess zu Beginn des Randomisierungstests, da hier die Unabhängigkeit der Werte von der Gruppe modelliert wird und das zufällige Aufteilen der Werte (Personen) auf zwei neue Gruppen durch das direkte, unabhängige Hintereinanderschalten zweier Bauteile visualisiert wird.

Teststatistik und Referenzverteilung

Wie bereits oben erwähnt, wird die Teststatistik durch den im Schema notierten beobachteten Unterschied implizit definiert. In den obigen Ausführungen wurde hier als

Teststatistik die Differenz des arithmetischen Mittels der Fischöl-Gruppe zum arithmetischen Mittelwert der „normale Öl“ Gruppe gewählt. Wurde alternativ die Differenz der Mediane der beiden Gruppen notiert, so ist dies hier als Teststatistik zu verwenden.

Im Beispiel der händischen Simulation mit Zetteln kann nun diese Differenz ermittelt werden. Dazu wird das arithmetische Mittel der beiden neuen Gruppen berechnet und die Differenz gebildet.

In der TinkerPlots-Simulation müssen ebenfalls die arithmetischen Mittel und die Differenzen gebildet werden. Dies lässt sich in einer Graphik realisieren, die genau so aufgebaut werden kann wie die in der Lernumgebung vorgegebene Graphik zu den beobachteten Daten. Mit Hilfe der Linealfunktion kann die Differenz der arithmetischen Mittel ermittelt werden wie in Abbildung 188.

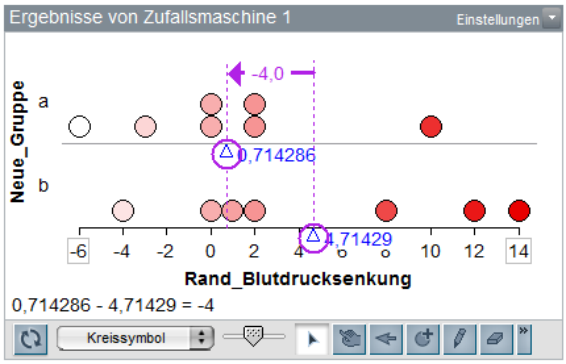


Abbildung 188: Ergebnis einer Randomisierung in TinkerPlots mit berechneter Differenz der Mittelwerte

Durch das häufige Wiederholen dieses Randomisierungsvorgangs und dazugehörigem Berechnen der Teststatistik lässt sich die Referenzverteilung erzeugen. Bei erfolgter händischer Simulation ist nun der Punkt gekommen, an welchem der Schritt in die Software vollzogen werden sollte, weil eine hohe Wiederholungsanzahl händisch kaum möglich ist.

In TinkerPlots lässt sich die Teststatistik als Messgröße definieren, deren Wert in einer eigenen Tabelle gesammelt wird (Abbildung 189), aus der sich anschließend die Referenzverteilung erzeugen lässt (Abbildung 190).

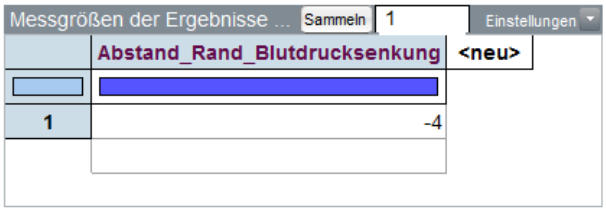


Abbildung 189: Messgrößentabelle zur Teststatistik: Differenz der Mittelwerte

Nachdem dieser Sammelvorgang beispielsweise 5000 Mal durchgeführt wurde, kann eine Referenzverteilung wie in Abbildung 190 erzeugt werden.

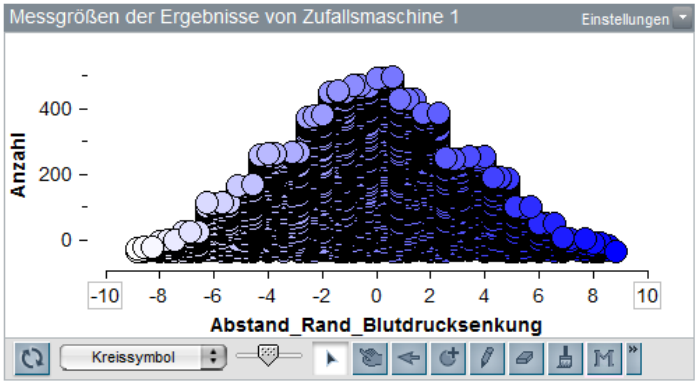


Abbildung 190: Referenzverteilung mit 5000 Werten

Für das Schema zur Durchführung eines Randomisierungstests kann von der simulierten Referenzverteilung eine grobe Skizze angefertigt werden, wie in Abbildung 191.

4. Teststatistik
Welcher Wert soll überprüft werden?

Wie sieht die Referenzverteilung aus?

Teststatistik X: arithmetisches Mittel der Fischöl Gruppe – arithmetisches Mittel der „normale Öl“ Gruppe

Skizze der Referenzverteilung:

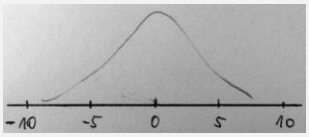


Abbildung 191: Mögliche Lösung des vierten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Mögliche Schwierigkeiten können sich bei der Definition der Teststatistik ergeben, wenn beispielsweise nicht klar ist, welcher Wert hier ausgewählt und berechnet werden muss. Die meisten Schwierigkeiten können sich hier auf der Softwareebene ergeben, dies resultiert aus der durchgeführten Werkzeuganalyse der Kapitel 3 und 4. Schwierigkeiten könnte die technische Umsetzung bereiten, wie eine Graphik wie in Abbildung 188 aufzubauen ist. Es wird eine Graphik benötigt, die einen Gruppenvergleich erlaubt. Dies erfordert eine Darstellung, in der der Mittelwert eingeblendet ist, dazu gehört, dass die Fälle nach den Gruppen getrennt und vollständig separiert werden. Das Auffinden des Mittelwertbuttons oder eine Verwechslung der Buttons für das arithmetische Mittel und den Median könnte sich als Schwierigkeit herausstellen. Weiterhin muss der Abstand zwischen den Mittelwerten gemessen werden, hier könnten sich Schwierigkeiten dadurch ergeben, dass zum Beispiel die Linealfunktion nicht mehr bekannt ist, oder ihre

Bedienung unklar ist. Weiter können sich Schwierigkeiten beim Definieren der Messgröße ergeben, wenn zum Beispiel nicht mehr gewusst wird, wie eine Messgröße ausgewählt werden kann. Die Anzahl der zu sammelnden Messgrößen könnte ebenfalls Schwierigkeiten bergen, wenn zu wenige Messgrößen gesammelt werden, um anschließend eine aussagekräftige Verteilung zu erhalten. Zu viele Messgrößen ergeben keine Schwierigkeit, denn zu viele können nicht gesammelt werden, weil TinkerPlots die Anzahl automatisch auf 5000 begrenzt, falls eine größere Zahl eingegeben wird. Beim Sammelvorgang vieler Messgrößen sind zur Beschleunigung einige Einstellungen vorzunehmen, hier wird jedoch nicht erwartet, dass die Lernenden dies noch wissen, da es sich um rein technische Einstellungen handelt, die für den Verstehensprozess keine Rolle spielen. Hier kann die Interviewerin auf der technischen Ebene eingreifen und diese Einstellungen vornehmen, um den Sammelprozess zu verkürzen.

P-Wert

Ist die Referenzverteilung erstellt, so ist im nächsten Schritt die Wahrscheinlichkeit dafür zu schätzen, ein Ergebnis wie das beobachtete oder noch extremer zu erhalten, unter der Voraussetzung, dass die Nullhypothese wahr ist. Hiermit lässt sich verorten, wie „extrem“ der beobachtete Unterschied des ursprünglichen Experiments ist, unter der Hypothese, dass die zufällige Einteilung der Versuchspersonen auf die Gruppen die Unterschiede hervorgerufen hat. Um diese Wahrscheinlichkeit zu ermitteln, muss in der Referenzverteilung der Anteil der Fälle ermittelt werden, die 7,7 oder größer sind.

Dazu kann in TinkerPlots eine Referenzlinie zur Verdeutlichung des beobachteten Unterschieds von 7,7 eingezeichnet werden, und mit einem Einteiler der Anteil der Fälle, die genauso groß oder extremer sind, bestimmt werden wie in Abbildung 192.

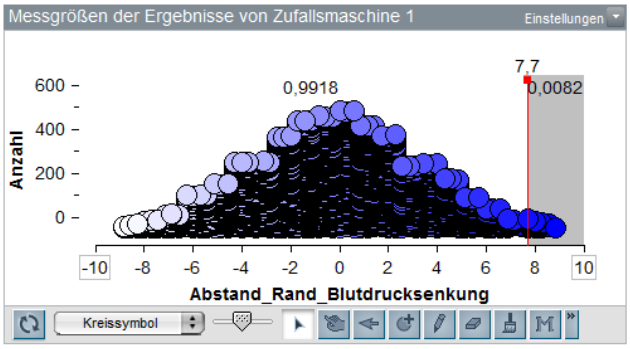


Abbildung 192: Referenzverteilung mit P-Wert

Der P-Wert ist hier geschätzt, da es sich um eine Simulation handelt:

$$P(X \geq X_{beob} | H_0 \text{ ist wahr}) = 0{,}0082 \approx 1\%.$$

Die Studierenden haben in der Veranstaltung die Grenze von 10% kennen gelernt, ab der sie ein Ergebnis als signifikant einstufen sollen. Mit ca. 1% liegt der P-Wert hier

deutlich unter dieser Grenze, es liegt also ein signifikantes Ergebnis vor. Im Kontext ausgedrückt hat man etwa ein Prozent Wahrscheinlichkeit dafür, einen Unterschied von 7,7 mmHg zwischen den Mittelwerten der beiden Gruppen zu erhalten, wenn die zufällige Gruppeneinteilung die Erklärung für den beobachteten Unterschied ist. In 1% kommt also zufällig ein Ergebnis wie das beobachtete oder ein noch extremeres vor.

5. P-Wert Wie groß ist der P-Wert? Signifikanz?	Der P-Wert beträgt ca. 1%, somit liegt ein signifikantes Ergebnis vor.

Abbildung 193: Mögliche Lösung des fünften Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Bezüglich des P-Werts können viele mögliche Schwierigkeiten auftreten. Zuerst einmal ist es möglich, dass die Lernenden nicht mehr wissen, was der P-Wert ist, also dass es sich um eine Wahrscheinlichkeit bezüglich des beobachteten Werts handelt, unter der Voraussetzung, dass die Nullhypothese wahr ist. Es könnte auch sein, dass nicht klar ist, welcher Wert als Grenze zu verwenden ist und in welcher Richtung von diesem Wert an der Anteil bestimmt werden muss. Dies wären Schwierigkeiten auf der statistischen Ebene. Hier kann es wie im Schritt zuvor auch zu technischen Schwierigkeiten bezüglich der TinkerPlots-Bedienung kommen. Möglicherweise ist den Studierenden nicht mehr bewusst, wie der Einteiler zu verwenden ist oder wie überhaupt der Anteil eines bestimmten Bereichs einer Verteilung bestimmt werden kann. Ist der Einteiler gesetzt, so kann auch es sein, dass nicht klar ist, welcher Bereich des Einteilers nun den P-Wert darstellt und so kann eventuell das Gegenereignis, also alle Werte, die kleiner als 7,7 sind, irrtümlich als P-Wert bestimmt werden. Dies hängt damit zusammen, wenn wie oben erwähnt, nicht mehr klar ist, welches Konzept hinter dem P-Wert steht, sondern nur noch gewusst wird, dass „der P-Wert ermittelt werden muss“. Wird sich wiederum auf statistischer Ebene nicht mehr an die Grenze erinnert, ab der ein Ergebnis als signifikant eingestuft wird, so können auch bei der Einordnung Schwierigkeiten auftreten. Wenn der P-Wert ausformuliert wird, so ist eine typische Fehlvorstellung, dass es sich um eine Hypothesenwahrscheinlichkeit handelt, also zum Beispiel um die Wahrscheinlichkeit dafür, dass die Nullhypothese richtig ist oder um die Wahrscheinlichkeit, dass die Forschungshypothese falsch ist (vgl. Garfield & Ben-Zvi 2008a, Krauss & Wassner 2001, Wagenmakers 2007).

Schlussfolgerungen

Im letzten Schritt ist nun die Interpretation des Tests gefordert und Schlussfolgerungen sollen gezogen werden. Es wird erwartet, dass die Studierenden mit einem P-Wert eine Entscheidung darüber treffen, ob die Nullhypothese abgelehnt werden kann. Da hier ein

P-Wert von unter einem Prozent erhalten wurde, lässt sich die Nullhypothese, dass der beobachtete Unterschied von 7,7 zwischen den arithmetischen Mitteln der beiden Gruppen an der zufälligen Gruppeneinteilung liegt, auf dem 1% Niveau ablehnen. Somit kann die Forschungshypothese angenommen werden, und es lässt sich die Schlussfolgerung ziehen, dass Fischöl wirksam ist, um den Blutdruck der Teilnehmer des Experiments zu senken. Eine kausale Schlussfolgerung bezüglich der Wirksamkeit von Fischöl ist hier möglich, weil beim Design des Experiments, wie oben ausgeführt, die Versuchspersonen zufällig auf die beiden Gruppen aufgeteilt wurden und somit weitere Einflussfaktoren ebenfalls zufällig verteilt sind und systematische Fehler, soweit bekannt, vermieden wurden. Allerdings kann nicht geschlossen werden, dass Fischöl nun bei allen Menschen wirksam ist, um Blutdruckreduktionen hervorzurufen, weil die Versuchspersonen männliche Freiwillige waren und keinesfalls repräsentativ. Die Schlussfolgerung muss also auf diesen Personenkreis eingeschränkt werden. Zieht man diese Schlussfolgerungen, so bleibt eine Restunsicherheit in Höhe des P-Werts dafür, sich falsch entschieden zu haben. In ca. einem Prozent der Fälle kommt ein Ergebnis wie das beobachtete oder ein noch extremeres vor, wenn die Nullhypothese wahr ist. Das bedeutet auch, dass man in ca. einem Prozent der Fälle die Nullhypothese irrtümlich ablehnt, was auch als Fehler erster Art bezeichnet wird und in der Veranstaltung thematisiert wurde. Eine mögliche Formulierung der Ergebnisse findet sich in Abbildung 194.

6. Schlussfolgerung Erläuterung der Ergebnisse Ablehnen? Kausal? Population?	Es liegt ca. 1% Wahrscheinlichkeit dafür vor, ein Ergebnis wie das beobachtete oder ein noch extremeres zu erhalten, wenn die Nullhypothese wahr ist. Somit liegt ein signifikantes Ergebnis vor und die Nullhypothese kann abgelehnt werden. Es kann geschlossen werden, dass die Forschungshypothese gilt und dass Fischöl wirksam ist, um den Blutdruck zu senken. Dies kann jedoch nur für die Teilnehmer des Experiments geschlossen werden, da es sich nicht um eine Zufallsstichprobe handelt.

Abbildung 194: Mögliche Lösung des sechsten Schritts des Schemas zur Durchführung eines Randomisierungstests

Mögliche Schwierigkeiten

Mögliche Schwierigkeiten, die hier auftreten können, sind auf der statistischen Ebene verortet. Sofern das Ausformulieren des P-Werts als Wahrscheinlichkeit nicht im Schritt zuvor geschehen ist, so sollte es hier geschehen, wobei die oben erwähnten möglichen Schwierigkeiten auftreten können. Bezüglich der Schlussfolgerungen können weitere Schwierigkeiten auftreten. Zum Beispiel kann der P-Wert irrtümlich dahingehend inter-

pretiert werden, dass bewiesen sei, dass die Nullhypothese falsch ist, obwohl diese Fehlinterpretation explizit in der Veranstaltung thematisiert wurde. Möglicherweise bereitet auch das Konzept Schwierigkeiten, dass ein Ablehnen der Nullhypothese bedeutet, dass die Forschungshypothese angenommen wird. Möglicherweise schließen die Studierenden trotz des kleinen P-Werts nicht auf die Wirksamkeit von Fischöl, weil das Studiendesign der randomisierten Zuordnung der Versuchsteilnehmer zu Gruppen als alleinige Voraussetzung, um diesen Schluss ziehen zu dürfen, für sie nicht ausreicht, wie dies von einigen Teilnehmern der Studie von Pfannkuch et al. (2015) berichtet wird. Es kann auch passieren, dass die Studierenden Schlussfolgerungen auf eine Population ziehen, obwohl dies ebenfalls als nicht möglich in der Veranstaltung thematisiert wurde. Möglicherweise wird der Zufall als Ursache nicht ausgeschlossen (Batanero 2000).

9.3 Geplanter Ablauf der Interviewstudie

Das Interview orientiert sich bei der Strukturierung an dem in Kapitel 6.3 aufgestellten Dreischritt, nach welchem zunächst über die erhobenen Daten nachgedacht werden soll, anschließend die Daten randomisiert und simuliert werden und am Ende Schlussfolgerungen gezogen werden. Ein wesentlicher Bestandteil der Interviewstudie ist das eigenständige Durchführen des Randomisierungstests durch die Teilnehmer. Um ein tieferes Verständnis von den Argumentationen der Teilnehmer zu erhalten, soll in einem weiteren Abschnitt des Interviews nach weiteren Erläuterungen bezüglich der einzelnen Konzepte gefragt werden. So gliedert sich die Laborstudie in drei Abschnitte. Im ersten kurzen Abschnitt werden Fragen zum Schritt *Nachdenken über die Daten* gestellt, um eine Beschäftigung mit dem Thema zu erhalten, bevor zu schnell mit Software gearbeitet wird, wie aus anderen Studien bekannt. Anschließend soll in einem zweiten Abschnitt die Aufgabe möglichst eigenständig von den Teilnehmern bearbeitet werden. Hier soll spontan seitens des Interviewers reagiert werden, sollten sich Fragen oder Schwierigkeiten durch die Teilnehmer bezüglich der Umsetzung in der Software ergeben. Im dritten Abschnitt sollen weiterführende Fragen gestellt werden. Mit den Fragen dieses Abschnitts, die idealerweise nach Durchführung des Randomisierungstests gestellt werden, sollen die Argumentationen bezüglich der einzelnen Schritte des Randomisierungstests und der Umsetzung in der Software erfragt werden. Der letzte Abschnitt des Fragenkatalogs beinhaltet Fragen zu den Schlussfolgerungen, die aus den Ergebnissen gezogen werden können. Hierbei wird bewusst nicht eine Form des stimulated-recall Interviews verwendet, wie zum Beispiel von Busse und Borromeo-Ferri (2003), da nicht nur darauf eingegangen werden soll, was die Teilnehmer an einzelnen Stellen gedacht haben, sondern darüber hinausführende Fragen nach Begründungen anhand eines vorher aufgestellten Fragenkatalogs gestellt werden sollen. Das gesamte Interview wird mit der Software Camtasia aufgezeichnet, wodurch die Bearbeitungen am Bildschirm zusammen mit den verbalen Äußerungen der Beteiligten aufgenommen werden. Zusätzlich

werden die Teilnehmer videographiert, um die Gesten und Aktionen während der Bearbeitung aufzunehmen (vgl. Abbildung 196).

Abschnitt 1 des Leitfadeninterviews: Erste Interviewfragen

Vor dem eigentlichen Interview soll ein Einstieg stattfinden, in welchem die Teilnehmer begrüßt und ggf. einander vorgestellt werden und das Prozedere des Interviews erklärt wird, welches aus einem Teil Interviewfragen und einem Teil an selbständiger Bearbeitung einer Randomisierungstestaufgabe besteht. Danach beginnt das Interview mit dem Vorlesen des Aufgabenblatts (Abbildung 180), woran sich unmittelbar der erste Abschnitt des Fragenkatalogs anschließt. In diesem ersten Abschnitt werden Fragen zu den erhobenen Daten gestellt. Zu den einzelnen Fragen werden optionale Nachfragen formuliert, falls die Teilnehmer die Frage nicht zu beantworten wissen. Wie oben ausgeführt, dient der Fragenkatalog als Orientierung für die einzelnen Abschnitte des Interviews. In der konkreten Situation müssen die im Katalog formulierten Fragen ggf. durch die Interviewerin angepasst, variiert, ergänzt, vorgezogen oder weggelassen werden (vgl. Döring & Bortz 2016 und Mayring 2016).

Tabelle 20: Abschnitt 1 des Leitfadeninterviews

Abschnitt 1: Nachdenken über die Daten	
N0	*Vorlesen des Aufgabenblatts*
N1	Wieso wurden die Versuchspersonen zufällig auf die beiden Gruppen aufgeteilt?
N1O	Optionale Nachfrage: Was wäre, wenn zum Beispiel alle Personen, die viel Sport in ihrer Freizeit machen, in die Fischöl-Gruppe eingeteilt worden wären?
N2	Welche möglichen Erklärungen gibt es für die beobachteten Unterschiede zwischen den beiden Gruppen, die in der Abbildung 1 zu sehen sind?

Die Aufgabe soll zu Beginn durch die Interviewerin vorgelesen werden, um einen gemeinsamen Einstieg zu kreieren. Mit der ersten Frage N1 soll herausgearbeitet werden, inwiefern die Teilnehmer das Konzept des zufälligen Aufteilens der Versuchspersonen auf die beiden Gruppen verstehen. Eine mögliche Begründung für das zufällige Aufteilen lautet zum Beispiel: Durch die zufällige Aufteilung der Versuchspersonen auf die beiden Gruppen werden bestimmte Eigenschaften, die die Versuchspersonen haben und die von Versuchsperson zu Versuchsperson schwanken, zufällig auf die Gruppen aufgeteilt. So wird vermieden, dass Störfaktoren die Ergebnisse des Experiments systematisch verzerren. In einer möglichen Antwort sind zwei Aspekte wichtig. Zum einen sollte formuliert werden, dass der Blutdruck von weiteren Faktoren außer von der eingenommen Öl-Sorte beeinflusst werden kann und dass dies von Versuchsperson zu Versuchsperson unterschiedlich sein kann. Als zweites sollte formuliert werden, dass die zufällige Aufteilung der Versuchspersonen zu den Gruppen eine systematische Verzerrung durch bestimmte Merkmale, wie beispielsweise dem Alter, vermeidet. Sollte eine Begründung nicht in diese Richtung formuliert werden, so kann mit der optionalen

Nachfrage N1O versucht werden, die Teilnehmer durch eine „Was wäre wenn?“ Frage in diese Richtung zu lenken.

Als zweite Frage N2 dieses Abschnitts wird nach den möglichen Erklärungen für die beobachteten Unterschiede gefragt. Diese Frage dient der Vorbereitung einer späteren möglichen Anknüpfung an die Null- und die Forschungshypothese. Eine mögliche Erklärung für die beobachteten Unterschiede ist, dass die beobachteten Werte schwanken und die zufällige Aufteilung der Versuchspersonen auf die Gruppen für den beobachteten Unterschied gesorgt hat. Die andere mögliche Erklärung wäre, dass Fischöl wirksam ist, um den Blutdruck zu senken, dass es also an der durchgeführten Behandlung liegt. Eine dritte mögliche Erklärung wäre, dass andere Einflussfaktoren systematisch für die beobachteten Unterschiede gesorgt haben. Mit Bezug auf die zufällige Aufteilung sollte diese Erklärung jedoch abgelehnt werden.

Abschnitt 2 des Leitfadeninterviews: Durchführen des Randomisierungstests

Anschließend soll der zweite Abschnitt beginnen, in welchem die Teilnehmer den Randomisierungstest durchführen und im Schema zur Durchführung eines Randomisierungstests (Abbildung 181) dokumentieren.

Tabelle 21: Abschnitt 2 des Leitfadeninterviews

Abschnitt 2: Durchführen des Randomisierungstests	
D0	*Austeilen des Schemas zur Durchführung eines Randomisierungstests mit der Anweisung:* Ihr habt das Verfahren des Randomisierungstests kennen gelernt, um solche Experimente zu beurteilen. Jetzt bitte ich euch, dass ihr einen Randomisierungstest durchführt. Hier habt ihr das Schema, das wir in der Vorlesung verwendet haben, um einen Randomisierungstest durchzuführen. Das könnt ihr verwenden, um jetzt hier für die Daten des Fischöl-Experiments einen Randomisierungstest durchzuführen. Als weiteres Material steht euch mein Laptop zur Verfügung und hier liegen auch leere Zettel, falls ihr die benötigt. Wenn ihr fertig seid, dann werden wir einzelnen Aspekten durch weitere Fragen nachgehen.

Im zweiten Abschnitt sollen die Studierenden den Randomisierungstest durchführen. Dazu bekommen sie als Material das Schema zur Durchführung eines Randomisierungstests, leere Zettel, um eine händische Randomisierung wie im Seminar gelernt durchzuführen und einen Laptop mit TinkerPlots und den Daten des Experiments, um zu simulieren.

Sollte es während des Durchführens des Randomisierungstests inhaltliche Schwierigkeiten auf Seiten der Teilnehmer geben, so können Fragen aus dem dritten Abschnitt vorgezogen werden, um dadurch die kognitive Beschäftigung der Teilnehmer mit den problematischen Konzepten anzuregen. Technische Schwierigkeiten im Umgang mit TinkerPlots können durch minimale Hilfen direkt gelöst werden. Beispielsweise wird erwartet, dass die Teilnehmer das Ausschalten der Animation beim Messgrößensammeln nicht beherrschen, da es nur am Rande der Veranstaltungen thematisiert wurde und es sich um eine reine technische Einstellung zum Beschleunigen des Vorgangs handelt, der

keine inhaltliche Bedeutung hat. Weiterhin können Schwierigkeiten auftreten beim Einstellen der Zufallsmaschine, ob mit oder ohne Zurücklegen gezogen werden soll in der Form, dass diese Einstellung nicht vorgenommen wird. Per Default ist in TinkerPlots mit Zurücklegen eingestellt, was für einen Randomisierungstests umgestellt werden muss. Sollte diese Einstellung von den Teilnehmern nicht vorgenommen werden, so kann durch die Interviewerin eingegriffen werden und die entsprechende Frage A2b aus dem dritten Abschnitt vorgezogen werden.

Abschnitt 3 des Leitfadeninterviews: Aufdecken der Argumentationen für den Randomisierungstest

Der dritte Abschnitt beinhaltet nun einen längeren Fragenkatalog, der dazu dienen soll, die Argumentation der Teilnehmer in einzelnen Details hervorzubringen, sofern dies nicht kommunikativ während des zweiten Abschnitts geschehen ist. Hierzu gehören auch Fragen bezüglich der einzelnen Aktionen in TinkerPlots, um die Verknüpfung der technischen Umsetzung in der Software mit der statistischen Ebene zu erfragen. Eine Reihe von Fragen bezüglich der Schlussfolgerungen steht am Ende des Interviews. Bei der Konzipierung der Fragen dieses Abschnitts wird davon ausgegangen, dass zuvor der Randomisierungstest (eventuell mit Hilfe) erfolgreich durchgeführt wurde und das Schema zur Durchführung eines Randomisierungstests ausgefüllt wurde.

Tabelle 22: Abschnitt 3 des Leitfadeninterviews

Abschnitt 3: Nachfragen zu den Argumentationen und Schlussfolgerungen	
A1	Ihr habt als Forschungshypothese und als Nullhypothese folgendes gewählt (→ vorlesen). Welche Erklärung von den vorher genannten steckt da jeweils dahinter?
A2	Bitte beschreibt, was in der Zufallsmaschine geschieht. a. Wieso habt ihr 14 Durchgänge gewählt? b. Wieso habt ihr „ohne Zurücklegen“ ausgewählt? c. Wieso sind sieben <Elemente> und sieben <Elemente> in dem zweiten Bauteil? d. Wieso habt ihr die Merkmale umbenannt/nicht umbenannt?
A3	Bitte erläutert die Referenzverteilung. a. Wofür steht dieser einzelne Punkt über der Null in der Verteilung? b. Wofür steht dieser einzelne Punkt über der <8 – möglichst extrem> in der Verteilung? c. Was bedeutet es, dass hier ein Unterschied von <8 – möglichst extrem> simuliert wurde? d. Ist ein solcher Punkt über der <8 – möglichst extrem> überraschend für euch?
A4	(Sofern nicht bereits geschehen) Was lässt sich nun mit einem P-Wert von etwa 1 Prozent schließen?
A5	(Sofern nicht bereits geschehen) Bitte formuliert Eure Schlussfolgerung im Kontext des Fischölexperiments.
A6	Lässt sich nun sagen, dass Fischöl eine blutdrucksenkende Wirkung hat? Wieso bzw. wieso nicht?
A7	Wie sicher seid ihr bei eurer Einschätzung (bezüglich der letzten Frage)?
A8	Könnt ihr eine alternative Interpretation zu dem P-Wert von 1% geben? (Bei der nicht auf die Wirksamkeit von Fischöl geschlossen wird?)
A9	Nehmt an, dass der P Wert nun 30% beträgt. Was lässt sich jetzt schließen?

Der Fragenkatalog soll an dem Moment ansetzen, in dem die Teilnehmer erklären, dass sie mit der Durchführung des Randomisierungstests und dem Ausfüllen des Schemas fertig sind.

Im Laufe des zweiten Abschnitts sollten im Idealfall von den Teilnehmern eine Null- und eine Forschungshypothese formuliert worden sein. Mit der Frage A1 soll geklärt werden, ob die Teilnehmer einen Bezug zwischen den Hypothesen und den vorher genannten Erklärungen für die beobachteten Unterschiede herstellen können. Die Frage A2 setzt am Nullmodell an, das in der Zufallsmaschine von TinkerPlots ausgedrückt wird. Die Unterfragen beziehen sich auf die Einstellungen, die für eine korrekte Simulation in der Zufallsmaschine vorgenommen werden müssen. Es wird erwartet, dass einige dieser Einstellungen ohne viel Kommunikation der Teilnehmer im zweiten Abschnitt vorgenommen werden, da es sich um die durchzuführenden Einstellungen zu Beginn einer Simulation handelt. Um nun zu erfahren, welche statistischen Begründungen hinter den vorgenommenen Einstellungen stecken, werden die Fragen A2a bis A2c gestellt. Frage A2d stammt daher, dass Noll und Kirin (2017) in ihrer Studie festgestellt haben, dass aus dem Nichtumbenennen der Merkmale in der Zufallsmaschine Schwierigkeiten für die Lernenden entstanden sind. Um zu erfahren, was die Motivation der Teilnehmer war, eine solche Umbenennung der Merkmale vorzunehmen oder auch nicht, dient diese Frage. In der späteren Analyse soll als Teilaspekt mit aufgenommen werden, ob eine Nichtumbenennung der Merkmale ähnlich wie bei Noll und Kirin (2017) zu Schwierigkeiten der Teilnehmer geführt hat. Während der Lerneinheit wurde hierzu keine explizite Norm gesetzt, in allen präsentierten Beispielen wurden die Merkmale jedoch umbenannt.

Die Fragen, die sich unter A3 finden, dienen dem Erfragen des Verständnisses zu einzelnen Punkten der Referenzverteilung. Einzelne Punkte sollen markiert werden und von den Teilnehmern soll erklärt werden, was diese Punkte bedeuten, beispielsweise soll erläutert werden, dass ein Punkt über der Null bedeutet, dass hier zufällig ein Unterschied zwischen den Mittelwerten der neuen Gruppen von Null aufgetreten ist. Durch die zufällige Neuzuordnung der Werte zu zwei neuen Gruppen steht dieser Fall dafür, dass kein Unterschied zwischen den Mittelwerten der beiden neuen Gruppen beobachtet wurde, dass also der Blutdruck weder erhöht noch gesenkt wurde. Die letzte Frage A3d soll die Erwartung der Studierenden beleuchten, da ein Wert von acht noch extremer als der in den Originaldaten beobachtete Unterschied ist.

Bei den Fragen A4 und A5 wird erwartet, dass diese bereits während des zweiten Abschnitts beantwortet werden. Sollte dies jedoch nicht, oder nur fragmenthaft geschehen sein, so kann an dieser Stelle noch einmal nachgefragt werden, was sich aus einem P-Wert von 1% auf statistischer Ebene schließen lässt (A4) und wie dies im Kontext zu formulieren ist (A5).

Die Frage A6 zielt darauf ab, herauszubekommen, ob die Studierenden eine eindeutige Entscheidung für oder gegen die Nullhypothese treffen und ob sie Einschränkungen bezüglich der Verallgemeinerung machen. Eine sinnvolle Einschränkung wäre beispielsweise zu sagen, dass die blutdrucksenkende Wirkung von Fischöl nur für die Versuchspersonen des Experiments gilt und nicht auf eine Grundgesamtheit verallgemeinert werden darf, weil die 14 Versuchspersonen keine Zufallsstichprobe einer Grundgesamtheit darstellen. Hier lässt sich der Kreis schließen zu den anfänglich gemachten Beobachtungen zu den Daten.

Bezüglich der Frage A7 gibt es keinen Erwartungshorizont, da dergleichen nicht in den Veranstaltungen thematisiert wurde.

Mit der Frage A8 soll geprüft werden, ob die Teilnehmer auch eine Interpretation liefern können, bei der die Nullhypothese nicht abgelehnt wird. Möglich wäre eine Interpretation wie *In einem Prozent der Fälle tritt bei zufälliger Zuordnung von Versuchspersonen zu den Gruppen ein Unterschied wie der beobachtete oder ein noch extremerer auf.* Wichtig hieran ist die Erkenntnis, dass es eine gewisse Wahrscheinlichkeit gibt (ausgedrückt durch den P-Wert), durch die zufällige Zuteilung der Versuchspersonen zu den Gruppen einen Wert wie den beobachteten oder einen extremeren zu erhalten.

Die Frage A9 dient dazu zu erfahren, wie die Teilnehmer einen P-Wert jenseits der 10% Grenze interpretieren würden.

9.4 Konkreter Ablauf der Interviewstudie

An der Interviewstudie haben sechs Studierende (drei Paare) des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung teilgenommen. Alle Teilnehmer haben die Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ besucht und insbesondere auch an dem Baustein Inferenzstatistik teilgenommen. Die Teilnehmer wurden einzeln eingeladen, durften allerdings Wünsche äußern, mit wem sie an dem Interview teilnehmen mochten. So kam es dazu, dass ein Paar vorher noch nie zusammengearbeitet hatte und die anderen beiden Paare bereits im Vorfeld miteinander gearbeitet hatten. Termine wurden individuell etwa acht Wochen nach der letzten Vorlesung vereinbart. Aufgrund des gewählten Zeitraums für die Interviews, dieser lag in den Semesterferien, war die Antwortquote bezüglich einer Interviewteilnahme sehr gering. Dies ging aus den abschlägigen Antworten vieler Studierender hervor. Eingeladen wurde mit dem Text der folgenden Abbildung 195.

Liebe/r XY,

dies ist eine Einladung für Dich zur Teilnahme an einer wissenschaftlichen Studie im Rahmen meiner Dissertation, die in den nächsten Wochen stattfindet.

Es sind ca. 60minütige Interviews geplant an folgenden Tagen. […]

Die Interviews sollen immer zu zweit stattfinden. Inhaltlich schließt es an die Veranstaltung Modellieren, Größen, Daten und Zufall I an und an das Werkzeug TinkerPlots. Eine spezielle Vorbereitung ist nicht nötig.

Bitte gib mir Bescheid, ob Du an einem der oben genannten Termine an der Studie teilnehmen wirst. Gerne kannst Du einen Vorschlag bezüglich einer Kommilitonin/eines Kommilitonen machen, mit dem Du gemeinsam am Interview teilnehmen möchtest. Falls Du nicht teilnehmen kannst, gib mir bitte auch kurz Bescheid.

Bei Fragen einfach fragen, gerne per Email!

Ich würde mich sehr freuen, Dich als Studienteilnehmer/in begrüßen zu dürfen!

Liebe Grüße,

Susanne

Abbildung 195: Einladungstext zur Teilnahme an der Studie

Die Teilnehmer der Studie sind somit keinesfalls als repräsentativ anzusehen, da es sich um eine freiwillige Teilnahme handelte und somit davon ausgegangen werden kann, dass hierdurch bereits von Seiten der Studierenden eine Selektion stattfand. So wird sich vermutlich kein Studierender für die Studie gemeldet haben, der nicht an den Themen der Veranstaltungen interessiert war. Bewusst wurde im Text das Thema Inferenzstatistik oder Randomisierungstest nicht erwähnt, um den potentiellen Kreis von interessierten Studierenden nicht noch stärker einzugrenzen.

Es wurden zu Beginn zwei Probeinterviews mit vier weiteren Teilnehmern geführt, um den Fragenkatalog anzupassen. Alle Interviews fanden in einem Laborraum der Universität Paderborn statt (J2.143) und wurden von der Autorin dieser Arbeit durchgeführt. Interviewerin und Teilnehmer saßen gemeinsam an einem Tisch, die Teilnehmer hatten einen Laptop vor sich, an den ein Monitor angeschlossen war, der das Bild des Laptops übertrug. Dies ermöglichte der Interviewerin dem Geschehen der Teilnehmer direkt zu folgen, ohne ihnen „über die Schulter sehen zu müssen“. Die Bildschirmaktivitäten wurden mit der Software Camtasia aufgezeichnet. Seitlich war eine Videokamera aufgestellt. Diese zweite Aufnahmequelle diente zum einen dem Aufzeichnen der Gesten der Teilnehmer, zum anderen als Backup. Das Arrangement ist der folgenden Abbildung 196 zu entnehmen. Vor Beginn des Interviews wurden die Teilnehmer begrüßt und das Arrangement wurde ihnen erklärt. Es wurde ihnen zum ersten Mal das Thema des Interviews genannt, nämlich dass es um Inferenzstatistik und die Durchführung eines Randomisierungstests durch die Teilnehmer geht. Gleichermaßen wurde klargestellt, dass das Ziel des Interviews ist, einen Einblick in die Denk- und Argumentationsweisen der Teilnehmer zu bekommen und dass es keine Testsituation sein solle. Der geplante Ablauf wurde erläutert und betont, dass auch Fragen seitens der Teilnehmer erlaubt sind, beispielsweise bei Schwierigkeiten. Desweitern wurde auf die beiden Auf-

nahmequellen hingewiesen und darauf aufmerksam gemacht, dass die aufgenommenen Daten vertraulich behandelt werden, dass beispielsweise Namen anonymisiert werden, etc.

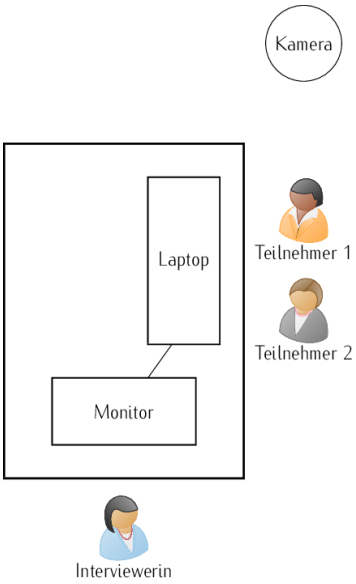


Abbildung 196: Interviewarrangement

Nach dieser Einführung wurden die Aufnahmen gestartet und das Aufgabenblatt vorgelesen.

9.5 Die Teilnehmer

Alle Teilnehmer des Interviews waren weiblich, was nicht verwunderlich ist, da knapp 90% aller Studierenden der Veranstaltungen weiblich waren. Das Alter der Teilnehmer wurde nicht erhoben, es wird auf ca. 21-26 Jahre geschätzt. Alle Teilnehmer befanden sich mindestens im vierten Hochschulsemester, hatten allerdings außer den beiden hier erwähnten Veranstaltungen keine universitäre Vorerfahrung mit Stochastik. Die folgende Tabelle 23 gibt Auskunft über die erreichten Punkte der Teilnehmer im Test, der am Ende der Vorlesung „Modellieren, Größen, Daten und Zufall I“ als Nachweis der Studienleistung geschrieben wurde. Insgesamt gab es 44 Punkte, bestanden war der Test ab 22 Punkten.

Tabelle 23: Testergebnisse (Studienleistung) der Teilnehmer am Ende der Veranstaltung „Modellieren, Größen, Daten und Zufall I“

Paar	Namen[46]	Erreichte Punkte
1	Rebecca	30,5 (69%)
	Selina	38,5 (87,5)
2	Fabia	23 (52%)
	Laura	37 (84%)

[46] Die Namen der Teilnehmer sind anonymisiert.

3	Mandy	23,5 (53%)
	Alisa	25 (57%)

Alle Teilnehmer des Interviews haben demnach den Test zur Studienleistung bestanden, drei von ihnen jedoch nur knapp.

9.6 Datenerhebung

Als Daten stehen zum einen die Aufnahmen der Bildschirmaktivitäten zusammen mit den sprachlichen Äußerungen durch die Software Camtasia und zum anderen die Videoaufzeichnung der Teilnehmer durch die Videokamera im Vordergrund. Die Aufzeichnungen durch Camtasia wurden im Anschluss transkribiert. Darüber hinaus wurden die TinkerPlotsdateien am Ende des Interviews gesichert und die schriftlichen Bearbeitungen eingesammelt.

Somit liegt folgendes Material vor:

- Camtasiaaufzeichnung des gesamten Interviews,
- Videoaufzeichnung des gesamten Interviews,
- Transkript des gesamten Interviews,
- ausgefüllte Zettel „Schema zur Durchführung eines Randomisierungstests“,
- Notizzettel zur händischen Durchführung einer Randomisierung,
- TinkerPlotsdateien.

Die Notizzettel wurden nur von einem der drei Paare verwendet.

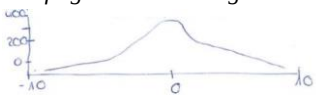
9.7 Transkription

Die Camtasiaaufnahmen wurden mit Hilfe der Software f4 transkribiert und durch eine zweite unabhängige Person abgeglichen und korrigiert. Durch die Transkription wurde die Sprache geglättet, um den Inhalt der Redebeiträge in den Vordergrund zu stellen. Tonhöhenverläufe wurden nicht dokumentiert, Betonungen nur in Extremfällen, besonders lautes Sprechen wurde durch Großbuchstaben gekennzeichnet. Bei der Transkription wurde sich an dem „einfachen Transkriptionssystem“ von Dresing und Pehl (2015, 21ff) orientiert, die sich ihrerseits auf Kuckartz, Dresing, Rädiker und Stefer (2008) berufen. So steht nach diesem System die wörtliche Transkription im Vordergrund, nicht die lautsprachliche, außerdem soll, soweit möglich, in Hochdeutsch transkribiert werden. Wortverschleifungen werden nicht aufgenommen, sondern geglättet. Die Interpunktion wird zugunsten der Lesbarkeit geglättet, es werden also öfter Punkte gesetzt, um ein Absenken der Stimme zu verdeutlichen. Nonverbale Äußerungen, wie beispielsweise Lachen, werden in Klammern aufgenommen. Unverständliche Wörter werden durch (unv) gekennzeichnet, Wort- oder Satzabbrüche werden durch / gekennzeichnet. Kürzere Pausen werden durch (.), längere Pausen durch (…) oder die Angabe der

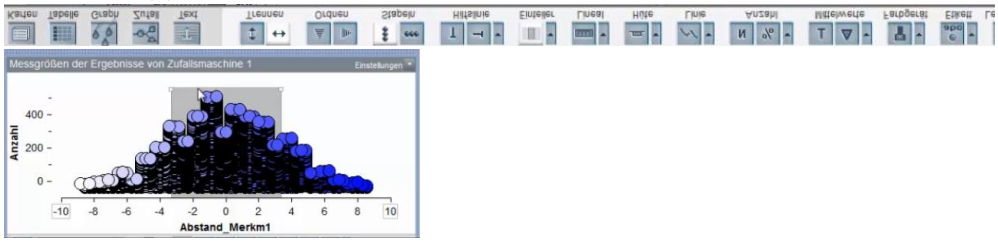
Sekunden, z. B. (5sec) für eine fünf sekündige Pause, dargestellt. Zusätzlich zu diesen Regeln werden die am Laptop durchgeführten Aktionen in das Transkript aufgenommen und in kursiver Schreibweise beschrieben. Dazu werden Screenshots eingefügt, um das Verständnis zu erhöhen. Gleichfalls werden in kursiv Beschreibungen anderer Aktivitäten vorgenommen, wenn beispielsweise Notizen gemacht werden. Bei aller Detailtreue muss bedacht werden, dass es sich beim Transkribieren bereits um eine Interpretation des Materials handelt (vgl. Voigt 1984).

Abbildung 197 zeigt einen beispielhaften Ausschnitt aus dem Transkript von Rebecca (R) und Selina (S).

421 S: Ja soll ja nicht schön sein (lacht), soll ja nur eine Skizze sein.
422 R: Ja nicht so krass aber/.
423 S: /Ja doch das trifft es ja schon ganz gut.
424 R: Ungefähr (lacht).
425 S: Ja also Minimum (unv.) Maximum.
426 S: Ich kann ja hier nochmal, ne/
427 R: Null.
428 S: Genau. Damit man das so ein bisschen einfacher erkennen kann.
429 R: Ja. Das ist dann hier ungefähr die Vierhundert.
430 *Die folgende Verteilung wird skizziert:*



431 S: Wie groß ist der P-Wert? (seufzt) Weißt du noch wie man den berechnet?
432 R: Ich meine man könnte das über den Einteiler machen, dass man den auf 7,7 stellt (.). Dann kann man sich ja die Prozente einblenden lassen, dann sieht man ja im Prinzip wie viele Werte gleich oder eben extremer sind.
433 *Die Studentin (R) klickt oben in der Leiste den Button "Einteiler" an.*



434 *Der Einteiler erscheint und die Studentin zieht das rechte Ende weiter nach rechts.*

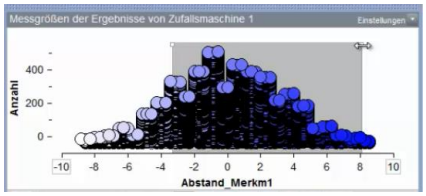


Abbildung 197: Ausschnitt aus dem Transkript von Rebecca und Selina

In diesem Beispielausschnitt findet sich zu Beginn jeden Beitrags die Zeilennummer, um in den späteren Analysen einen eindeutigen Bezug auf einen Gesprächsbeitrag herstellen zu können. In der zweiten Spalte beginnen die Sprecherkürzel (S) und (R) jeden Redebeitrag. Nonverbale Handlungen werden in kursiver Schrift beschrieben, um das Geschehen möglichst vollständig zu dokumentieren und für die spätere Analyse aufzubereiten. Um diese Beschreibungen anzufertigen, wurde teilweise auf die Videoaufzeichnungen der externen Kamera zurückgegriffen oder auf die von den Teilnehmern angefertigten Notizen.

10 Methodologische Grundlagen der Auswertungen

Seit den 1960er Jahren hat qualitative Forschung an Bedeutung gewonnen. Die durch den kritischen Rationalismus geprägten quantitativen Methoden wurden in Frage gestellt und die Diskussion um Forschungsmethoden wurde aufgenommen (Garz 1995). Durch diese Umbrüche erlangte qualitative Forschung an Einfluss und ist heute in vielen Forschungszweigen quantitativen Methoden gleichgestellt.

In der Mathematikdidaktik gibt es seit den 1970er Jahren eine interpretative Richtung qualitativer Forschung (Jungwirth 2003) und es hat sich die interpretative (Unterrichts-) Forschung als eigene Richtung der empirischen Forschung etabliert, wie zum Beispiel der Querschnitt von Beck und Maier (1994) und die Zusammenstellung von Jungwirth und Krummheuer (2006) zeigen. Da der Forschende keinen Zugang zum Denken der beobachteten Personen hat, werden die beobachteten Handlungen interpretiert. Auf Grundlage des Verhaltens und der Äußerungen der Beobachteten werden Interpretationen konstruiert, die möglichst gut dazu passen. Hierzu ist die Analyse der Äußerungen (und Handlungen) der Beobachteten nötig. Durch die Verschriftlichung von Gesprächen, zum Beispiel in Transkripten, können Interpretationen von Äußerungen besser herausgearbeitet werden als zum Beispiel in Interviews.

Um die Kommunikations- und Bearbeitungsprozesse gemäß den gestellten Forschungsfragen (Kapitel 9.1) analysieren zu können bedarf es interpretativer Methoden. Die verschiedenen, in Kapitel 11 verwendeten Methoden werden hier erläutert. Bei den verwendeten Methoden handelt es sich um qualitative Forschungsmethoden, die dazu dienen, Einzelfallanalysen durchzuführen (Flick 2016). In der vorliegenden Arbeit stehen mathematisch geprägte Gespräche im Forschungsinteresse, in denen Lernende fachlich-inhaltsbezogen argumentieren und gemeinsam eine Aufgabe bearbeiten. Einerseits sollen mit Hilfe von Ansätzen aus der Linguistik Gesprächsstrukturen aufgedeckt werden, aus denen sich ein Gesprächsabschnitt zusammensetzt. Auf der anderen Seite soll der im Gespräch gemeinsam hervorgebrachte Sinn erfasst, beschrieben und analysiert werden. Drittens sollen die kollaborativen Bearbeitungen nachvollzogen werden.

Um die Kommunikationsstrukturen der Studierenden zu analysieren, wird in Ansätzen die Gesprächsanalyse von Deppermann (2008) und das Gesprächsphasenmodell von Brinker und Sager (2010) verwendet. Um die Kommunikationsprozesse bezüglich der Thematisierung eines fachbezogenen Inhalts zu rekonstruieren, wird auf die Interaktionsanalyse zurückgegriffen, wie sie von Krummheuer und Naujok (1999) dargestellt wird, zusammen mit der Argumentationsanalyse nach Toulmin (1975). Dies lässt sich mit dem Forschungsanliegen explorativer Forschung (wie in Kapitel 7.3 beschrieben) sehr gut in Einklang bringen (Lamnek & Krell 2016, 48). Hier sei erwähnt, dass eine

solche Analyse eine fast mikroskopische Darstellung der Episoden benötigt, um die Feinheiten der Analysen nachvollziehbar darstellen zu können.

10.1 Aushandlung von Bedeutung als konstruktiver, interaktionistischer Prozess

Den in Kapitel 11 durchgeführten Analysen liegt die Theorie zu Grunde, dass Lernen ein sozialer, konstruktiver Prozess ist, nach welchem jeder Lerner sein Wissen konstruiert im Rahmen seiner Erfahrungen, wofür insbesondere Interaktionen mit anderen ausschlaggebend sind (Neubert, Reich & Voß 2001). Die hierzu notwendige soziale Interaktion hat ihre Ursprünge in der Ethnomethodologie (Garfinkel 1969) und im Symbolischen Interaktionismus (Blumer 1969). Durch die Teilnahme an Argumentationsprozessen gelingen interaktionale Bedeutungsfindungen für mathematische Zusammenhänge (Miller 1986). Nach Miller geht es um Konsensbildung, somit muss eine von allen Seiten akzeptierte Lösung argumentativ ausgehandelt werden (Krummheuer & Brandt 2001, Miller 1986, 2006). Miller betont außerdem, dass über den einzelnen Argumentationsbeitrag hinaus die kollektive Argumentation mehr hervorbringt, als die Summe der Einzeläußerungen (Miller 1986, 25). Dadurch findet ein Aushandlungsprozess statt, der Lernchancen bietet, indem „etwas zu Beginn einer Argumentation noch kollektiv Strittiges (…) in etwas kollektiv Geltendes überführt [wird]“ (Miller 1986, 25). In diesem interaktionalen Aushandlungsprozess wird also mithilfe einer kollektiven Argumentation die Lösung einer Aufgabe kleinschrittig verhandelt, so dass eine als geteilt geltende Deutung des Lösungswegs im Kollektiv erzeugt wird (Krummheuer 2003, Krummheuer & Brandt 2001, Voigt 1991). Dabei werden die subjektiven Vorstellungen der Gesprächsteilnehmer permanent angepasst und Erwartungen und Deutungen werden gemeinsam entwickelt und kontrolliert. „Das Ergebnis ist nicht ein gemeinsames mathematisches Wissen, sondern ein *als gemeinsam und als mathematisch geltendes Wissen*“ (Voigt 1991, 163, Hervorhebung im Original). Krummheuer und Brandt (2001) zeigen jedoch auch, dass im Mathematikunterricht nicht nur „Strittiges“ verhandelt wird, sondern auch Argumentationen nach Aufforderung durch die Lehrkraft entwickelt werden. Diese werden von anderen nicht in Frage gestellt, sondern durch eigene Beiträge ergänzt, mit weiteren Formulierungsideen angereichert und können somit eine konstruktive Weiterentwicklung einer Lösungsidee bilden (Brandt 2006, Krummheuer & Brandt 2001).

10.2 Kommunikationsprozesse analysieren

Wie in Kapitel 9.1 erwähnt, sollen die Fallstudien auf zwei Ebene ausgewertet werden. Erstens soll auf linguistischer Ebene geschaut werden, wie die Kommunikationsstrukturen einzelner Gesprächsepisoden gestaltet sind. Zweitens wird mit interpretativen Methoden der Sinngehalt der gemeinsam hervorgebrachten Deutungen rekonstruiert.

10.2.1 Struktur von Gesprächen

Die grundlegende Einheit eines Gesprächs ist der Turn (Gesprächsbeitrag eines Sprechers). Die Organisation von Sprecherwechseln in einem Gespräch ist im berühmten Artikel von Sacks, Schegloff und Jefferson (1974) beschrieben. Hier wird erklärt, aus welchen Gründen es zu Überlappungen, Pausen, zu Turnergänzungen oder zu Turnabbrüchen kommt, mit dem Fazit, dass der Sprecherwechsel überwiegend reibungslos geschieht. Am Ende jeden Turns wird das Rederecht lokal durch Fremd- oder Selbstwahl von den am Gespräch Beteiligten zugewiesen. Die zugehörigen sozialen Interaktionen sind sequenziell organisiert. „Sie bestehen aus aufeinander folgenden Interaktionsbeiträgen verschiedener Beteiligter, die aufeinander reagieren“ (Deppermann 2014, 28). In einem sequentiell organisierten Gespräch lassen sich gewisse Makrostrukturen finden, die jeden Gesprächsabschnitt charakterisieren. In einem solchen Gesprächsabschnitt soll von den Interaktanten eine bestimmte Aufgabe bewältigt werden (Deppermann 2008), zu der mehrere Schritte (Turns) nötig sind. Diese Gesprächsabschnitte sind gekennzeichnet durch wiederkehrende Abläufe, für die drei Phasen charakteristisch sind.

> „Sie beginnen mit einer Eröffnungsphase, in der die Aktivität angekündigt und von den Beteiligten ratifiziert wird, es folgt der Durchführungskern, in dem die eigentliche Aktivität vollzogen wird, und sie schließen mit einer Beendigungsphase ab.“ (Deppermann 2008, 75)

Die Eröffnungsphase dient darüber hinaus der Herstellung der Gesprächsbereitschaft und der Definition des Themas. In der Kernphase wird das Thema entfaltet und die Gesprächsziele werden verfolgt. Die Beendigungsphase ist gekennzeichnet durch die gegenseitig akzeptierte Auflösung der Gesprächsbereitschaft (vgl. Brinker & Sager 2010). „Sowohl Beginn und Abschluß der Bearbeitung einer Interaktionsaufgabe als auch der Übergang von einem Teilschritt innerhalb eines Sequenzmusters zum nächsten werden oft durch *Grenzmarkierungen* angezeigt.“ (Deppermann 2008, 76), wobei diese Grenzmarkierungen explizit formuliert sein können („Wir beginnen jetzt mit…“) oder nur Markierungen im Sinne eines einzelnen Wortes sein können (ebd).

Um eine Gesprächsstruktur zu verdeutlichen, empfiehlt es sich, eine strukturelle Beschreibung des Gesprächs vorzunehmen, wodurch „die makroskopische Gliederung des Gesprächs in Themen- und Handlungssegmente [rekonstruiert wird]“ (Deppermann 2008, 52). Dies kann als Ausgangspunkt für die mikroskopische Interaktionsanalyse verwendet werden, die im Folgenden vorgestellt wird.

10.2.2 Interaktionsanalyse nach Krummheuer und Naujok

Um die Bedeutungsaushandlung im mathematisch geprägten Gespräch zu rekonstruieren, ist ein interpretativer Rekonstruktionsansatz nötig. Ein solcher wurde von Krummheuer und Naujok (1999) mit der Interaktionsanalyse entwickelt. Es handelt sich dabei um ein mehrschrittiges Verfahren, das die thematische Entwicklung eines Gesprächs zu rekonstruieren versucht. Krummheuer und Naujok nennen selbst als Ziel:

„In der Interaktionsanalyse soll rekonstruiert werden, wie die Individuen in der Interaktion als gemeinsam geteilt geltende Deutungen hervorbringen und was sie dabei aushandeln“ (Krummheuer & Naujok 1999, 68).

Dabei ist die Reihenfolge der abzuarbeitenden Schritte wie folgt:

1. „Gliederung der Interaktionseinheit
2. Allgemeine Beschreibung
3. Ausführliche Analyse der Einzeläußerungen – Interpretationsalternativen (re-)konstruieren
4. Turn-by-Turn Analyse
5. Zusammenfassende Interpretation“ (Krummheuer & Naujok 1999, 68)

Dabei kann das Hin- und Zurückspringen zwischen einzelnen Schritten bereichernd sein. Die einzelnen Schritte werden nun näher erläutert.

Gliederung der Interaktionseinheit

Für die Analyse werden zunächst Ausschnitte aus einem Video ausgewählt. Diese Ausschnitte werden Interaktionseinheit genannt. Eine Interaktionseinheit kann nach unterschiedlichen Kriterien gegliedert werden. Mögliche Kriterien sind:

- „Fachspezifische/fachdidaktische (z. B. von Beginn bis Ende der Bearbeitung einer bestimmten Aufgabe,
- Interaktionstheoretische (z. B. vom Auftritt bis zum Abtritt einer Interaktantin/eines Interaktanten oder von Beginn bis Ende einer Interaktionsform wie Hilfe) oder
- Linguistische (z. B. von einem bis zum nächsten zäsierenden Marker wie „so“). (Krummheuer & Naujok 1999, 69)

Die strukturelle Beschreibung eines Gesprächs nach Deppermann (2008) knüpft an den ersten Punkt an und lässt sich so sinnvoll verwenden, um die Interaktionseinheiten in der vorliegenden Arbeit zu definieren.

Allgemeine Beschreibung

Hier soll eine „erste mehr oder weniger spontane und oberflächliche Schilderung“ geschehen (Krummheuer & Naujok 1999, 69). Ein erster vermuteter Sinngehalt soll hier benannt werden und für einen am Kontext interessierten Leser dargestellt werden.

Ausführliche Analyse der Einzeläußerungen

Nach der allgemeinen Beschreibung soll eine ausführliche sequentielle Analyse erfolgen, in der alternative Interpretationsmöglichkeiten entwickelt werden. Dabei soll sie sich an der sequentiellen Struktur des Gesprächs orientieren und folgende Eigenschaften haben:

1. „Die Äußerungen werden eine nach der anderen in der Reihenfolge ihres Vorkommens interpretiert, womit die Interpretation nach vorne offen bleibt.
2. Plausibilisierungen dürfen und können nur rückwärts gewandt erfolgen.
3. Interpretationen müssen sich im Verlauf der Interaktion bewähren.“ (Krummheuer & Naujok 1999, 69)

Für interpretativ besonders schwer zugängliche Stellen schlagen die Autoren vor, potentielle Folgehandlungen zu entwerfen, die im Licht der nachfolgenden Äußerungen beleuchtet werden können. In diesem Schritt sollen auch theoretische Modelle für die Konstruktion von Interpretationsalternativen herangezogen werden. Dies ist konform mit der übergeordneten Konversationsanalyse (z. B. Bergmann 1981, Heller & Morek 2016) und der darauf aufbauenden Gesprächsanalyse nach Deppermann (2008), die ähnliche Postulate für die Interpretationen aufstellen.

Turn-by-Turn Analyse

Angelehnt an die Konversationsanalyse geschieht in diesem Schritt ggf. wieder eine Einschränkung der Deutungsalternativen aufgrund einer Turn-by-Turn Analyse. Durch die Analyse einer auf A folgenden Äußerung B soll verglichen werden, ob die Äußerung B des zweiten Turnnehmers eine oder mehrere Deutungsalternativen stützt. Hierdurch können Deutungsalternativen entfallen oder neue hinzukommen. Da der erste Interaktant gemäß der Konversationsanalyse nun wiederum die Möglichkeit hat, die Interpretation von B zu korrigieren, wird bei einem solchermaßen unterlassenen Korrekturversuch davon ausgegangen, dass A sich von B ausreichend verstanden meint. Zusammenfassen lässt sich die Turn-by-Turn Analyse durch die Frage „Wie reagieren andere Interaktanten auf eine Äußerung, wie scheinen sie die Äußerung zu interpretieren, was wird gemeinsam aus der Situation gemacht?“ (Krummheuer & Naujok 1999, 70).

Zusammenfassende Interpretation

Die am besten zu begründende Gesamtinterpretation wird hier zusammengefasst und kann eine Theoriegenese anstoßen. Da die Schritte 3 und 4 sehr umfangreich sein können, sollen eventuell nur diese zusammenfassenden Interpretationen dargestellt werden mit Bezug auf die vorher durchgeführten Analysen.

Um die Interaktionsanalyse zu ergänzen, schlagen Krummheuer und Naujok (1999) vor, bei Bedarf weitere Analysen in Form von Ergänzungen vorzunehmen. Eine mögliche solche Ergänzung ist es, im Hinblick auf gemeinsam hervorgebrachte Argumentationen eine Argumentationsanalyse ausgewählter Inhalte vorzunehmen. Hierzu notieren die Autoren „die Analyse des Verlaufs und der Inhalte der kollektiven Argumentation kann Antworten auf Fragen zum fachlichen Lernen bringen“ (Krummheuer & Naujok 1999, 73). In diesem Sinne wird ergänzend die funktionale Argumentationsanalyse von Toulmin (1975)vorgestellt.

10.2.3 Funktionale Argumentationsanalyse nach Toulmin

Ein Ansatz, um Argumente strukturell zu analysieren, liefert die funktionale Argumentationsanalyse von Toulmin (1975), in der ein Argumentationsprozess in einzelne Be-

standteile aufgespaltet wird. Eine Argumentation lässt sich nach Toulmin (1975) schematisch darstellen wie in Abbildung 198.

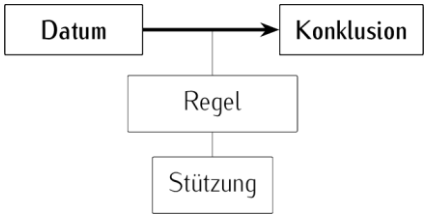


Abbildung 198: Allgemeine Struktur einer Argumentation nach Toulmin (eigene Darstellung, angelehnt an Toulmin 1975)

Ausgangspunkt einer Argumentation nach Toulmin (1975) ist ein „Datum“, aus welchem die zu begründende „Konklusion“ gefolgert werden soll. Dies wird durch allgemeingültige (Schluss-)Regeln begründet, die durch weitere Begründungen gestützt werden.

Im Hinblick auf Randomisierungstests lässt sich die toulminsche Argumentationsstruktur so interpretieren, dass das Datum die anfänglich gemachte Beobachtung eines Experiments ist, aus der weitere Schlüsse gezogen werden sollen. Die Konklusion ist die Schlussfolgerung, die am Ende eines Randomisierungstests steht, zum Beispiel die Ablehnung der Nullhypothese. Als Regeln werden hier Schritte verstanden, die als „Brücken“ dienen können, um eine solche Schlussfolgerung zu erlauben (Toulmin 1975, 89). Diese können laut Toulmin als Antwort auf die Frage, wie man zu diesem Schluss kommt, verstanden werden. Die Stützungen dazu hängen vom jeweiligen situativen Kontext ab und sollen die Frage beantworten, weshalb die formulierte Schlussfolgerung allgemein als zulässig akzeptiert werden soll (Toulmin 1975, 94f).

Das Schema von Toulmin (1975) kann hier helfen, die Struktur der Argumentation von Studierenden im Interaktionsprozess des Durchführens eines Randomisierungstests aufzuzeigen. Mit Hilfe der funktionalen Argumentationsanalyse kann mikroskopisch in die Aussagen Studierender geblickt werden, um die Struktur der Argumentationen bezüglich einzelner, zum Randomisierungstest gehöriger Konzepte (vgl. Kapitel 6.4) nachzuzeichnen.

Anwendungen der funktionalen Argumentationsanalyse, um Argumentationen von Lernenden in der Mathematik zu analysieren, finden sich z. B. in Bezold (2009), Fetzer (2006), Höck (2015), Krummheuer (1997), Krummheuer und Fetzer (2005) oder Schwarzkopf (2000).

10.3 Generieren der Unterforschungsfragen

Es wird erwartet, dass die Teilnehmer die Konzepte, die zu einem Randomisierungstest gehören, anwenden und formulieren können, so wie dies im Baustein Inferenzstatistik

der Veranstaltungen gelernt wurde (vgl. Kapitel 8.2.2). Dabei ist von Interesse, wie sie die einzelnen Konzepte sprachlich adressieren und wie sie im gemeinsamen Gespräch die zugehörige Deutung konstituieren. Dabei soll die Software TinkerPlots eingesetzt werden, um einen Randomisierungstest möglichst selbständig durchzuführen. Das Randomisierungstestschema soll hierzu als Orientierung im Prozess genutzt werden, um die einzelnen Schritte durchzuführen. Als Fragestellung sind die in Kapitel 9.1 aufgeworfenen Forschungsfragen formuliert (die erste Forschungsfrage bezieht sich auf die designte Lerneinheit und findet sich in Kapitel 7.1).

2. Inwieweit treten die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen auf? Welche unterwarteten Problemstellen können identifiziert werden?

3. Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots können festgestellt werden?

4. Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?

Zu Forschungsfrage 2 werden folgende Unterforschungsfragen für die Auswertung gestellt. Hierbei wird die a priori Analyse der Aufgabe (Kapitel 9.2.2) als Bezug genommen.

- Welche Problemstellen treten auf? Welche davon sind unerwartet? Welche Gründe lassen sich identifizieren für die unerwarteten Problemstellen?

Da nach der a priori Analyse der Einfluss der kollaborativen Prozesse des gemeinsamen Aushandelns von Bedeutung nicht bekannt ist, sollen diese mit den folgenden Unterforschungsfargen zu Forschungsfrage 3 untersucht werden. Im Sinne der Zwischenstufe (Flick 2016) sollen Unterforschungsfragen zum Vergleich der Paare für Forschungsfrage 3 gestellt werden, um hierdurch robustere Aussagen treffen zu können als bei reinen Einzelfallanalysen. Die Unterforschungsfragen lauten:

- Wie stark sind die Phasen der Gesprächsführung ausgeprägt? Wie gestalten die Paare die Kommunikation untereinander? Was bedeutet das in Bezug auf die kollaborativen Bearbeitungsprozesse?
- Arbeiten die Teilnehmer der Studie kollaborativ zusammen? An welchen Merkmalen lässt sich die Kollaboration erkennen?

- Können die Teilnehmer die verschiedenen, in Kapitel 6.4 identifizierten Konzepte (einen Unterschied in den beobachteten Daten identifizieren, Nullhypothese aufstellen, etc.), die zur Argumentation eines Randomisierungstests gehören, verwenden und erläutern? Welche Schwierigkeiten treten dabei auf?
- In welcher Weise gelingt es den Teilnehmern, die drei Ebenen Kontext, Statistik und Software aufeinander zu beziehen?
- Können die Teilnehmer die Software TinkerPlots als Instrument einsetzen, um einen Randomisierungstest durchzuführen? An welchen Stellen bei der Benutzung der Software TinkerPlots benötigen die Teilnehmer Hilfe?

10.4 Auswertungsmethodologie zur Fallstudie – angepasstes Vorgehen

Für die einzelnen Paaranalysen soll die Analyse in folgenden Schritten ablaufen:

1. Gliederung der einzelnen Fallstudien in Interaktionseinheiten
2. Interpretative Rekonstruktion des Lösungsprozesses
3. Rekonstruktion einzelner Argumentationen
4. Untersuchung der TinkerPlots-Benutzung

Daran anschließen soll sich zur Beantwortung der Unterforschungsfragen noch ein letzter Schritt:

5. Vergleichende und zusammenfassende Analysen

10.4.1 Gliederung in Interaktionseinheiten

Um das durchgeführte Interview und die Phase des selbständigen Bearbeitungsprozesses zu strukturieren, werden 15 Interaktionseinheiten nach Krummheuer und Naujok (1999) definiert.

Die Auswahl der Interaktionseinheiten orientiert sich an fachspezifischen Charakteristika, hier wird sie vom Beginn einer (Teil-)Aufgabenbearbeitung bis zu deren Ende gewählt. Dabei wurden die Interaktionseinheiten zunächst grob nach den drei Abschnitten des Leitfadeninterviews gegliedert (Kapitel 9.4) und anschließend innerhalb der Abschnitte fein gegliedert entlang der Interviewfragen (Abschnitt 1 und 3) bzw. entlang der in Kapitel 6.4 identifizierten Konzepte für die Durchführung eines Randomisierungstests (Abschnitt 2). Zur Abgrenzung der Interaktionseinheiten wird das Gesprächsphasenmodell nach Deppermann (2008) verwendet, um den Übergang von einer zu nächsten Einheit auf linguistischer Ebene zu kennzeichnen. Hierfür sind vor allem die Eröffnungsphase bzw. Marker, die die Eröffnung eines neuen Abschnitts signalisieren, eine deutliche Markierung eines Wechsels. Auch wenn diese, wie sich in den Analysen in Kapitel 11 herausstellen wird, häufig sehr kurz ist. Eröffnungsphasen sind ent-

weder gekennzeichnet durch Fragen, die die Interviewerin gemäß des Leitfadens stellt, z. B. aus dem Interview von Mandy und Alisa:

3	I: […] Meine erste Frage ist: Wieso wurden die Versuchspersonen, die hier teilgenommen haben, zufällig auf die zwei Gruppen aufgeteilt?

Oder Eröffnungsphasen sind, vor allem während des Teils der selbständigen Bearbeitung der Studierenden, gekennzeichnet durch kurze Marker, die anzeigen, dass die Teilnehmer zum nächsten Konzept übergehen, wie z. B. in dem Interview von Rebecca und Selina zum Einstieg in die Diskussion um das Nullmodell:

82	S: Forschungshypothese ist ja wieder, ob Fischöl/ (.) dass das Fischöl gegen Bluthochdruck wirkt oder den Blutdruck senkt.

Hier stellt das Wort „Forschungshypothese“ den Marker dar, der benutzt wurde, um die Eröffnung der Interaktionseinheit zum Nullmodell einzuleiten.

Es ergeben sich aus dem Leitfaden des Interviews (Kapitel 9.4) und den zum Randomisierungstest gehörigen Konzepten (Kapitel 6.4) 15 Interaktionseinheiten, nach denen die drei Transkripte strukturiert werden. Diese sind in Tabelle 24 zu sehen. Zu den Interpretationen in Kapitel 11 der jeweiligen Interaktionseinheiten werden die Zeilennummern der Transkripte angegeben, um für den Leser transparent zu machen, welchen Umfang eine Interaktionseinheit hat. Hierbei ist zu beachten, dass im Transkript neben der vollständigen Kommunikation der Interviewteilnehmer auch Aktivitäten in TinkerPlots und andere Handlungen, wie beispielsweise das schriftliche Festhalten von Ergebnissen im Randomisierungstestschema aufgenommen sind (vgl. Kapitel 9.7).

Tabelle 24: Definierte Interaktionseinheiten zur Strukturierung der Fallstudien

Interaktionseinheit	Thematischer Inhalt
1	Vorstellung und Vorlesen des Aufgabenblatts
2	Begründung für die zufällige Aufteilung auf die Gruppen
3	Mögliche Erklärungen für die beobachteten Unterschiede
4	Überleitung zum Randomisierungstest
5	Forschungsfrage notieren
6	Beobachtete Daten
7	Nullmodell
8	Simulieren/Zufallsmaschine erstellen
9	Teststatistik definieren und Referenzverteilung erstellen
10	Ermitteln des P-Werts
11	Schlussfolgerungen ziehen
12	Zusammenhang zwischen den formulierten Hypothesen und den möglichen Erklärungen

13	Beschreiben der Zufallsmaschine
14	Verständnis über einzelne Punkte der Referenzverteilung
15	Schlussfolgerungen ziehen

Da manche Interaktionseinheiten sehr lang sind, werden als darunter liegende Strukturierungseinheit Episoden definiert. Diese sind inhaltlich definiert am Beginn und Ende eines thematischen Arbeitsschritts (z. B. Planen des weiteren Vorgehens) oder anhand eines Arbeitsschritts in der Software (z. B. Messgrößen definieren). Für die verschiedenen Fallanalysen werden die Episoden induktiv aus dem Material heraus definiert, in Abhängigkeit von dem individuellen Vorgehen der Teilnehmer.

10.4.2 Interpretative Rekonstruktion des Lösungsprozesses

Der Hauptteil der Analyse besteht in der Interpretation der Kommunikation nach der Interaktionsanalyse von Krummheuer und Naujok (1999). Prinzipiell ist die Interaktionstheorie ausgerichtet auf verbale Komponenten. Hier wird der Analyseprozess ausgeweitet auf Aktionen, die mit der Software und weiterem Material (wie Zetteln) durchgeführt werden, da diese maßgeblich den Bearbeitungs- und Kommunikationsprozess mitgestalten. Die Ausweitung der Interaktionsanalyse auf die Benutzung der Software steht neben der Kommunikation im Mittelpunkt des Forschungsinteresses (vgl. Forschungsfrage 3).

In Anlehnung an Biehler et al. (2015) und wie in Kapitel 6 herausgearbeitet, wird im Interpretationsprozess untersucht, auf welcher der drei Ebenen Kontext, Statistik oder Software Kommunikation in den Interaktionseinheiten stattfindet. Hiermit soll ein grober Eindruck wiedergegeben werden, auf welcher sprachlichen Ebene sich die Kommunikation jeweils in den Interaktionseinheiten abspielt. Biehler et al. (2015) haben in ihrem theoretischen Schema die Ebenen getrennt, was von Noll und Kirin (2017) anders beobachtet wurde. Nach dem Schema von Biehler et al. (2015) werden zu Beginn des Randomisierens das reale Problem in Form der vorliegenden Situation und mit bestimmten Beobachtungen und am Ende des Randomisierungstestens die Schlussfolgerungen der Kontextebene zugeordnet. Das statistische Problem in Form der Nullhypothese, das Auswählen einer statistischen Methode, hier das Randomisierungstesten und statistische Schlussfolgerungen bezüglich der Signifikanz eines erhaltenen P-Werts werden der Statistikebene zugeordnet. Schließlich ordnen Biehler et al. die Simulation in TinkerPlots und die Ergebnisse der Simulation einschließlich des P-Werts der Softwareebene zu. Eine Schnittstelle zwischen den Ebenen befindet sich laut den Autoren zwischen der Statistikebene und der Softwareebene in Form des Nullmodels, das diese beiden Ebenen verbindet. Hierzu schreiben die Autoren nur, dass diese Welten ineinander liegen:

> „Here we distinguish between three worlds: The contextual world, the statistical world, and the world of software, each of which is embedded within the other." (Biehler et al. 2015, 138)

Dies wurde von Noll und Kirin (2017) so interpretiert, dass die drei Ebenen bis auf die Verbindung durch das Nullmodell komplett getrennt sind. Somit kommen diese beiden Autorinnen in der Zusammenfassung ihrer Studie zu dem Schluss, nachdem sie Lernende beim Erstellen von Zufallsmaschinen in TinkerPlots zur Durchführung eines Randomisierungstests beobachtet haben, dass Lernende zwischen den drei Welten in einem einzigen Schritt wechseln.

> „Students rarely followed any linear order in solving the Dolphin Therapy Problem and at times they moved back and forth between the three worlds identified by Biehler et al. (2015). Our research indicated that once students began working in the software world the context mediated any translating they did between the statistical world and software world." (Noll & Kirin 2017, 239)

In der Tat lassen sich bereits in den Beschreibungen von Biehler et al. (2015) jedoch Hinweise finden, dass keine strikte Trennung der drei Ebenen gemeint ist, denn beispielsweise für die Formulierung einer möglichen Nullhypothese, die nach ihrem Schema auf der Statistikebene angesiedelt ist, nehmen sie Bezug zum Kontext und verbinden somit die zwei Ebenen. Daran und an den weiteren Ausführungen lässt sich erkennen, dass mit der „Einbettung" der Welten ineinander vielmehr gemeint war, dass die jeweils außen liegende Welt zusätzlich mit einbezogen ist.

In der vorliegenden Studie soll nun noch einmal genauer untersucht werden, ob - und wenn ja, in welchen Interaktionseinheiten - die Ebenen sprachlich miteinander verbunden werden.

Ausführliche Analyse der Einzeläußerungen und *Turn-by-Turn Analyse*

Nachdem die Transkripte in Interaktionseinheiten und in Episoden gegliedert sind, sollen mit Hilfe der Schritte 3 *Ausführliche Analyse der Einzeläußerungen* und 4 *Turn-by-Turn Analyse* von Krummheuer und Naujok (1999) die einzelnen Abschnitte turnweise interpretiert werden. Diese Interpretation geschieht auf Basis der in Kapitel 8.2 beschriebenen Lerneinheit und der a priori Analyse der Aufgabe in Kapitel 9.2.2. Hinzukommend werden für die Interpretation die in Kapitel 6 gewonnenen Einsichten aus den empirischen Studien zu Randomisierungstests genutzt.

Bein Interpretieren während der Interaktionsanalyse ist zum einen die „extensive Interpretation der Äußerungen" (Jungwirth 2005, 5) ein wichtiges Merkmal. Diese „extensive Interpretation" zielt auf weitere Bedeutungen, die die Äußerungen „noch" haben könnten und jede Äußerung wird als auslegungsbedürftig angesehen. „Gegenüber den Intentionen des Sprechers enthält der Text ein ‚surplus of meaning', ein Mehr an Sinn, das in der Interpretation erschließbar ist, wenn gefragt wird, wie ‚man' die jeweilige Stelle noch verstehen kann" (Jungwirth 2003, 191). Das Ziel der Analyse sind Deutungshypothesen, die über Einzelereignisse hinausgehen. Durch den systematischen

Vergleich mehrerer Einzelfallanalysen kann der Allgemeinheitsgrad dieser vergrößert bzw. weiter ausdifferenziert werden (Jungwirth 2005, 6).

Im Sinne der Gesprächsanalyse von Deppermann (2008) handelt es sich hierbei um eine spezielle Form der Sequenzanalyse eines Gesprächs. Deppermann formuliert in einem weiteren Werk als Ziel der Sequenzanalyse:

> „Im Unterschied zu deduktiven Forschungsansätzen sind die sprachlich-kommunikativen Handlungen in den Daten nicht vorab definierten Kategorien zuzuordnen. Ziel der Rekonstruktion ist vielmehr herauszuarbeiten, wie die Details des Handelns systematisch auf den sich entfaltenden Gesprächsprozess zugeschnitten sind und zu ihm in bestimmter sinnhafter Weise beitragen.“ (Deppermann 2014, S. 23)

Dies stellt in gewissem Sinn das Ziel der Analysen dar - zu sehen, wie durch die Interaktion der Teilnehmer in sinnhafter Weise zum Gesprächsprozess beigetragen wird. Dies geschieht zum Einen im Hinblick auf die Kommunikationsprozesse der Teilnehmer bezüglich der zu einem Randomisierungstest gehörigen Konzepte und zum anderen bezüglich des Einsatzes der Software TinkerPlots im Bearbeitungsprozess (Forschungsfragen 2 und 3).

Zusammenfassende Interpretation

Im letzten Schritt der Interaktionsanalyse von Krummheuer und Naujok (1999)werden alle wichtigen Ergebnisse zusammengefasst, deshalb wird in Kapitel 11 nur dieser Teil des Analyseprozesses ausführlich dargestellt, wie von Krummheuer und Naujok (1999) vorgeschlagen. Dabei werden einzelne Passagen der Schritte 3 und 4 soweit dargestellt, wie es für die Transparenz des Interpretationsprozesses nötig ist. Zur Nachvollziehbarkeit wird der jeweilige Transkriptausschnitt angegeben, der zu einer Interaktionseinheit bzw. Episode gehört, wodurch die „intersubjektive Überprüfbarkeit der Interpretationsleistung“ (Bohnsack 2014, 132) gewährleistet wird.

10.4.3 Rekonstruktion einzelner Argumentationen

Die interpretativen Analysen sollen an ausgewählten Stellen ergänzt werden durch die funktionale Argumentationsanalyse nach Toulmin (1975), um einzelnen Argumentationen der Studierenden detailliert nachgehen zu können und diese in ihre vorgebrachten Bestandteile zu gliedern, was durch eine reine Interpretation nicht möglich wäre. Dies dient nicht nur als Ergänzung, wie Krummheuer und Naujok (1999) das vorschlagen, sondern als notwendige Analyse, um die Argumentationsstrukturen der Teilnehmer analysieren zu können. Konkret soll herausgearbeitet werden, welche Regeln und Stützungen die Teilnehmer verwenden, um von einem Datum zu einer Konklusion zu gelangen. Dabei muss jedoch beachtet werden, dass in mündlich geführten Argumentationen vieles implizit bleibt. Nicht alle von Toulmin herausgearbeiteten funktionalen Bestandteile eines Arguments lassen sich in Äußerungen wiederfinden. Dies bedeutet jedoch nicht, dass das Analyseschema von Toulmin ungeeignet zur Auswertung der Argumentations-

prozesse ist, sondern dass es typisch für mündliche geführte Argumentationen ist, beispielsweise Schlussregeln nicht zu nennen, was auch Toulmin selbst einräumt:

> „Einer der Gründe dafür, zwischen Datum und Schlussregeln zu unterscheiden, ist folgender: Auf Daten wird explizit Bezug genommen, auf Schlussregeln implizit.“ (Toulmin 1975, 91)

Diese impliziten Teile einer Argumentation zu rekonstruieren ist ein Bestandteil dieser Arbeit.

10.4.4 Analyse der TinkerPlots-Benutzung

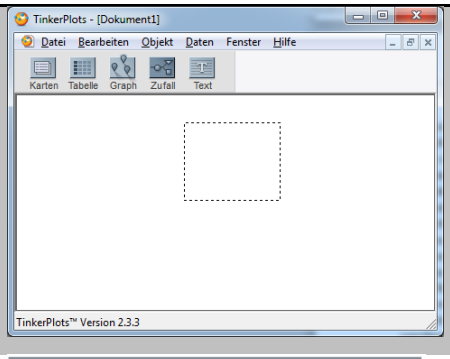
Es soll der Frage nachgegangen werden, wie die Studierenden TinkerPlots in ihren Lösungsprozess einsetzen und an welchen Stellen im Bearbeitungsprozess sie Unterstützung bei der Bedienung der Software benötigen (Unterfroschungsfragen zu Forschungsfrage 3). Hierfür wurden aus der Aufgabenanalyse (Kapitel 9.2.2) deduktiv Schritte abgeleitet, die während des Prozesses nötig sind, um einen Randomisierungstest mit TinkerPlots durchzuführen. Hierbei wurde sehr kleinschrittig vorgegangen, um in den späteren Analysen genau feststellen zu können, an welchen Stellen Studierende, die den Baustein Inferenzstatistik absolviert haben, Unterstützung benötigen und so auch gezielt Hinweise für ein Redesign der Lerneinheit geben zu können (Forschungsfrage 4).

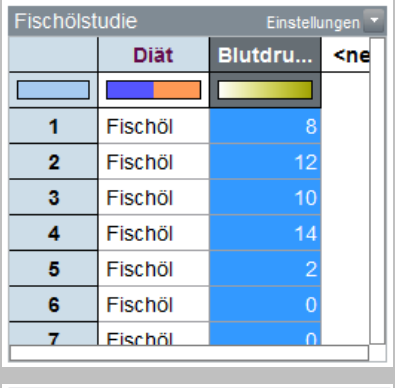
Die Arbeit mit der Software kann in der Interaktionseinheit 6 einsetzen, in welcher die beobachteten Daten des Experiments analysiert werden. Da allerdings die Daten graphisch als Punktdiagramm und als Boxplot mit eingeblendeten arithmetischen Mittelwerten und eingezeichneter Differenz zwischen den arithmetischen Mitteln vorliegen, wird vermutlich auch nur mit diesen Darstellungen gearbeitet und das Analysieren der Daten geschieht ohne Software. Unumgänglich ist der Einsatz der Software TinkerPlots in den Interaktionseinheiten 8, 9 und 10. Um die in der Nullhypothese ausgedrückte Situation zu simulieren, muss in Interaktionseinheit 8 TinkerPlots eingesetzt werden. Die Teststatistik und die Referenzverteilung werden in Interaktionseinheit 9 ebenfalls mit TinkerPlots erzeugt und schließlich geschieht noch das Ermitteln des P-Werts in Interaktionseinheit 10 mit TinkerPlots. Für alle anderen Interaktionseinheiten ist das Arbeiten mit der Software nicht nötig. Für die Interaktionseinheiten 8, 9 und 10 werden nun genau die Schritte aufgezeigt, die durchgeführt werden müssen, um die Simulation erfolgreich durchzuführen.

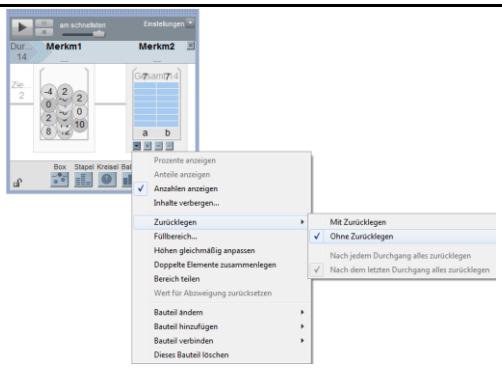
Das Übertragen der Nullhypothese in eine TinkerPlots-Simulation ist Bestandteil der Interaktionseinheit 8. Abgeleitet aus der Aufgabenanalyse (Kapitel 9.2.2) müssen folgende Schritte (Tabelle 25) durchgeführt werden. Dabei kann die Reihenfolge der Schritte variieren, beispielsweise kann die Durchgangsanzahl eingestellt werden, bevor die Bauteile gefüllt werden oder umgekehrt. Der gesamte Vorgang ist bezüglich der in der Software notwendig durchzuführenden Schritte operationalisiert, um eine vollständige Aufstellung zu erhalten und genau analysieren zu können, an welchen Stellen Un-

terstützungsbedarf herrscht. In der ersten Spalte von Tabelle 25 wird der Schritt beschrieben, der nötig ist. In der zweiten Spalte werden die zugehörigen Aktivitäten auf Softwareebene beschrieben und in der dritten Spalte graphisch repräsentiert. Die hier vorgestellte Abfolge von Schritten stellt eine prototypische Lösung dar, Varianten davon sind möglich, ohne dass jedoch Schritte ausgelassen werden können.

Tabelle 25: Notwendige Schritte in TinkerPlots zur Erstellung der Zufallsmaschine in Interaktionseinheit 8

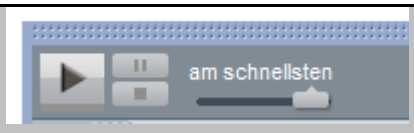
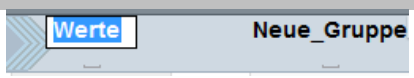
Schritt	Beschreibung der Aktivität	Graphische Darstellung
Erstellen der Zufallsmaschine	Anklicken des Icons Zufall und Ziehen in die Arbeitsfläche	
Vorhandene Kugeln der Zufallsmaschine löschen	Löschen aller vorhandenen Kugeln aus dem Bauteil Box durch Anklicken des Minus-Symbols	

Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren	Das Merkmal Blutdruckreduktion auswählen, mit STRG+C kopieren und mit STRG+V in das leere Bauteil der Zufallsmaschine einfügen Oder durch das Plus-Symbol 14 Kugeln hinzufügen und die Werte von Hand eintragen	
Ein zweites Bauteil der Zufallsmaschine hinzufügen	Ein zweites Bauteil (z.B. Stapel) auswählen und in Reihe (unabhängig vom ersten Bauteil) der Zufallsmaschine hinzufügen	
Zweites Bauteil füllen	Dem zweiten Bauteil die Gruppenetiketten hinzufügen, z. B. im Stapelbauteil zwei Stapel a und b erstellen und mit jeweils sieben Elementen füllen	

Die Einstellung ohne Zurücklegen bei beiden Bauteilen auswählen	In beiden Bauteilen die Einstellung ohne Zurücklegen über das Pfeilmenü auswählen	
Durchgangsanzahl festlegen	Für die Durchgangsanzahl den Wert 14 eintragen	

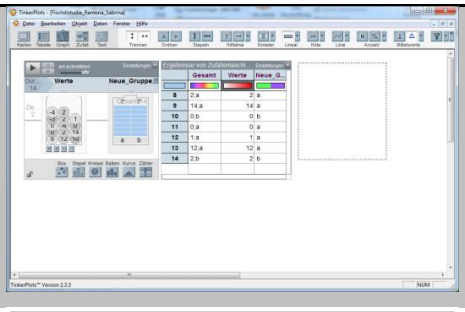
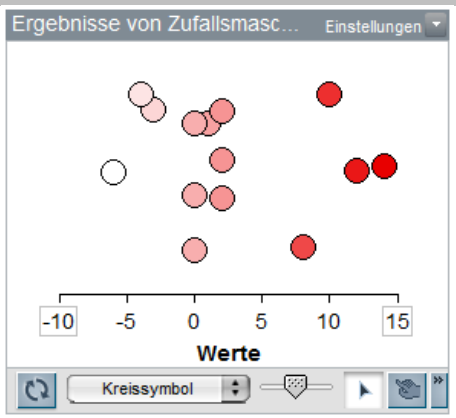
Optional sollte zusätzlich der Schritt „Einstellen der Geschwindigkeit der Zufallsmaschine“ durchgeführt werden (Tabelle 26), um später die Simulation zu beschleunigen. Dies ist jedoch kein notwendiger Schritt, um die Simulation erfolgreich auszuführen. Weiterhin können die von TinkerPlots automatisch benannten Merkmale *Merkm1* und *Merkm2* umbenannt werden (Tabelle 26), auch dies ist jedoch kein notwendiger Schritt.

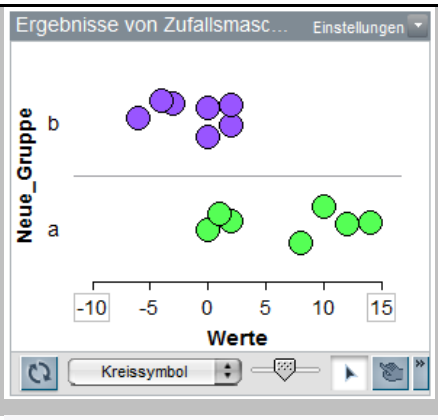
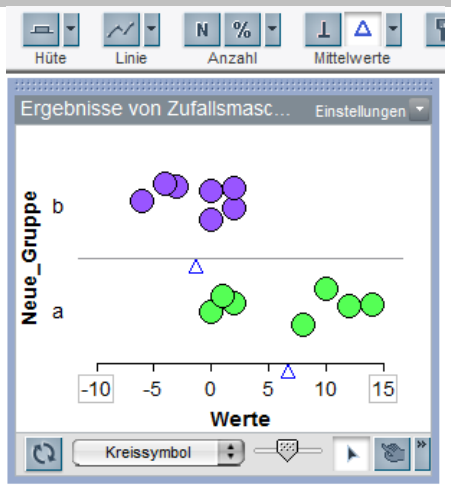
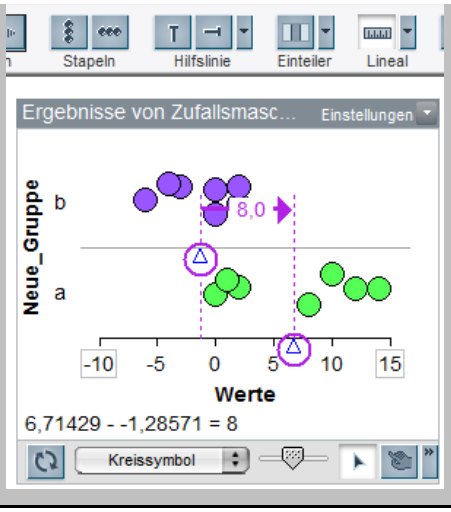
Tabelle 26: Optionale Schritte in TinkerPlots zur Erstellung der Zufallsmaschine in Interaktionseinheit 8

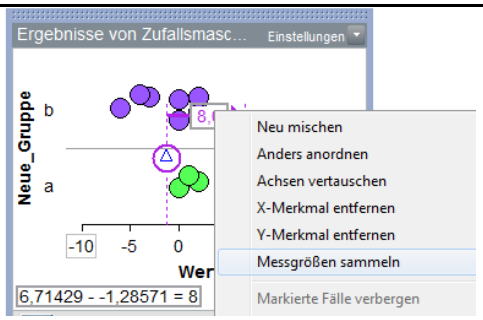
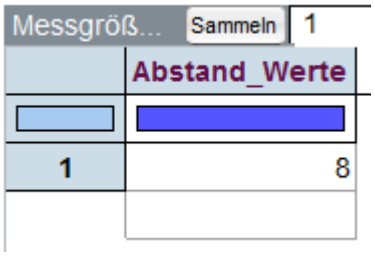
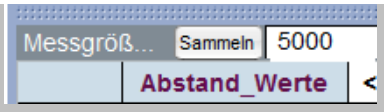
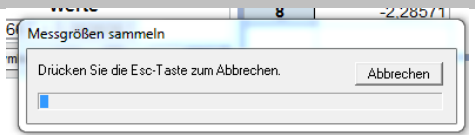
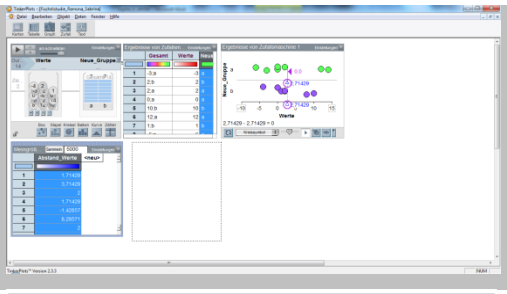
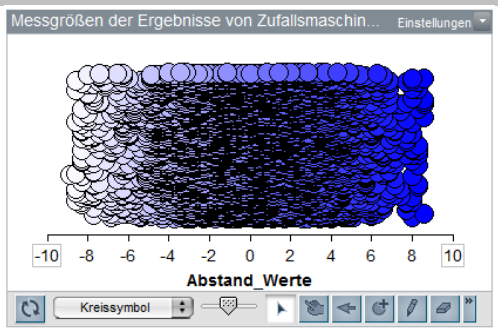
Schritt	Beschreibung der Aktivität	Graphische Darstellung
Einstellen der Geschwindigkeit	Einstellen der Geschwindigkeit *am schnellsten* in der Zufallsmaschine	
Umbenennen der Merkmale *Merkm1* und *Merkm2*	Das Merkmal *Merkm1* zum Beispiel in „Werte“ umbenennen und das Merkmal *Merkm2* zum Beispiel in „Neue_Gruppe“ umbenennen	

Die Teststatistik wird in Interaktionseinheit 9 definiert und dort wird auch die Referenzverteilung erstellt. Dazu ist es zunächst notwendig, die in einer Tabelle protokollierten Ergebnisse der Simulation aus Interaktionseinheit 8 in einer Graphik so darzustellen, dass die Teststatistik (Differenz der Mittelwerte der beiden neuen Gruppen) definiert werden kann. Diese muss in TinkerPlots als Messgröße gesammelt werden und die gesammelten Messgrößen werden in einer Graphik als Referenzverteilung dargestellt. Die hierzu notwendigen Schritte finden sich in Tabelle 27. Auch hierbei gilt wieder, dass die Reihenfolge der Schritte verändert werden kann, so ist es zum Beispiel gleich, ob zuerst im Graphen die beiden Gruppen dargestellt werden, oder ob zuerst die Werte komplett separiert werden. Voraussetzung hierfür sind die korrekt durchgeführten Schritte in Interaktionseinheit 8.

Tabelle 27: Notwendige Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung in Interaktionseinheit 9

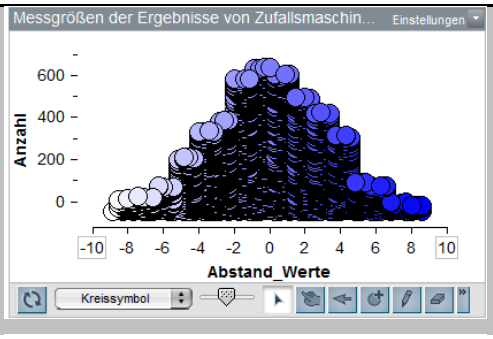
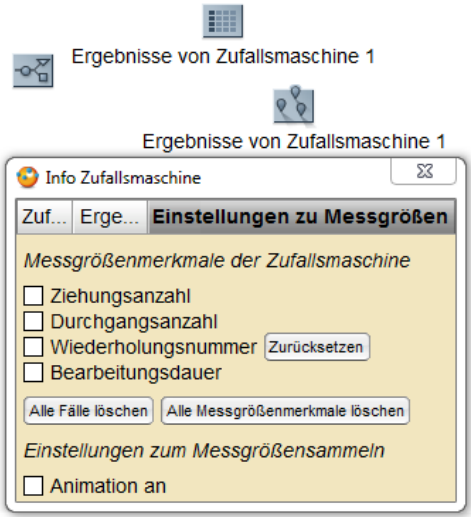
Schritt	Beschreibung der Aktivität	Graphische Darstellung
Erstellen einer Graphik zur Simulation	Anklicken des Icons Graph und Ziehen in die Arbeitsfläche	
Das erste Merkmal auswählen und darstellen	Ein Merkmal (z. B. *Merkm1* bzw. *Werte*) auswählen und komplett separieren	

Das zweite Merkmal auswählen und in der Graphik passend darstellen	Zweites Merkmal (*Merkm2* bzw. *Neue_Gruppe*) auswählen und auf der zweiten Achse darstellen	
Die Mittelwerte einblenden	Das Icon *arithmetisches Mittel* anklicken. Die Mittelwerte werden automatisch für beide Gruppen eingeblendet	
Differenz der Mittelwerte mit Hilfe der Linealfunktion messen	Das Icon Lineal anklicken und die beiden Enden des Lineals über die arithmetischen Mittel ziehen	

Messgröße definieren	Den gemessenen Wert des Lineals anklicken und per Rechtsklick als Messgröße festlegen. ODER den Button M in der Graphik anklicken und anschließend den gemessenen Wert des Lineals anklicken. Die Messgröße wird durch beide Aktionen automatisch definiert und in einer automatisch erzeugten Tabelle protokolliert	 
Anzahl der zu sammelnden Messgrößen eintragen	In die Tabelle der Messgrößen die Anzahl der zu sammelnden Messgrößen eintragen, z. B. 5000	
Messgrößen sammeln	Durch Anklicken von Sammeln in der Tabelle der Messgrößen den Sammelvorgang starten	
Erstellen einer Graphik zu den gesammelten Messgrößen	Anklicken des Icons Graph und Ziehen in die Arbeitsfläche	
Referenzverteilung darstellen	Auswählen des Merkmals *Abstand_Werte* zur Darstellung in der Graphik. Komplettes Separieren der Fälle	

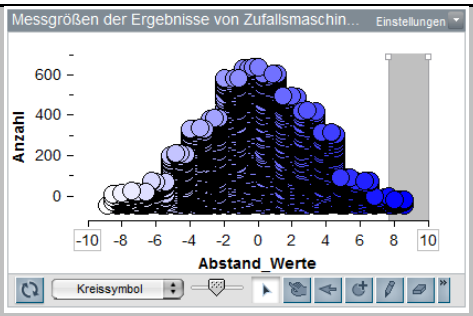
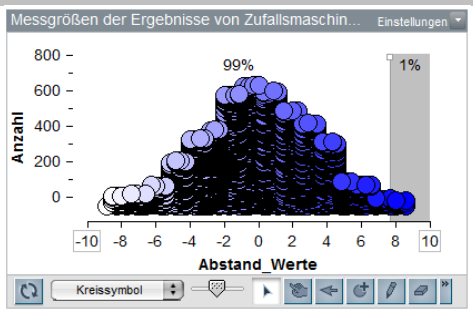
Hinzu kommen zwei Schritte (Tabelle 28), die zwar für die erfolgreiche Bearbeitung nicht notwendig, aber sehr sinnvoll sind. Das ist erstens das Stapeln der Fälle in der Referenzverteilung und zweitens das Ausstellen der Animation während des Messgrößensammelns.

Tabelle 28: Optionale Schritte in Interaktionseinheit 9

Schritt	Beschreibung der Aktivität	Graphische Darstellung
Fälle in der Graphik stapeln	In der Graphik der simulierten Ergebnisse die Fälle stapeln	
Beschleunigen des Messgrößensammelns	In der Tabelle der Messgrößen über das Einstellungsmenü die Animation zu den Messgrößen abwählen UND alle zur Simulation gehörigen Objekte (Zufallsmaschine, Tabelle der simulierten Ergebnisse, Graphik der simulierten Ergebnisse) minimieren	

Das Ende der Arbeit mit TinkerPlots besteht darin, den P-Wert aus der Referenzverteilung der gesammelten Messgrößen zu schätzen. Dies geschieht in Interaktionseinheit 10. Hierzu wird der Einteiler benötigt, von dem das linke Ende des grauen Bereichs auf den Wert 7,7 gelegt werden muss, um diesen Wert mit in den grauen Bereich aufzunehmen. Einblenden der Prozente ermöglicht die Schätzung des P-Werts aus der Simulation. Die hierzu notwendigen Schritte finden sich in Tabelle 29. Hierzu gibt es Alternativen, z. B. das Bilden passender Klassen, die jedoch im Baustein Inferenzstatistik nicht thematisiert wurden. Deshalb wird auch nicht davon ausgegangen, dass diese von den Teilnehmern der Studie verwendet werden. Voraussetzung sind die korrekt durchgeführten Schritte der Interaktionseinheit 9.

Tabelle 29: Notwendige Schritte in TinkerPlots zur Ermittlung des P-Werts in Interaktionseinheit 10

Schritt	Beschreibung der Aktivität	Graphische Darstellung
Einteiler verwenden	Anklicken des Icons *Einteiler* und Verschieben des rechten Endes des Einteilers bis an das rechte Ende der Verteilung und Verschieben des linken Endes des Einteilers bis zum Wert 7,7, beispielsweise durch Doppelklicken des linken Endes des Einteilers und Eingeben des Wertes 7,7	
Prozente einblenden	Anklicken des Icons *Prozent*	

Somit ergeben sich aus Tabelle 25, Tabelle 27 und Tabelle 29 neunzehn notwendige und aus Tabelle 26 und Tabelle 28 vier optionale Schritte, um den Randomisierungstest mit TinkerPlots durchzuführen. Bezüglich dieser Schritte soll analysiert werden, ob die Teilnehmer diese selbständig durchführen können oder ob sie Hilfe bei der Umsetzung der einzelnen Schritte in TinkerPlots benötigen. Da die Teilnehmer den Randomisierungstest unter Verwendung der Software TinkerPlots lediglich einmal unter Anleitung im Seminar „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ durchgeführt haben (vgl. Kapitel 8.2.2.2), wird davon ausgegangen, dass nicht alle benötigten Schritte fehlerfrei und selbständig verwendet werden (vgl. hierzu *mögliche Schwierigkeiten* in der a priori Analyse der Aufgabe in Kapitel 9.2.2).

Biehler, Frischemeier und Podworny (2017) haben ein sehr fein ausdifferenziertes Interventionssystem entwickelt, um Studierende in einer Interviewsituation, in welcher mit der Software TinkerPlots simuliert werden sollte, auf verschiedenen Ebenen (angelehnt an Leiss 2007) zu unterstützen. Ausgewertet wurde jedoch nur, ob für einen bestimmten in der Software umzusetzenden Schritt eine Intervention benötigt wurde oder nicht. Aus den Erfahrungen dieser Analysen wurden für die hier vorliegende Studie be-

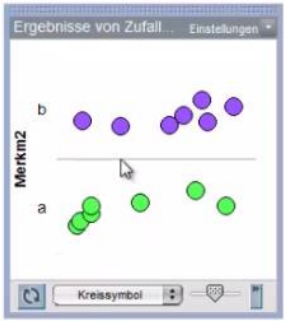
nötigte Interventionen für die Arbeit mit TinkerPlots auf drei Ebenen unterschieden. Ziel dieser Analyse ist ein genauer Blick entsprechend Forschungsfrage 3 in den Bearbeitungsprozess der Studierenden, um gezielt herauszufinden, welche Schritte in TinkerPlots selbständig umgesetzt werden können und für welche Schritte Hilfen benötigt werden. Es soll keine Häufigkeitsauswertung im Sinne beispielsweise der qualitativen Inhaltsanalyse nach Mayring (2010) geschehen, sondern lediglich abgegrenzt werden, bei welchen Schritten Hilfe benötigt wird. Es wird unterschieden in „keine Intervention wird benötigt“ (kurz *keine Intervention*), „leichte Intervention wird benötigt“ (kurz *leichte Intervention*) und „starke Intervention wird benötigt“ (kurz *starke Intervention*).

Die Kategorie „keine Intervention“ wird vergeben, wenn die Studierenden den entsprechenden Schritt selbständig durchführen und weder gefragt noch ungefragt eine Intervention durch die Interviewerin benötigt wird. Ein Beispiel hierfür ist der Beginn der Episode 9-1 von Fabia und Laura, in der ohne Intervention durch die Interviewerin eine Graphik erstellt wird, um die simulierten Ergebnisse darzustellen.

173 F: Jetzt hat man ja zumindest schon mal einmal randomisiert. (In der Ergebnistabelle wird hoch und runtergescrollt). Das ist schon so lange her mit TinkerPlots. (...) Und jetzt können wir das ja theoretisch in die zwei Gruppen einteilen. (...)

174 L: Da eine Graphik jetzt erst mal machen und da konnte man ja äh (...) irgendwo auf Ziehung nochmal machen.

175 *Ein Graph wird erstellt, die Fälle werden nach Merkm2 eingefärbt und werden getrennt.*



Die Kategorie „leichte Intervention“ wird vergeben, wenn die Studierenden einen Impuls durch die Interviewerin benötigen, um den entsprechenden Schritt durchzuführen, ohne dass jedoch explizit von der Interviewerin genannt werden muss, welche Aktionen für den jeweiligen Schritt auszuführen sind. Ein Ankerbeispiel dafür findet sich zu Beginn der Episode 9-7 von Mandy und Alisa, in welcher sie die Referenzverteilung darstellen möchten und hierzu eine Graphik benötigen. Mandy wendet sich an die Interviewerin (325) und provoziert so eine leichte Intervention von dieser (326), ohne dass jedoch eine konkrete Anweisung gegeben werden muss.

325 M: Und jetzt? (...) Jetzt bin ich verwirrt. Jetzt können wir die Punkte doch eigentlich auch noch mal in einen Plot ziehen, oder?

326 I: Probier doch.

327 A: Also einen Graph, ne?

328 M: Schön. Ja.
329 *Die Studentin öffnet einen neuen Graph.*

Die Kategorie „starke Intervention“ wird vergeben, wenn die Studierenden nicht wissen, wie ein Schritt durchgeführt werden muss, um weiterarbeiten zu können, wenn also genaue Anweisungen durch die Interviewerin benötigt werden. Ein Ankerbeispiel hierzu findet sich in Episode 8-3 von Rebecca und Selina, in welcher sie die Werte des ursprünglichen Experiments in ein Bauteil der Zufallsmaschine kopieren möchten. Zuvor haben Rebecca und Selina das Vorgehen erklärt, nun benötigen sie Hilfe bei der Umsetzung in TinkerPlots. In (109) wird eine starke Intervention von der Interviewerin gegeben in Form einer Anweisung, um die Zufallsmaschine für den Kopiervorgang vorzubereiten.

107 I: (.) Genau. Und was ihr mir gerade erklärt habt ist, ihr wollt einmal diese ganzen Werte in die Urne legen, wo zufällig rausgezogen wird, ist das richtig?
108 R&S: Genau, ja.
109 I: Gut, dann löscht ihr jetzt als erstes alle Kugeln raus. Das geht mit dem Minus.
110 *Die drei Kugeln werden gelöscht.*

10.5 Gütekriterien qualitativer Forschung

Ein wichtiger Aspekt qualitativer Forschung ist die Einschätzung der Ergebnisse anhand von Gütekriterien. Es können jedoch nicht einfach die etablierten Maßstäbe der quantitativen Forschung übernommen werden (Steinke 2013; Mayring 2016). Die Kriterien müssen die jeweilige Fragestellung, Methode und die stark eingeschränkte Standardisierbarkeit methodischer Vorgehensweisen berücksichtigen (Steinke 2013, 323). Deshalb schlägt die Autorin vor, einen breit angelegten Kriterienkatalog zu formulieren, der

eine Orientierung für qualitative Forschung bieten kann und entsprechend der Fragestellung und verwendeten Methode angepasst werden sollte (Steinke 2013, 324). Ein solcher Kriterienkatalog findet sich z. B. bei Mayring (2016), der sechs allgemeingültige Gütekriterien für qualitative Forschung aufstellt. Die sechs Kriterien von Mayring (2016) sind *Verfahrensdokumentation*, *argumentative Interpretationsabsicherung*, *Regelgeleitetheit*, *Nähe zum Gegenstand*, *kommunikative Validierung* und *Triangulation*.

Unter **Verfahrensdokumentation** versteht der Autor eine genaue, bis ins Detail gehende, Dokumentation, „um den Forschungsprozess für andere nachvollziehbar werden zu lassen" (Mayring 2016, 145). Hier werden vor allem die Explikation des Vorverständnisses, die Zusammenstellung des Analyseinstruments und die Durchführung und Auswertung der Datenerhebung genannt. Da Interpretationen eine wichtige Rolle in qualitativer Forschung spielen, aber nicht bewiesen oder nachgerechnet werden können, müssen Interpretationen argumentativ begründet werden. Eine Interpretation soll durch das Vorverständnis theoriegeleitet sein und in sich schlüssig. Das Suchen nach Alternativdeutungen ist hierbei ein wichtiger Aspekt.

Verfahrensdokumentation in dieser Arbeit

Der Forschungsprozess wird so detailreich wie möglich in dieser Arbeit dargestellt. Das Vorverständnis der Teilnehmer ist in Kapitel 8.3.2 dargestellt, die Analysemethoden sind in diesem Kapitel expliziert und die Auswertung der Daten geschieht im nächsten Kapitel. Die in Kapitel 11 ausgeführten Interpretationen werden entlang der jeweiligen Transkriptausschnitte getätigt und sind somit für den Leser nachvollziehbar und kritisierbar. Beim Interpretieren liegt das in Kapitel 6 aufgebaute Theorieverständnis zugrunde.

Qualitative Forschung soll zwar offen sein und ggf. vorgeplante Analyseschritte ändern, soll sich aber trotzdem an bestimmte Verfahrensregeln halten, beispielsweise durch das schrittweise, sequentielle Vorgehen beim Interpretieren, wodurch ein systematisches Vorgehen geschaffen wird (Mayring 2016, 146). Dies bestimmt die **Regelgeleitetheit** als Kriterium.

Regelgeleitetheit in dieser Arbeit

In Kapitel 10.4.1 sind die geplanten Analyseschritte ausführlich dargelegt und die Interaktionsanalyse nach Krummheuer und Naujok (1999) ebenso wie die Anpassung für die vorliegende Arbeit sind vorgestellt. Ein wesentliches Element der Interpretationen in Kapitel 11 ist das sequentielle Vorgehen, das das systematische Vorgehen zeigt.

Die **Nähe zum Gegenstand** als Kriterium zeichnet sich dadurch aus, dass versucht wird, eine Interessenübereinstimmung mit den Beforschten herzustellen und ein offenes, gleichberechtigtes Verhältnis angestrebt wird (Mayring 2016, 146).

Nähe zum Gegenstand in dieser Arbeit

Bei den Teilnehmern der Fallstudien handelt es sich um Studierende, die am Baustein Inferenzstatistik teilgenommen haben, der von der Autorin dieser Arbeit unterrichtet wurde. Als Ziel der Fallstudie wurde zu Beginn der Interviews deutlich gemacht, dass ein Wissenszuwachs auf beiden Seiten (Interviewerin und Interviewteilnehmer) angestrebt wird. Ein offenes Verhältnis wurde ebenfalls explizit angestrebt durch die Mitteilung zu Beginn, dass das Interview in Dialogform geführt werden soll und jederzeit Fragen durch die Interviewten zugelassen sind.

Die **kommunikative Validierung** soll im Wesentlichen dadurch geschehen, dass die Interpretationen den Beforschten vorgelegt und mit ihnen diskutiert werden. Allerdings darf dies nicht das ausschließliche Kriterium sein, da die Analyse sonst bei den subjektiven Aussagen der Beforschten stehen bliebe (Mayring 2016, 147). Vielmehr soll qualitative Forschung darüber hinausgehen.

Kommunikative Validierung in dieser Arbeit

Durch die besondere Form des Interviews in Form eines Leitfadeninterviews geschieht die kommunikative Validierung mit den Beforschten bereits im dritten Teil des Interviews, in welchem konkrete Nachfragen und Interpretationen der Interviewerin über das vorher Geschehene und Gesagte gemeinsam diskutiert werden. Darüber hinaus sind in einem weiteren Schritt alle Interpretationen aus Kapitel 11 mit einem unabhängigen Forscher durchdiskutiert worden, um einseitige bzw. subjektive Interpretationen zu öffnen und zu ergänzen. Zu ausgewählten Interpretationen fanden weitere Diskussionen mit Mitgliedern der AG Biehler statt.

Schließlich bildet die **Triangulation** das letzte Kriterium, durch das mehrere Analysegänge verbunden werden sollen. Dies kann auf unterschiedlichen Ebenen stattfinden, zum Beispiel durch verschiedene Datenquellen oder unterschiedliche Interpreten, um Ergebnisse vergleichen zu können. Wobei dabei keine völlige Übereinstimmung erreicht werden soll, sondern Stärken und Schwächen verschiedener Analysewege sollen aufgezeigt werden (Mayring 2016, 148).

Triangulation in dieser Arbeit

Wie bereits im Punkt vorher erwähnt, wurden die Interpretationen verschiedener Interpreten miteinander verglichen und abgestimmt, um getätigte Interpretationen zu bestätigen, um weiterführende Interpretationen zu finden, oder in seltenen Fällen, um getätigte Interpretationen zu verwerfen. Triangulation nach Flick (2016) wird auch erreicht durch die Kombination mindestens zweier sich im Analyseprozess ergänzender Methoden, was hier durch die Interaktionsanalyse nach Krummheuer und Naujok (1999) und durch die funktionale Argumentationsanalyse von Toulmin (1975) geschieht. Der Vergleich

dreier Einzelfallstudien miteinander stellt den letzten Punkt dar, wie Triangulation in der vorliegenden Arbeit erreicht wird.

11 Auswertung der explorativen Fallstudien „Randomisierungstest mit TinkerPlots“

In diesem Kapitel wird die Analyse der drei Paare dargelegt. Dabei wird das Vorgehen der Paare hinsichtlich der Forschungsfragen aus Kapitel 9.1 „Inwieweit treten die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen auf? Welche unerwarteten Problemstellen können identifiziert werden?“, „Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots können festgestellt werden?“ und „Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?“ nacheinander analysiert und abschließend vergleichend betrachtet. Als Leitfragen für die Auswertungen dienen die Unterforschungsfragen (Kapitel 10.3):

- Welche Problemstellen treten auf? Welche davon sind unerwartet? Welche Gründe lassen sich identifizieren für die unerwarteten Problemstellungen?
- Wie stark sind die Phasen der Gesprächsführung ausgeprägt? Wie gestalten die Paare die Kommunikation untereinander? Was bedeutet das in Bezug auf die kollaborativen Bearbeitungsprozesse?
- Arbeiten die Teilnehmer der Studie kollaborativ zusammen? An welchen Merkmalen lässt sich die Kollaboration erkennen?
- Können die Teilnehmer die verschiedenen, in Kapitel 6.4 identifizierten Konzepte (einen Unterschied in den beobachteten Daten identifizieren, Nullhypothese aufstellen, etc.), die zur Argumentation eines Randomisierungstests gehören, verwenden und erläutern? Welche Schwierigkeiten treten dabei auf?
- In welcher Weise gelingt es den Teilnehmern, die drei Ebenen Kontext, Statistik und Software aufeinander zu beziehen?
- Können die Teilnehmer die Software TinkerPlots als Instrument einsetzen, um einen Randomisierungstest durchzuführen? An welchen Stellen bei der Benutzung der Software TinkerPlots benötigen die Studierenden Hilfe?

Das Kapitel beginnt mit einem zeitlichen Überblick über den Bearbeitungsprozess der drei Paare, um das Vorgehen der Paare insgesamt zu beschreiben. Die anschließenden Fallstudien sind alle in gleicher Reihenfolge aufgebaut. Entlang der in Kapitel 10.4.1 identifizierten Interaktionseinheiten werden die zugehörigen Transkriptausschnitte präsentiert und die zusammenfassenden Interpretationen dargestellt. Die ausführliche Analyse der Einzeläußerungen und Interpretationsalternativen sowie die Turn-by-Turn Analyse werden hier in dem Umfang berichtet, wie sie für die zusammenfassende Interpre-

tation im Hinblick auf die Forschungsfragen nötig sind, da eine noch ausführlichere Darstellung den Rahmen dieser Arbeit sprengen würde. Für die zusammenfassende Interpretation der jeweiligen Interaktionseinheiten wird Bezug genommen auf die Kontextebene, die Statistikebene und die Softwareebene, auf denen die Kommunikation stattfindet. An einigen ausgewählten Stellen wird die funktionale Argumentationsanalyse nach Toulmin (1975) ergänzend verwendet, um die Aussagen der Probanden zu strukturieren. Bezüglich des Umgangs mit TinkerPlots wird am Ende jeder Fallstudie in einem tabellarischen Überblick berichtet, an welchen Stellen die Paare die Software selbständig bedient haben und an welchen Stellen sie Hilfestellungen durch die Interviewerin benötigten, um einen Einblick zu bekommen, wie die Studierenden TinkerPlots einsetzen (Forschungsfrage 2). Im Anschluss an jede Fallstudie findet sich eine knappe Fallzusammenfassung, in der besondere Auffälligkeiten des Paares berichtet werden. Im Anschluss an die Einzelfallstudien werden die Bearbeitungsprozesse der drei Paare vergleichend betrachtet (Kapitel 11.5) und die Unterforschungsfragen werden beantwortet.

Wenn in den zusammenfassenden Interpretationen der durchgeführten Interaktionsanalyse wörtliche Rede der Sprechenden zitiert wird, werden nicht immer ganze Turns zitiert, sondern nur die relevanten Teile. Wenn sprachliche Einheiten eines Turns in einem Zitat ausgelassen wurden, so wird dies durch „…“ in einem Zitat angezeigt.[47]

11.1 Überblick über den zeitlichen Bearbeitungsprozess aller Paare

Bevor nun die Analysen der einzelnen Paare detailliert vorgestellt werden, wird ein zeitlicher Überblick über den Bearbeitungsprozess gegeben. In Abbildung 199 ist die Dauer der 15 Interaktionseinheiten für jedes Paar abgebildet. Zur Erinnerung: Die ersten vier Interaktionseinheiten sind gegliedert entlang des ersten Teils des Leitfadeninterviews (N0 bis N2, Kapitel 9.3) und beinhalten Dialoge zwischen den Teilnehmerinnen und der Interviewerin. Ab der fünften Interaktionseinheit beginnt das selbständige Bearbeiten der gestellten Aufgabe durch die Teilnehmerinnen, bis hin zur elften Interaktionseinheit. Die letzten vier Interaktionseinheiten sind nun wieder in einer Interviewsituation angesiedelt, in denen die Interviewerin die Fragen A1 bis A9 entlang des Leitfadens (Kapitel 9.3) stellt.

[47] Es ist zu beachten, dass Punkte in Klammern entsprechend den Transkriptionsregeln Pausen kürzerer Dauer (.) oder längerer Dauer (…) anzeigen. Diese sind nicht zu verwechseln mit den drei Punkten für das Anzeigen von Auslassungen. Sprachliche Ergänzungen werden in eckigen Klammern vorgenommen, zum Beispiel „wie haben wir das [gemacht]“.

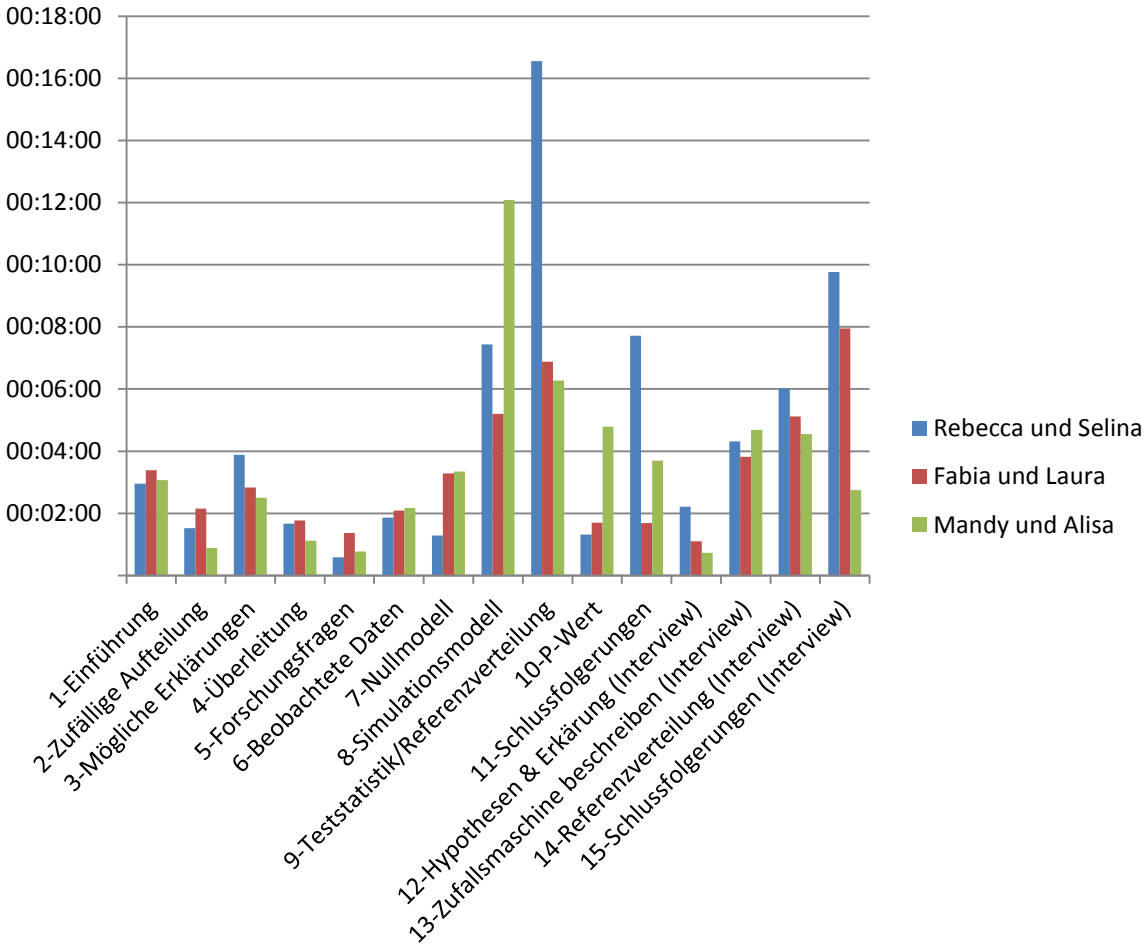


Abbildung 199: Zeitlicher Überblick über den Bearbeitungsprozess aller drei Paare

Bezüglich der ersten sieben Interaktionseinheiten ähneln sich die Paare im Hinblick auf die Zeit, in der sie sich mit dem jeweiligen Inhalt auseinandersetzen. In der selbständigen Bearbeitung des Randomisierungstests in den Einheiten 8-11 lassen sich stärkere Unterschiede in den Bearbeitungszeiten ausmachen. Gründe hierfür werden durch die Interaktionsanalyse aufgedeckt und in dem abschließenden Vergleich der Paare aufgearbeitet. Die deutlichsten Unterschiede lassen sich in den drei Interaktionseinheiten ausmachen, in denen die Simulation in TinkerPlots durchgeführt und ausgewertet wird (Interaktionseinheiten 8 bis 11). Gründe hierzu werden ebenfalls in der folgenden Analyse der Paare herausgearbeitet und abschließend verglichen. Die Interaktionseinheiten 12 bis 15, die wieder in einer Interviewsituation angesiedelt sind, ähneln sich von der Dauer her wieder stärker, abgesehen von der letzten Interaktionseinheit, in der den Schlussfolgerungen in sehr unterschiedlicher Dauer noch einmal nachgegangen wird.

11.2 Fallstudie Rebecca und Selina

In diesem Abschnitt werden die zusammenfassenden Interpretationen der 15 Interaktionseinheiten von Rebecca und Selina geliefert. Die zugehörigen Transkriptausschnitte werden den Interpretationen vorangestellt, um die Interpretation für den Leser nachvollziehbar zu machen. Einige Einheiten (z. B. die erste) werden ohne vorangestellten Transkriptausschnitt zusammengefasst, da in diesen Einheiten keine inhaltlichen Aspekte diskutiert werden. Lange Interaktionseinheiten, wie beispielsweise die achte, werden in thematisch orientierte Episoden untergliedert, dies wird zu Beginn jeder dieser Einheiten erläutert.

11.2.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Rebecca und Selina

Interaktionseinheit 1: Vorstellung und Vorlesen des Aufgabenblatts (Z. 1-12)

In der ersten Einheit geschieht wenig Interaktion. Zunächst wird geklärt, dass die beiden Teilnehmerinnen noch nicht zusammen gearbeitet haben. Anschließend wird von der Interviewerin das gesamte Aufgabenblatt vorgelesen. Die sich anschließende Frage an die Studierenden, ob es noch Fragen gäbe, wird von beiden verneint. Wie in der a priori Analyse der Aufgabe (Kapitel 9.2.2) erwartet, artikulieren die Teilnehmerinnen keine Schwierigkeiten zum Textverständnis, es wird davon ausgegangen, dass sie den Text verstanden haben. Somit wird in der ersten Interaktionseinheit eine Basis für die weitere inhaltliche Arbeit gelegt.

Interaktionseinheit 2: Begründung für die zufällige Aufteilung auf die Gruppen (Z. 13-20)

In der zweiten Interaktionseinheit wird der ersten Frage N1 des Interviewleitfadens nachgegangen. Die inhaltlich relevante Passage dazu findet sich im folgenden Transkriptausschnitt.

13	I: Gut. Ähm wir sehen jetzt hier oben in dieser Darstellung, dass wir einen Unterschied haben zwischen den beiden Mittelwerten. Da wurden die arithmetischen Mittel berechnet und ein Unterschied von 7,7. Und jetzt kommt meine erste Frage, bevor wir uns das näher in TinkerPlots angucken: Wieso wurden die Versuchspersonen zufällig auf die zwei Gruppen aufgeteilt?
14	R: Soll ich mal versuchen?
15	S: Ja, dann ergänze ich.
16	R: Ähm also ich hätte jetzt gesagt, dass man eben durch die zufällige Zuordnung versucht irgendwelche ganz persönlichen, ja personengebundenen Eigenschaften irgendwie auch wirklich zufällig auf die Gruppen zu verteilen. Ich weiß jetzt nicht, ob bei so einer Diät vielleicht so was wie das Alter eine Rolle spielt oder vielleicht auch Gewicht und dass man eben schaut, wenn man das zufällig einteilt, dass sich dann das ungefähr eben gleich auf beide Gruppen aufteilt und nicht in einer Gruppe eben ähm ganz viele nur ältere Menschen drin sind und in der anderen nur eben ganz viele nur Jüngere. Dass das so ganz ungleich in sich aufgeteilt wäre
17	S: Ja also genau das würde ich auch sagen, weil ansonsten, wenn man das nicht zufällig macht, könnte man so eine Studie oder so ein Ergebnis ja auch indirekt beeinflussen, indem ich eben in die eine Gruppe zum Beispiel nur junge Probanden und in die andere nur Ältere. Und dann hab

ich natürlich ein gefälschtes Ergebnis.

Direkt im Anschluss an die Frage der Interviewerin (13) wieso die Versuchspersonen zufällig auf die Gruppen eingeteilt wurden, wendet sich Rebecca an ihre Partnerin und fragt „Soll ich mal versuchen?“ (14). Hierdurch drückt sie sprachlich aus, dass sie einen Ansatz hat, um die Frage zu beantworten, gleichzeitig jedoch nicht sicher ist, ob das, was folgt, richtig sein wird. Selina nimmt dies auf, denn sie antwortet mit „Ja, dann ergänze ich.“ (15). Sie nimmt also wahr, dass die nun folgende Erklärung von Rebecca eventuell unvollständig sein wird und drückt gleichzeitig aus, dass sie sich für kompetent genug hält, etwaige Unstimmigkeiten oder Leerstellen zu ergänzen (oder auch zu korrigieren). Der nächste Turn (16) von Rebecca zeigt, dass sie die Selbsteinschätzung von Selina nicht kommentiert. Gemäß Erkenntnissen aus der Konversationsanalyse gilt der vorherige Turn somit als stillschweigend akzeptiert (Deppermann 2008). Dieses kurze Aushandeln zu Beginn darüber, wer die Frage beantwortet, lässt sich darauf zurückführen, dass sich die Teilnehmerinnen dieser Fallstudie zuvor nicht kannten und somit auch keine Hintergrundinformationen über das fachliche Vorwissen voneinander haben. Rebecca wendet sich nun der Beantwortung der Frage zu.

Auf Kontextebene nennt Rebecca in (16) „irgendwelche ganz persönlichen, ja personengebundene Eigenschaften“, die zufällig auf die Gruppen verteilt werden sollen. Als Beispiel dieser personengebundenen Eigenschaften nennt sie Alter und Gewicht, die „ungefähr eben gleich auf beide Gruppen“ aufgeteilt werden sollen, damit „nicht in einer Gruppe eben ganz viele nur ältere Menschen drin sind“. Dies ist eine Argumentation, die möglicherweise direkt aus der Lerneinheit übernommen wurde, denn dort wurde das Alter als eine Möglichkeit verwendet, nach welchem ein Versuchsleiter Gruppen nichtzufällig einteilen kann. Dadurch würde möglicherweise eine systematische Verzerrung der Ergebnisse entstehen. Ob Rebecca diese Konklusion zieht, kann aus dem Gesprochenen nicht erschlossen werden. Möglicherweise meint sie auch, dass nach der zufälligen Aufteilung geprüft werden solle, ob bestimmte Merkmale in beiden Gruppen gleichverteilt sind, wie beispielsweise das Alter. Mit Blick auf die a priori Analyse (Kapitel 9.2.2) fehlt in dieser Begründung, dass die von Rebecca genannten „personengebundenen Eigenschaften“ von Person zu Person schwanken, was im Sinne der Argumentationsanalyse nach Toulmin (1975) die Stützung der Regel *zufällige Aufteilung der Versuchspersonen auf die Gruppen* darstellt. Mit der zufälligen Aufteilung der Personen auf die Gruppen geht für Rebecca einher, dass diese dazu führt, dass „sich dann das ungefähr eben gleich auf beide Gruppen aufteilt“. Ihrem Konzept von zufälliger Aufteilung liegt also die Vorstellung zugrunde, dass dadurch alle Eigenschaften möglichst gleich auf die Gruppen aufgeteilt werden. „Ungefähr eben gleich“ als Formulierung spricht dafür, dass sie hier mit einer gewissen Streuung rechnet, dass also keine exakt gleiche Aufteilung erwartet wird. Andererseits kann dies auch als berechtigte Kritik am Verfahren des Randomisierungstests interpretiert werden, da nachgeprüft werden müss-

te, ob die zufällige Zuordnung zu einer Gleichverteilung bereits als relevant erkannter Merkmale geführt hat. Dass durch eine bewusste Gruppeneinteilung nach bestimmten Merkmalen eine Verzerrung der Ergebnisse möglich wäre, wird im nächsten Turn (17) von Selina, wie angekündigt, ergänzt. Zunächst zeigt Selina, dass sie mit der von Rebecca genannten Erklärung einverstanden ist („Ja, also genau das würde ich auch sagen“) und präzisiert die Folgen, die eine nichtzufällige Einteilung hätte in dem Sinne, dass man dadurch „so eine Studie oder so ein Ergebnis ja auch indirekt beeinflussen [könnte]“ (17). Zur Stützung greift sie das von Rebecca genannte Merkmal Alter auf, womit sie erneut zu erkennen gibt, dass sie die von Rebecca genannte Erklärung akzeptiert. Selinas Selbsteinschätzung, Ergänzungen zu der Erklärung von Rebecca liefern zu können, trifft zu.

Strukturell stellt sich die hier von Rebecca und Selina gemeinsam aufgebaute Argumentation wie in Abbildung 200 dar, wobei nur Selina den Einfluss des Versuchsleiters thematisiert.

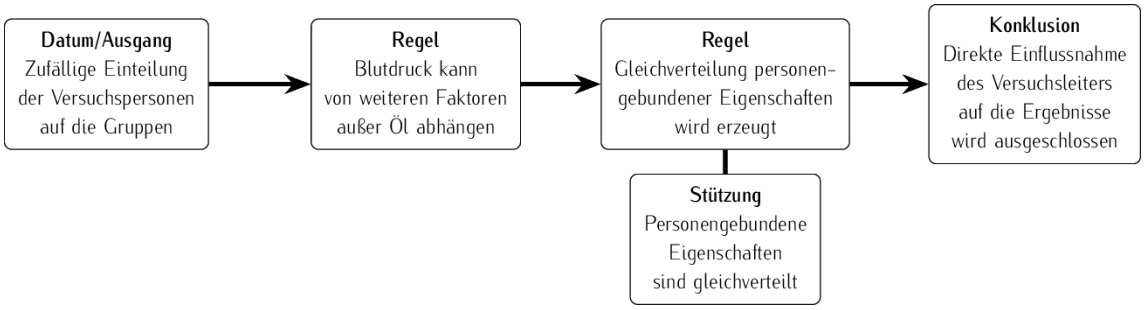


Abbildung 200: Rekonstruierte Argumentation von Rebecca und Selina zur zufälligen Aufteilung der Versuchspersonen auf die Gruppen

Interaktionseinheit 3: Mögliche Erklärungen für die beobachteten Unterschiede (Z. 21-46)

Der Frage N2 des Leitfadens nach möglichen Erklärungen für die beobachteten Unterschiede wird in der dritten Interaktionseinheit nachgegangen. Der folgende Transkriptausschnitt zeigt die relevanten Dialoge.

21 I: Mhm (.) Ok. (.) Gut. (..) Dann (.) die nächste Frage ist: Welche Erklärung gibt es denn für die Unterschiede, die man hier beobachten kann? Also wenn ihr jetzt einfach darüber nachdenkt/ (..) Wir sehen einfach es gibt Unterschiede, wenn wir die Mittelwerte betrachten. Da ist ein Unterschied von 7,7 zwischen den Mittelwerten der Fischölgruppe und der normalen Öldiätgruppe. Was sind mögliche Erklärungen, woran kann das liegen, dass das so unterschiedlich ist?

22 S: Ja also zum einen natürlich äh die/ (.) soll ich jetzt schon was zur Forschungshypothese sagen?

23 I: Gerne. Also wie du möchtest.

24 S: Also hier wird ja wahrscheinlich die Forschungshypothese sein, dass Fischöl wirksamer zur Blutdrucksenkung ist, als normales Öl. Und also das kann natürlich daran liegen, dass das Fischöl tatsächlich wirksamer ist, dann müsste man dementsprechend natürlich auch noch die Nullhypothese in Betracht ziehen. Was dann da wieder sagt, dass das Öl da überhaupt gar keine äh (.) Relevanz hat, sondern einfach nur die Zuordnung der Gruppen. Also da macht man dann ja

	im Anschluss den Randomisierungstest.
25-46	(…) Es wird in kurzen Turns geklärt, dass dies die beiden Erklärungen sind, die Rebecca und Selina einfallen. Inhaltlich wird nicht weiter diskutiert.

Es findet diesmal kein Aushandeln statt, wer die Frage nach möglichen Erklärungen (21) zunächst beantwortet. Im Gegensatz zur letzten Interaktionseinheit übernimmt Selina hier den Turn. Sie beginnt damit, einen Satz anzufangen, der eine mögliche Erklärung liefert, unterbricht sich jedoch selbst und zeigt einen Klärungsbedarf an, nämlich, ob jetzt bereits der Zeitpunkt gegeben sei, etwas zur Forschungshypothese zu sagen (22). Dies zeigt, dass sie eine Verbindung zieht zwischen einer möglichen Erklärung auf Kontextebene und dem, was sie in der Veranstaltung als Forschungshypothese als Element der Statistikebene kennen gelernt hat. Eine weitere mögliche Deutung ist, dass sie merkt, dass sie für eine Antwort die Ebenen Kontext und Statistik (vgl. Kapitel 6.4 bzw. Biehler et al. 2015) benötigt und zuvor die Absicht hatte, rein auf einer der beiden Ebenen, vermutlich der kontextuellen Ebene zu antworten. Durch die eigene Unterbrechung des Turns (22) lässt sich interpretieren, dass ihr diese Trennung der Ebenen bewusst ist und sie nicht sicher ist, auf welcher Ebene eine Antwort erwartet wird. Sie nutzt im Folgenden das gelernte Fachvokabular, um zwei mögliche Erklärungen zu beschreiben. Als Forschungshypothese formuliert sie, „dass Fischöl wirksamer zur Blutdrucksenkung ist als normales Öl“ (24). Als nächstes formuliert sie die Nullhypothese, die von ihr zwar nicht als weitere Erklärung betitelt wird, in der Interpretation durch den weiteren Gesprächsverlauf jedoch als zweite Erklärung angesehen werden kann. Die Nullhypothese wird von Selina formuliert als „dass das Öl da überhaupt gar keine äh (.) Relevanz hat, sondern einfach nur die Zuordnung der Gruppen.“ (24), mit dem Nachsatz, dass man daran anschließend einen Randomisierungstest mache. Selina sagt dies mit sicherer Stimme und ohne großes Stocken bis auf die kurze Pause vor „Relevanz“. Das lässt sich interpretieren daraufhin, dass sie bezüglich des Inhalts der Nullhypothese sicher ist. Die grammatikalisch unvollständige Formulierung der Nullhypothese und das kurze Stocken in dieser Passage deuten allerdings eine Unsicherheit bezüglich der Formulierung an. In der ausführlichen Analyse der Einzeläußerungen wurden verschiedene Deutungsalternativen erarbeitet, um den unvollständigen Nebensatz „sondern einfach nur die Zuordnung der Gruppen“ zu vervollständigen. Als plausibelste Deutung erschien hierbei, dass dies die Erinnerung an eine Formulierung ist, die in der Lerneinheit verwendet wurde, um die Nullhypothese auszudrücken, ohne dass diese in Tiefe verstanden wurde, um sie hier flüssig übertragen zu können. Der nächste Satz „Also da macht man dann ja im Anschluss den Randomisierungstest“ (24) zeigt, dass sie einen Schritt weiter denkt und die mögliche Erklärung, die sie in der Nullhypothese ausdrückt, als Nullmodell einstuft, das im Folgenden getestet werden soll.

Im Vergleich wird die Forschungshypothese zu Beginn des Turns (24) flüssiger und sprachlich verständlicher hervorgebracht. Durch die Sprache lässt sich interpretieren,

dass die Formulierung der Forschungshypothese als eine mögliche Erklärung Selina leichter fällt als die Formulierung der Nullhypothese als weiterer möglicher Erklärung. Trotzdem kann sie jedoch auf Anhieb beide gesuchten Erklärungen nennen. Die sich anschließenden Redebeiträge (25-46) bestätigen im Wesentlichen die von Selina genannten Erklärungen und drehen sich noch einmal um die zufällige Aufteilung der Personen auf die Gruppen, ähnlich zu den bereits zuvor gemachten Äußerungen der Interaktionseinheit 2.

Im Wesentlichen nennt Selina in ihrem zweiten Turn (24) die beiden möglichen Erklärungen. Dies verbindet sie sofort mit dem gelernten Fachvokabular von Forschungshypothese und Nullhypothese. Diese werden im Laufe der Einheit weiter diskutiert. Rebeccas Beiträge beschränken sich auf kurze Zustimmungen zu Selinas Ausführungen. Dies lässt sich deuten, dass sie den Erklärungen von Selina zustimmt. Dass die dritte mögliche Erklärung, dass es an Störvariablen liegt, nicht genannt wird, mag eine Ursache darin haben, dass dies in der Interaktionseinheit direkt zuvor geklärt wurde und durch das Design des Experiments ausgeschlossen ist als mögliche Erklärung. Kommunikation findet hier sowohl auf Kontext- als auch auf Statistikebene statt.

Interaktionseinheit 4: Überleitung zum Randomisierungstest (Z. 47-56)

In dieser Interaktionseinheit wird den beiden Teilnehmerinnen das Material (Randomisierungstestschema, Stift, leere Zettel, Laptop) überreicht, mit dem sie arbeiten sollen. Von Selina wird bestimmt, dass Rebecca den Stift nimmt und somit die Aufgabe erhält, das Randomisierungstestschema auszufüllen. In dieser Einheit gibt es weder einen Bezug zum Inhalt der Aufgabe noch zur Statistik, weshalb sie aus der ausführlichen Interpretation ausgelassen wird.

Interaktionseinheit 5:Forschungsfrage notieren (Z. 57-63)

Der folgende Transkriptausschnitt zeigt die Interaktionseinheit zum Konzept der Forschungsfrage.

57	S: Ja, welche Frage soll beantwortet werden?
58	R: Ja ich würde sagen, ob ähm/ also ist eine Fischöldiät/ oder ist Fischöl wirksam, um eine Blutdrucksenkung herbeizuführen?
59	S: Ja.
60	R: Ich schreib das mal auf. (der Stift funktioniert nicht)
61	S: So jetzt, jetzt tut er es.
62	R: Mhm (bejahend).
63	*Es wird notiert: „Ist Fischöl wirksam, um eine Blutdrucksenkung herbeizuführen?“*

Hierbei handelt es sich um eine der kürzesten Interaktionseinheiten des gesamten Gesprächs. Es findet ein direkter Einstieg in die Aufgabe anhand des Schemas auf Kontextebene statt. Selina eröffnet die Interaktionseinheit und übergibt den Turn an Rebecca (57), die ihn aufnimmt und einen Formulierungsvorschlag für die Forschungsfrage macht (58). Der Kontext scheint ihr so klar zu sein, dass sie mit einem kurzen Anlauf

direkt eine Formulierung findet, die sofort von Selina bestätigt wird (59). Anschließend drückt sie ihr Vorhaben aus, dies direkt aufzuschreiben (60), woraufhin Selina ihre Zustimmung durch ein bejahendes „Mhm“ ausdrückt (59). Weitere Diskussionen oder Alternativformulierungen der Forschungsfrage finden nicht statt.

Interaktionseinheit 6: Beobachtete Daten (Z. 64-81)

Dem Konzept der Beschreibung der beobachteten Daten wird in der sechsten Interaktionseinheit nachgegangen. Diese Interaktionseinheit wird geteilt in zwei Episoden. In der ersten Episode wird das Design des Experiments im Hinblick auf die erfolgte Randomisierung von den Teilnehmerinnen diskutiert. Die zweite Episode dreht sich um den Unterschied, der zwischen den beiden Gruppen des Experiments auszumachen ist.

Episode 6-1: Design des Experiments (Z. 64-71)

64	R: Ja dann bei Randomisierung (.) ähm würde ich sagen dass man sich in dem Moment einfach fragen muss wie, ähm war das denn am Anfang erst mal als die Gruppen zugeteilt werden. Wie haben die Forscher das gemacht. Also haben sie da schon randomisiert oder haben sie eben genau gesagt, wegen der Eigenschaft gehst DU in DIE Gruppe oder DU in DIE oder war da eben auch schon eine zufällige Zuordnung/
65	S: /Und die war halt zufällig/
66	R: /Vorhanden. Deswegen aufgrund der zufälligen Zuordnung. Es ist zwar jetzt nicht genau gesagt, wie sie das zufällig gemacht haben, ob sie jetzt eine Münze geworfen haben oder (.). Aber man kann ja zumindest schon mal sagen, dass eine Randomisierung irgendwo stattgefunden hat.
67	S: Ja.
68	R: Ich schreib da einfach nochmal zu: Zufällige Zuordnung.
69	S: Zufällige Zuordnung würde ich einfach schreiben.
70	*Es wird notiert: Zufällige Zuordnung.*
71	R: Ja.

Das Paar lässt sich von dem Schema und den dort notierten Fragen leiten. Das macht der Beginn von Rebeccas Turn (64) deutlich, in welchem sie zu Beginn das Wort „Randomisierung“ vorliest. Sie versucht, dem Design des Experiments auf den Grund zu gehen, indem sie sich auf Kontextebene in die Lage der Forscher versetzt („Wie haben die Forscher das gemacht.“). Sie zeigt zwei Möglichkeiten auf, wie die Forscher die Gruppen eingeteilt haben können, entweder randomisiert oder eine Einteilung nach bestimmten Merkmalen. Auch hier zeigt sich in der Interaktion, dass ihre Partnerin ihren Gedankengängen folgt, denn sie vervollständigt sofort die Ausführungen, mit „und die war halt zufällig“ (65). Im folgenden Turn (66) beendet Rebecca ihren letzten Turn (64), der durch Selinas Einwurf (65) unterbrochen wird. Rebecca hinterfragt hier möglicherweise die Aufgabenstellung, in der die zufällige Aufteilung der Versuchspersonen auf die Gruppen ohne weitere Details angegeben wurde. Sie versetzt sich hier nun noch weiter in die Situation und bezieht die Statistikebene ein, indem sie überlegt, wie eine Randomisierung stattgefunden haben könne, zum Beispiel durch einen Münzwurf. Der Münzwurf als eine Möglichkeit, Personen zufällig auf zwei Gruppen aufzuteilen, wurde in der Vorlesung als Beispiel genannt und wird hier offensichtlich von Rebecca abgeru-

fen. Die Auswahl der Stichprobe für das Experiment wird nicht von den beiden diskutiert. Hierzu gibt das Schema keine Anweisung vor, möglicherweise fehlt deshalb der Anlass, sich damit zu beschäftigen, welcher Art die Stichprobe ist.

Episode 6-2: Unterschied zwischen den beiden Gruppen (Z. 72-81)

72	S: Ja der Unterschied, der sich feststellen lässt ist dann ja, dass Fischöl den Blutdruck höher senkt als normales Öl. Oder?
73	R: Ja zumindest auf jeden Fall, dass ähm (..) im arithmetischen Mittel ähm die (.) Blutdrucksenkung bei der Fischölgruppe um diese sieben/.
74	S & R: /Komma 7/
75	S: /Ja.
76	R: Millimeter Quecksilber/
77	S: /größer ist.
78	R: Soll ich das so [notieren]?
79	S: In Stichpunkten. Kurz und prägnant.
80	R: Ja.
81	*Es wird notiert: „- zufällige Zuordnung, - Blutdrucksenkung bei Fischöl-Gruppe im ar. Mittel 7,7mm QS. größer“*

In der nächsten Episode beginnt Selina zunächst mit einer Schlussfolgerung im Kontext, in der sie eine Kausalbeziehung zwischen dem Fischöl und den Blutdrucksenkungen formuliert. Eine korrekte Formulierung hier wäre, dass es zwischen den beiden Gruppen folgende Unterschiede gibt und für die Erklärung zwei Hypothesen. Dies geschieht jedoch nicht sorgfältig. Nach dieser Aussage (72) schränkt sie selbst ihre Interpretation mit dem nachfolgenden „Oder“, das fragend hervorgebracht wird, wieder ein. Sie fordert mit diesem turn-ausleitenden „Oder?“ gleichzeitig eine Reaktion von Rebecca ein, vielleicht auch aus dem Grund, dass sie ihre eigene Interpretation in Frage stellt. Andererseits kann sie hiermit auch die Absicht hervorbringen, nun den Unterschied in den beobachteten Daten diskutieren zu wollen, der als nächstes im Randomisierungstestschema als Anweisung notiert ist. Mit dem Turn (72) bietet sie nun Rebecca einen Diskussionsanlass. Rebecca bestätigt zunächst die Interpretation von Selina mit „Ja …“ (73), schränkt sie dann allerdings konkret auf den Unterschied zwischen den Mittelwerten der beiden Gruppen ein, ohne weiter auf die von Rebecca zuvor konkret formulierte Schlussfolgerung einzugehen. Mit schnellen Wechseln der Redebeiträge werden gemeinsame Sätze vervollständigt und nur der im Aufgabentext bereits vorgegebene Unterschied der arithmetischen Mittel von 7,7 formuliert und präzise als Vergleich notiert. Die gemeinsam hervorgebrachte Beobachtung ist im Sinne des Erwartungshorizonts sehr gut.

Nach der Argumentationsanalyse beginnt Selina diese Episode mit der Konklusion, die sie hier quasi als Behauptung, angezeigt durch das nachgefügte „Oder?“ zur Debatte stellt. Daraufhin formuliert Selina die Ausgangslage, nämlich die beobachtete Differenz der arithmetischen Mittel der beiden Gruppen, die in den folgenden Turns mit der Angabe des Werts von 7,7 mmHg konkretisiert wird.

Interaktionseinheit 7: Nullmodell (Z. 82-88)

Null- und Forschungshypothese werden in der siebten Interaktionseinheit aufgestellt.

82	S: Forschungshypothese ist ja wieder, ob Fischöl/ (.) dass das Fischöl gegen Bluthochdruck wirkt oder den Blutdruck senkt.
83	R: Ja.
84	*Es wird notiert: „F: Fischöl ist hilfreich zur Blutdrucksenkung“*
85	R: Ja Nullhypothese dann, dass die zufällige Zuordnung zu diesen BEIDEN Gruppen eben die Ursache für diese beobachteten Unterschiede/
86	S: /Ja.
87	R: /Dann darstellt.
88	*Es wird notiert: „N: Zufällige Gruppenzuordnung als Ursache für beobachtete Unterschiede.“*

Wiederum ohne sprachliche Überleitung wenden sich Selina und Rebecca dem nächsten Schritt des Schemas zu. Selina leitet mit den ersten Worten das statistische Thema dieser Gesprächseinheit ein „Forschungshypothese ist ja wieder …“ (82). Das „wieder“ macht deutlich, dass sie diese bereits zuvor formuliert haben und hier wieder aufgreifen. Die Formulierung auf Kontextebene wird im Folgeturn von Rebecca bestätigt und notiert. Daraufhin wird die Nullhypothese ebenso kurz von Rebecca formuliert (85) und notiert. Die beiden zeigen sich sicher im Umgang mit der Null- und der Forschungshypothese. Beide Hypothesen werden im Kontext formuliert.

Interaktionseinheit 8: Simulieren/Zufallsmaschine erstellen (Z. 89-210)

In der Interaktionseinheit 8 beschäftigen sich die beiden Teilnehmerinnen mit dem Erstellen der Simulation in TinkerPlots. Diese Interaktionseinheit wird inhaltlich in sechs Episoden unterteilt. In der ersten Episode planen die Teilnehmerinnen ihr Vorgehen und bemerken ihre eigenen Unzulänglichkeiten bezüglich des Umgangs mit TinkerPlots. In der zweiten Episode erklärt Selina den Vorgang des Randomisierens. Die dritte Episode ist die umfangreichste, in dieser erstellen die Studierenden die Zufallsmaschine mit viel technischer Hilfe durch die Interviewerin. Anschließend beschäftigen sich die Studierenden mit der Benennung der neuen Gruppen, aus analysierender Sicht ist dies eine der interessantesten Episoden, da hier viel von den kognitiven Denkweisen zum Randomisieren offenbart wird. In der fünften Episode wird die Reproduktion der Gruppengrößen thematisiert und in der Software abgebildet. Gleiches geschieht für die Experimentdurchführung, abgebildet in der Software durch die Durchgangsanzahl in Episode sechs, zu der auch das anschließende Laufenlassen der Zufallsmaschine gezählt wird.

Episode 8-1: Planen des Vorgehens (Z. 89-98)

89	R: Ja gut beim Randomisieren. Willst du was sagen oder?
90	S: Ja (lachend) also wir könnten das ja jetzt über TinkerPlots (.) simulieren.
91	R: Ja genau.
92	S: Oder wir könnten es händisch machen, indem wir einfach die beiden Gruppen nochmal zufällig äh zuordnen und dann daraus wieder das arithmetische Mittel ziehen und dann/.
93	R: Genau das mehrmals dann/.
94	S: /Mehrmals ganz, ganz viel/ Kannst du das bei TinkerPlots?

95	R: Mhm (verneinend). Ich weiß nicht mehr wie das geht (lachend).
96	S: Ich auch nicht (lachend).
97	R: Ich weiß irgendwie nur noch Messgrößen sammeln, um die Referenzverteilung zu erstellen (lachend) aber/.
98	S: /Dafür war echt die Zeit zu knapp/ um das so richtig. Aber da können wir ja ganz viele Wiederholungen auf einmal machen, um dann einen aussagekräftigen Wert zu finden. Wenn wir es jetzt händisch machen würden, dann (.) würden wir ja ein bisschen hier sitzen (lachend) also (..).

Rebecca leitet diese Episode ein und startet direkt mit der Turnübergabe an Selina (89). Dies kann gedeutet werden entweder, dass sie sich unsicher fühlt und nicht weiß, was nun für das „Randomisieren“ zu tun ist. Oder sie sich auch gerade sicher in diesem Bereich fühlen und möchte Selina die Gelegenheit geben, ihre Gedanken hier auszuführen. Beide Deutungen werden gestützt durch die durchgeführte Turn-by-Turn Analyse, die einen Rückbezug erlaubt, da in der folgenden Episode Rebecca diejenige ist, die ausführlich das Vorgehen des Randomisierens beschreibt. Ihr Prozesswissen zum Randomisierungstest kommt dabei zum Vorschein, es wird aber auch deutlich, dass sie nicht (mehr) weiß, wie sie TinkerPlots als Instrument in diesem Prozess einsetzen kann. Selina übernimmt den Turn, das paraverbale Lachen zu Beginn der Turnübernahme (90) lässt die Deutung zu, dass sie weiß, dass Rebecca das Randomisieren selbst hätte ausführen können oder die alternative Deutung, dass sie davon ausgeht, die Aufgabe selbst gut bewältigen zu können. Sie zeigt zunächst die Möglichkeit auf, dass sie das mit TinkerPlots simulieren könnten. Rebeccas Antwort „Ja genau“ (91) stützt die Interpretation, dass sie diese Antwort erwartet und ihre erste Anfrage in (89) selbst hätte beantworten können. Daraufhin übernimmt wieder Selina und erwähnt die zweite Möglichkeit, die sie gelernt hat, um einen Randomisierungstest durchzuführen, die händische Simulation. Dies beschreibt sie etwas unpräzise als „… indem wir einfach die beiden Gruppen nochmal zufällig äh zuordnen und dann daraus wieder das arithmetische Mittel ziehen und dann“ (92) und bricht dann ab. Hier wird nicht deutlich formuliert, wie die beiden Gruppen neu zugeordnet werden sollen, aber positiv kann vermerkt werden, dass dieser Vorgang zufällig geschehen soll. Auch ist ihr klar, dass sie das arithmetische Mittel betrachten soll. Dass jedoch die Differenz der Mittelwerte als Messgröße gesammelt und in der Referenzverteilung dargestellt werden soll, wird hier noch nicht genannt. Ein expliziter Bezug beispielsweise zu dem Seminarinhalt des Bausteins Inferenzstatistik, in welchem eine händische Simulation durchgeführt wurde, geschieht hier nicht. Rebecca ergänzt, dass dies mehrmals geschehen solle (93) und Selina fällt bestätigend ein und verstärkt dies sogar zu „ganz, ganz viel“ (94). Es lässt sich interpretieren, dass das Wissen vorhanden ist, dass eine sehr hohe Anzahl an Wiederholungen gemacht werden soll, allerdings lässt sich nicht schließen, ob das Wissen ist, das während des Bausteins Wahrscheinlichkeit erworben wurde, da hier viel mit Simulationen gearbeitet wurde, oder ob dies explizit Wissen ist, das bezüglich des Randomisierungstests aufgebaut wurde. Selina fragt im gleichen Turn noch nach, ob Rebecca dies bei TinkerPlots könne (94). Aus der Satzbetonung kann interpretiert werden, dass sie damit ihre eigene Unsi-

cherheit bezüglich TinkerPlots zum Ausdruck bringt und nun wissen möchte, ob Rebecca das kann. Möglicherweise kann sie es auch selbst und könnte an dieser Stelle übernehmen. Ebenfalls kann interpretiert werden, dass sie kognitiv den Transfer vom händischen Simulieren zum Simulieren mit Software leistet, um eine hohe Wiederholungszahl durchführen zu können. Rebecca antwortet auf ihre Frage verneinend, erneut lachend (95). Eine Deutungsalternative hierzu ist, dass beiden bewusst ist, dass dies Inhalt des letzten Bausteins des Kurses war, und sie dort die nötigen Kompetenzen hätten erwerben können. In den nächsten Turns werden einzelne Bruchstücke in den Raum geworfen, mal auf statistischer Ebene (Referenzverteilung), mal auf Ebene der Software (Messgrößen sammeln), diese werden jedoch nicht in einen Zusammenhang eingeordnet. In (97) sagt Selina, dass die Zeit zu knapp war, „um das so richtig“. Sie bricht den Satz an dieser Stelle ab und fährt fort damit, dass sie TinkerPlots gegenüber der händischen Simulation bevorzugen würde, weil sie dort „ganz viele Wiederholungen auf einmal machen [können], um einen aussagekräftigen Wert zu finden“ (98). Dies ist eine Aussage sowohl auf statistischer als auch auf Softwareebene, in Bezug auf die Statistikebene ist das allerdings ungenau. Es bleibt unklar, was „so richtig“ hätte gelernt werden sollen, aus dem Zusammenhang lässt sich hier schließen, dass sie die Simulation in TinkerPlots meint. Der Hinweis auf die knappe Zeit während des Bausteins zur Inferenzstatistik lässt die Deutung der vorigen Turns plausibel erscheinen, dass ihre instrumentelle Genese zur Verwendung von TinkerPlots für einen Randomisierungstest noch nicht genügend aufgebaut war. Gleichzeitig ist es eine Rechtfertigung, warum sie die Simulation jetzt nicht selbständig durchführen können.

Diese Episode drückt auch aus, dass beiden Teilnehmerinnen klar ist, was von ihnen erwartet wird, nämlich dass sie simulieren sollen. Sie reflektieren hier jedoch ihre eigenen Kompetenzen und kommen beide zu dem Schluss, dass sie die Simulation in dieser Situation nicht selbständig erstellen können. Im überwiegenden Teil der Episode wird auf statistischer Ebene kommuniziert, ohne dass ein Bezug zum Kontext hergestellt wird. Die statistische Ebene wird am Ende auf die Softwareebene transferiert.

Episode 8-2: Erklären des Vorgangs (Z. 99-104)

99 I: /(lacht) (.) Das heißt, vielleicht könnt ihr einmal kurz erklären, was ihr machen wollt und ich helfe euch dann einfach dabei das quasi in TinkerPlots zu übersetzen.

100 R: Genau. Danke. Also. Was wir jetzt im Prinzip machen wollen ist, dass wir nochmal wieder/ wir haben jetzt wieder diese Werte aus den Gruppen und wir wollen jetzt nochmal wieder die Werte, die wir hier eben haben/ diese vierzehn Stück sind es glaube ich/ wieder in zwei neue Siebenergruppen einteilen. Das heißt im Prinzip würden wir, wenn wir das jetzt händisch machen würden, die ganzen Werte auf diese Datenkarten packen, in eine Urne legen und dann immer wieder (.) sieben rausziehen. Das heißt wir hätten (.). Zum Beispiel nachher diesen Wert acht aus der Fischöldiätgruppe zum Beispiel mit diesem minus drei aus der Normale Öldiätgruppe eventuell in einer Kontrollgruppe A. Und dann ist es ja im Grunde so, dass wir dann von diesen Gruppen, dass wir dann von diesen Gruppen eben wieder in dem Fall das arithmetische Mittel bilden würden und dann eben wieder die Differenz bestimmen würden, wie es eigentlich genau hier (zeigt auf das Arbeitsblatt) gemacht wurde. Und um da dann eben eine Aussage ziehen

zu können, ist es eben so hilfreich das dann mit TinkerPlots und nicht händisch zu machen, damit wir das eben entsprechend häufig/.
101 S: /Genau.
102 R: Wiederholen können.
103 I: Ok.
104 S: Nur die Software bedienen (.).

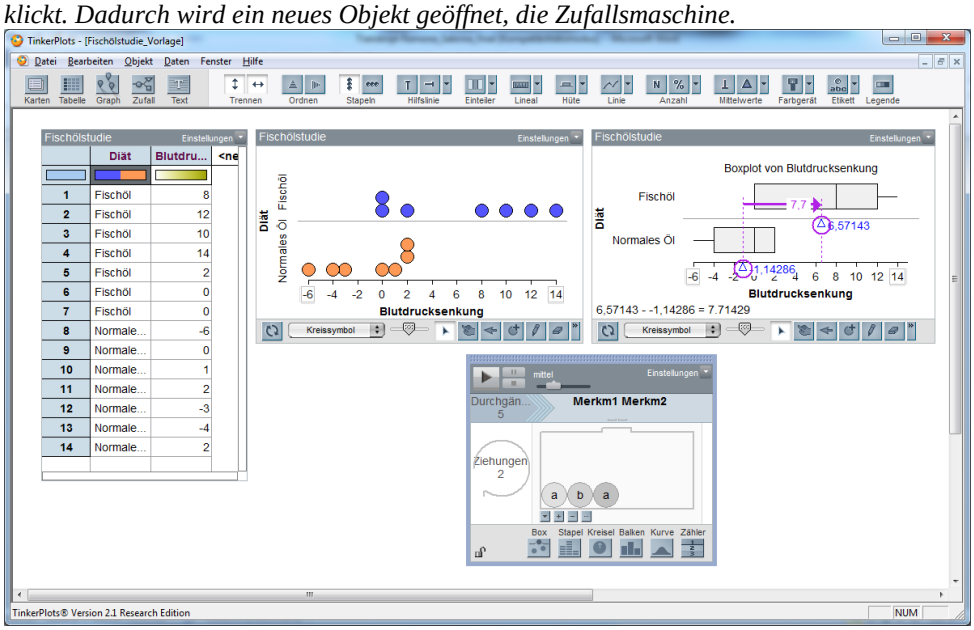
Nach den verbalisierten Schwierigkeiten in der letzten Episode greift die Interviewerin mit einer strategischen Intervention ein und bietet Hilfe für die Umsetzung in TinkerPlots an (99). Dazu fordert sie eine Erklärung dessen, was die Studierenden machen wollen. Rebecca übernimmt direkt den Turn und erläutert den Vorgang (100). Diese recht kompetente Beschreibung führt zu der zuvor beschriebenen Interpretation ihrer Aussagen in Episode 1 dieser Einheit. Sie beginnt ihre Ausführungen damit, dass die vorhandenen 14 Werte in zwei neue Siebenergruppen eingeteilt werden sollen. Sie beschreibt den Vorgang, wie sie ihn mit Datenkarten machen würde. Die Werte sollten auf Datenkarten notiert werden, diese in eine Urne gelegt und daraus immer sieben gezogen werden. Hier wird ein Bezug zum Baustein Inferenzstatistik deutlich, in welchem sie dieses Vorgehen als händische Simulation kennen und auf die Software zu übertragen gelernt hat. An einem Beispiel mit einzelnen Werten zeigt sie, dass sie die Zuordnung zu den ursprünglichen Gruppen als aufgehoben ansieht, da sich Werte aus beiden ursprünglichen Gruppen nun beispielsweise in einer „Kontrollgruppe A“ befinden können. Hier zeigt sich erneut die Übernahme der Wortwahl aus der Lerneinheit. Sie erläutert weiter, dass von diesen Gruppen das arithmetische Mittel und die Differenz dazwischen gebildet werden müsse, wie auf dem Arbeitsblatt für das ursprüngliche Experiment. Mit Bezug auf die häufige Durchführung dieses Vorgangs nennt sie TinkerPlots als hilfreich am Ende des Turns. Bei dem Wort „häufig“ fällt ihr Selina bestätigend ins Wort (101). Von der Interviewerin wird diese Ausführung mit „Ok“ als akzeptiert angezeigt (103). Daraufhin wirft Selina noch einmal die Softwarebedienung ein (104). Ihre Stimmlage lässt die Interpretation zu, dass sie mit dem von Rebecca beschriebenen Prozess einverstanden ist und diesen bestätigt, aber dass sie Schwierigkeiten auf der Ebene der Software erwartet. Diese Episode zeigt, dass eine präzise Erinnerung an die händische Simulation besteht, an die Benutzung von TinkerPlots jedoch nicht. Das stützt das durchgeführte Lehrkonzept, in welchem es vorrangig um Verständnisaufbau zum Randomisierungstest ging, stark.

Soweit hier Aspekte genannt wurden, scheinen die beiden Teilnehmerinnen sowohl die Ebene des Kontexts als auch der Statistik gut beschreiben zu können. Sie selbst schätzen die Ebene der Software als problematisch ein.

Episode 8-3: Die Zufallsmaschine erstellen (Z. 105-132)

105 I: Ja dann unterstütze ich euch einfach bei der Softwarebedienung. Also als erstes braucht ihr eine Zufallsmaschine. Das war da, wo Zufall steht.
106 *Oben in der Leiste befindet sich der Button mit der Beschriftung "Zufall". Dieser wird ange-*

klickt. Dadurch wird ein neues Objekt geöffnet, die Zufallsmaschine.



107 I: (.) Genau. Und was ihr mir gerade erklärt habt ist, ihr wollt einmal diese ganzen Werte in die Urne legen, wo zufällig rausgezogen wird, ist das richtig?

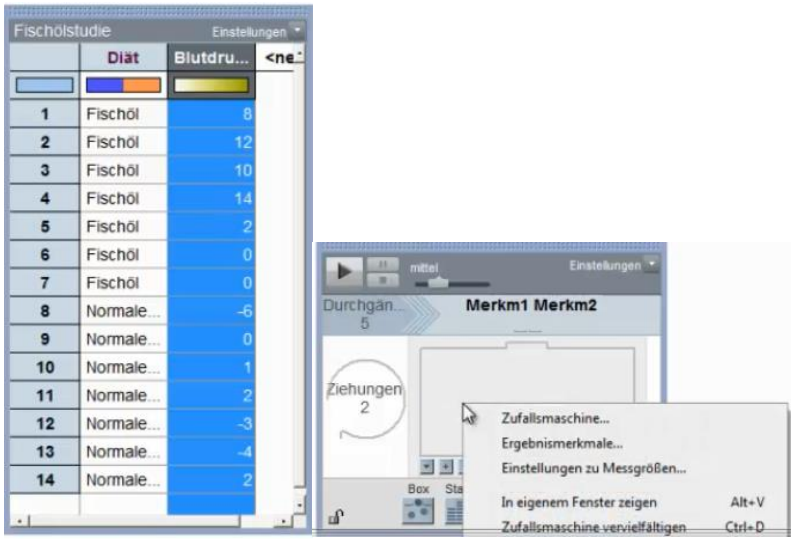
108 R&S: Genau, ja.

109 I: Gut, dann löscht ihr jetzt als erstes alle Kugeln raus. Das geht mit dem Minus.

110 *Die drei Kugeln werden gelöscht.*

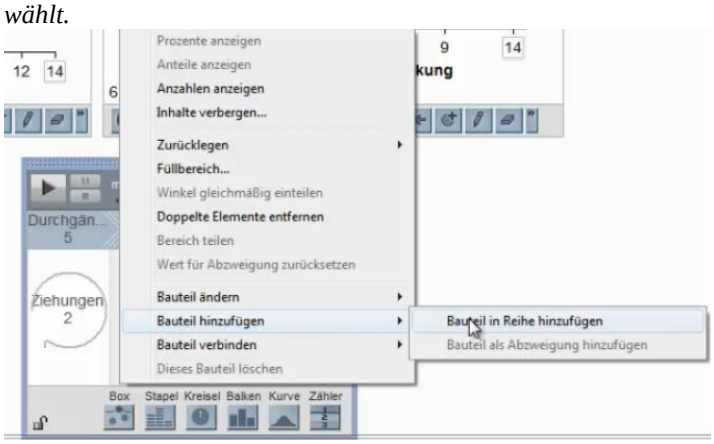
111 I: Und markiert die Spalte mit den Werten, die wir haben. Einmal oben auf Blutdruck, genau. Und jetzt auf „Steuerung c“, dann kopiert ihr die ganzen Werte und legt die dann da rein. Einmal in die Zufallsmaschine klicken, wo leer steht. Noch (.) Nein.

112 *Die Studentin hat zunächst die rechte Spalte "Blutdruck" markiert und kopiert und anschließend mit der Maus mit einem Rechtsklick auf den Teil der Zufallsmaschine geklickt, der als "leer" betitelt ist. Es öffnet sich das Kontextmenü.*

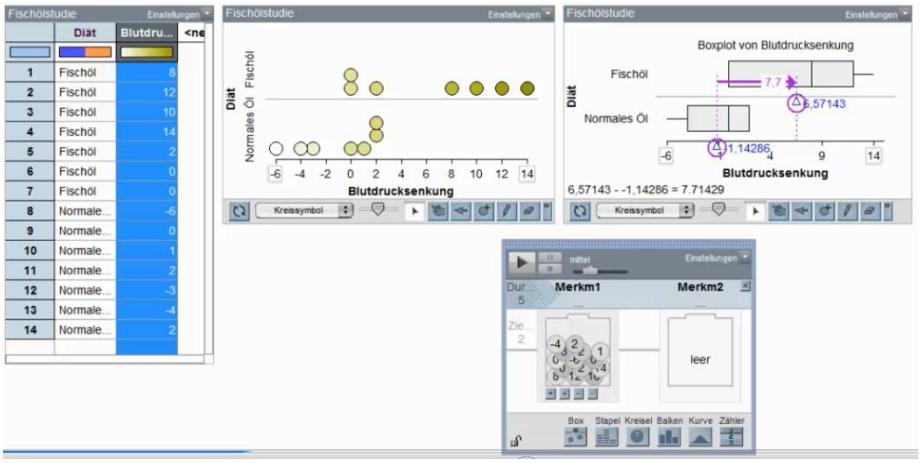


113 I: Halt, einfach nur normal reinklicken.
114 R: Genau, dann mach ich jetzt einfügen und dann sind die schon mal alle drin.
115 S: Ja.
116 *Die Studentin klickt auf den entsprechenden Teil des Objekts, drückt Steuerung +v und die Werte erscheinen.*

117 I: Ist das das, was ihr im ersten Schritt machen wolltet?
118 R: Genau. Dass wir die alle in einer (.) Urne wieder zusammen haben und nicht mehr in diese zwei Gruppen aufgeteilt.
119 S: Mhm (bejahend) Und jetzt bräuchten wir natürlich aber diese Aufteilung auf zwei Gruppen.
120 R: Genau.
121 S: Das heißt wir bauen einfach noch ein Bauteil dahinter (.) wo wir die anderen quasi hineinlegen. Wenn wir jetzt einfach nochmal/
122 *Das Kontextmenü wird geöffnet und der Eintrag "Bauteile hinzufügen" (in Reihe) wird ausgewählt.*



123 *Es erscheint ein weiteres Bauteil in der Zufallsmaschine.*



124 S: Bauteile hinzufügen war das, oder?
125 I: Genau.
126 S: In Reihe. Und jetzt?
127 I: Mhm (bejahend) so, jetzt kann man noch unten wählen, was ihr für ein Bauteil möchtet. Da kannst du zum Beispiel dieses Stapelbauteil nehmen.
128 *Die Studentin klickt zunächst den Button "Stapel" an, hält ihn anschließend gedrückt und zieht ihn in die leere Urne.*

129 I: Ja, das musst du gedrückt halten und da rein schieben. Genau jetzt machst du zwei Mal plus.
130 *Die Studentin klickt zwei Mal auf das "Plussymbol" unterhalb der leeren Urne und es erscheinen zwei Felder, die automatisch von TinkerPlots mit a und b beschriftet sind.*

131 R: Das sind jetzt sozusagen die zwei Gruppen.
132 S: Mhm (bejahend).

Diese Episode wird in weiten Teilen zusammengefasst, da hier der meiste Dialog technische Umsetzungen in der Software betrifft.

Die Interviewerin gibt auf Softwareebene kurze Hinweise, welche Operationen in TinkerPlots ausgeführt werden müssen, um die Zufallsmaschine zu erstellen und die Studierenden führen diese Aktionen kompetent und ohne weitere Nachfragen aus. Zunächst

werden auf Anweisung alle Kugeln aus einer Standardzufallsmaschine gelöscht (105-110) und die Werte der Blutdruckreduktionen in sie kopiert, wie es in Kapitel 10.4.4 beschrieben ist (111-116).

Abbildung 201: Erstes Zwischenergebnis der Zufallsmaschine von Selina und Rebecca

Bis zur Erstellung der Zufallsmaschine, wie sie in Abbildung 201 zu sehen ist, befinden sich die Gesprächsbeiträge auf Softwareebene. Hieran schließt sich ein Wechsel auf die Statistikebene an (117-121).

Um sich zu versichern, dass die in Abbildung 201 gezeigte Zufallsmaschine den Vorstellungen der Studierenden entspricht, fragt die Interviewerin hier nun nach (117), da bis hierher im Wesentlichen ihren Anweisungen gefolgt wurde. Rebecca antwortet zunächst auf Statistikebene und betont noch einmal, wie in der Episode zuvor auch schon, dass hier alle (Blutdruck-)Werte zusammen sein sollen „… und nicht mehr in diese zwei Gruppen aufgeteilt.“ (118). Eine Deutung ist, dass das „diese“ hier auf die Gruppen des ursprünglichen Experiments referiert, die nun nicht mehr beachtet werden sollen. Selina übernimmt den Turn, bestätigt kurz das zuvor Gesagte und führt weiter aus „… jetzt bräuchten wir natürlich aber diese Aufteilung auf zwei Gruppen.“ (119). Durch das Weglassen des Demonstrativpronomens „diese“ bei „auf zwei Gruppen“ kann interpretiert werden, dass Selina nicht die Gruppen des ursprünglichen Experiments meint, sondern zwei neue Gruppen, zum Beispiel die, von denen eine vorher als „Kontrollgruppe A“ betitelt wurde (100). Auf statistischer Ebene können die beiden interaktionistisch beschreiben, wie der weitere Prozess aussehen soll und wechseln dabei auch auf die Softwareebene (121). Ein weiteres Bauteil wird ohne Hilfe durch die Interviewerin der Zufallsmaschine in Reihe hinzugefügt (122-128), für die Befüllung des zweiten Bauteils fordert die Interviewerin auf, zweimal das Plus-Zeichen zu betätigen (129), wodurch eine Zufallsmaschine wie in Abbildung 202 erstellt wird.

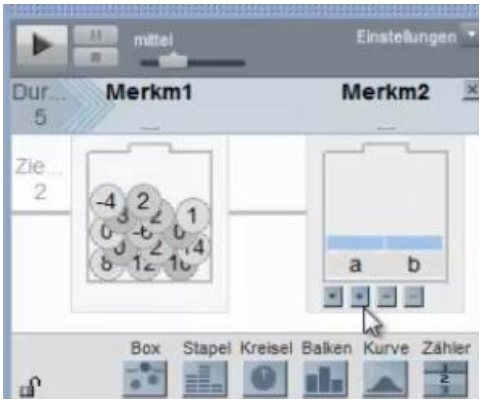


Abbildung 202: Weiteres Zwischenergebnis der Zufallsmaschine von Rebecca und Selina

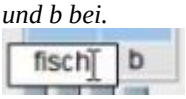
Hier wechselt Rebecca aus der Ebene der Software in die Ebene der Statistik und erläutert dies (131). Dies mag auch dazu dienen, ihren jetzigen Wissensstand bzw. ihr Verständnis der vorliegenden Zufallsmaschine auszudrücken und möglicherweise erwartet sie Bestätigung, dass sie dies richtig verstanden hat. Diese Bestätigung liefert ihr Selina (132), so dass es sich nun um geteiltes Wissen handelt, dass die zwei (neuen) Gruppen nun implementiert sind.

Episode 8-4: Umbenennung der beiden neuen Gruppen (133-153)

133 I: Die könnt ihr jetzt benennen. Da wo a und b steht, die könnt ihr umbenennen, wenn ihr möchtet.

134 S: Wir können ja eine Fischöl und die andere/. Hier die eine nennen wir "Fisch" und die andere "Normal".

135 *Die Studentin beschriftet den ersten Stapel zunächst neu, behält dann jedoch die Buchstaben a und b bei.*



136 R: Müssen wir die denn so nennen, weil eigentlich teilen wir die ja jetzt so auf dass wir da nachher zum Beispiel/. Also ich bin nicht ganz sicher, aber wir machen ja jetzt gerade diesen zweiten Schritt, dann kann es ja sein dass wir/.

137 S: /Stimmt, brauchen wir eigentlich nicht.

138 R: Eine Fischölperson mit einer Normalölperson/.

139 S: /Du hast Recht/.

140 R: /Zusammen. Ich finde a und b ist eigentlich ganz neutral.

141 S: Man kann eigentlich a und b nehmen, weil das dann mehr. Wir wollen ja jetzt nachweisen, dass es nichts mehr mit dem Öl zu tun hat.

142 R: Ja genau wir werden es sagen/.

143 S: /Also muss man/

144 R: Wir sind Vertreter der Nullhypothese.

145 S: Eben, weil dann kommen in die eine Gruppe möglicherweise immer beide, eine Person mit Fischöl und eine mit normalem Öl. Es geht ja jetzt nur um die Blutdrucksenkung.

146 R: Ja genau das kann es ja sein/.

147 S: /Ja.

148 R: Dass du hiervon (*zeigt auf die Fischölgruppe im Plot in TinkerPlots*) vielleicht fünf hast und davon (*zeigt auf die „normale Öl“ Gruppe im Plot*) dann (.) den Rest. Also es kann ja mal so mal so/.

149 S: /Du hast Recht. (lacht) Also ich würde dir ja spontan zustimmen. (lachend)

150 R: Ok (lachend).
151 I: Das heißt ihr wollt bei a und b bleiben?
152 S: Wir bleiben bei a und b, weil es nicht mehr relevant ist.
153 R: Ja. (…)

An dieser Stelle regt die Interviewerin an, die beiden Gruppen umzubenennen (133). Selina übernimmt sofort den Turn und die Anregung und schlägt eine Benennung im Kontext in „Fisch“ und „Normal“ vor (132). Hierzu ergeben sich mehrere Deutungsalternativen. Zunächst einmal erscheint dies einfach als Rückgriff auf die Gruppennamen, die im ursprünglichen Experiment vorlagen. Aus welchem Grund sie diese Namen vorschlägt, bleibt unklar. Es lässt sich interpretieren, dass hier die Schwierigkeit vorliegt, zwischen dem ursprünglichen Experiment und der Zuordnung der Werte auf neue Gruppen zu unterscheiden. Alternativ lässt sich deuten, dass der Wunsch vorliegt, oder durch die Interviewerin angeregt wurde, den beiden neuen Gruppen Namen zu geben, wissend, dass die Namen der neuen Gruppen nicht von Bedeutung sind, solange sie unterschieden werden können. Dies lässt sich an dieser Stelle nicht rekonstruieren, die schnell formulierte Umbenennung und die nachfolgende Aktivität (135), die dann unmittelbar wieder rückgängig gemacht wird, lässt jedoch darauf schließen, dass zunächst (vielleicht gedankenlos?) die Bezeichnung wie im ursprünglichen Experiment genommen werden sollte, ihr allerdings während des Beschriftens klar wird, dass eine Beschriftung mit den gleichen Gruppennamen nicht sinnvoll ist. Rebecca übernimmt nun den Turn und stellt die von Selina vorgeschlagenen Bezeichnungen in Frage (136). Wiederum bricht sie mitten im Satz ab und formuliert ihre eigene Unsicherheit bezüglich des Vorgehens („Aber ich bin nicht ganz sicher.“). Im Folgenden nimmt sie Bezug auf „den zweiten Schritt“, präzisiert jedoch nicht, was sie damit meint. Es lässt sich einerseits deuten, dass sie den zweiten Schritt aus dem Dreischritt 1-Nachdenken über die Daten, 2-Randomisieren und Wiederholen und 3-Schlussfolgerungen ziehen meint. Den Punkt 1-Nachdenken über die Daten haben die beiden abgeschlossen und befinden sich jetzt im Prozess des Randomisierens und Wiederholens. Der andere zweite Schritt, auf den sie Bezug nehmen könnte, ist der aus dem Schema zur Durchführung eines Randomisierungstests. Hier ist der zweite Schritt das Formulieren der Forschungs- und der Nullhypothese. Die Turn-by-Turn Analyse der nächsten Turns lassen den Schluss zu, dass Rebecca die Nullhypothese im Kopf hat, denn sie formuliert, mit Unterbrechungen von Selina, über die nächsten drei Turns: „kann es ja sein, dass wir“ (136) „eine Fischölperson mit einer Normalperson“ (138) „Zusammen.“ (140). Es lässt sich logisch ergänzen *in einer Gruppe erhalten*. Rebecca schließt den Turn (140) mit „Ich finde a und b ist eigentlich ganz neutral“. Selina lässt sich sofort überzeugen, denn bereits in (137) bestätigt sie mit „Stimmt, brauchen wir eigentlich nicht“ und folgt Rebecca in diesem Gedankengang offensichtlich, denn sie wirft bestätigende Kommentare in deren turnübergreifende Argumentation ein. In (141) ändert Selina ihre Gesprächsrolle vom Reagieren ins Agieren und zeigt, dass ihr die Unabhängigkeit des Wirkstoffs von den

erreichten Blutdruckwerten deutlich ist, denn sie formuliert: „Wir wollen ja jetzt nachweisen, dass es nichts mehr mit dem Öl zu tun hat.“ In (144) nennt Rebecca explizit den Bezug zur Nullhypothese, mit ihrer Wortwahl „Wir sind Vertreter der Nullhypothese“ macht sie deutlich, dass ihr bewusst ist, dass sie (temporär) eine spezielle Sicht für den Test einnimmt, nämlich die der Nullhypothese, beziehungsweise der von Selina zuvor ausgedrückten Unabhängigkeit vom eingenommenen Öl. Selina nähert sich im nächsten Turn weiter der Bedeutung der zwei neuen Gruppen und expliziert sehr gut, dass nun in eine der beiden Gruppen möglicherweise Personen aus beiden ursprünglichen Gruppen des Experiments gelangen können. Sie schließt mit „Es geht ja jetzt nur um die Blutdrucksenkung“ (145). Rebecca scheint dies auch noch einmal für sich formulieren zu wollen, denn sie beginnt in (146) ein Beispiel und führt in (148) fort, wie eine mögliche neue Gruppe zusammengesetzt sein kann. Hier lässt sich deuten, dass auch sie, nachdem sie in (136) ihre Unsicherheit zum Ausdruck gebracht hatte, noch einmal die Notwendigkeit sah, sich diesen Umstand klarzumachen. Das anschließende Lachen in den beiden Turns (149) und (150) von Selina und Rebecca scheint etwas Befreiendes zu haben, nachdem sie nun gemeinsam die Bedeutung der zwei neuen Gruppen erarbeitet haben. Sie kommen zum Schluss, dass sie bei den Bezeichnungen a und b für die neuen Gruppen bleiben, „weil es nicht mehr relevant ist“ (152). Möglicherweise meint Selina hier, dass die vorherige Gruppenzugehörigkeit nicht mehr relevant ist oder dass die Bezeichnung der neuen Gruppen aufgrund der zuvor erfolgten Diskussion nicht mehr relevant ist. Dies bleibt offen.

Durch eine Argumentationsanalyse ergibt sich hier folgende Struktur (Abbildung 203). Nachdem Selina eine Benennung in Fischöl und Normal vorgeschlagen hat (134), stellt Rebecca diese Umbenennung in Frage (135) und Selina formuliert dann den Ausgangspunkt für die nachfolgende Diskussion „Stimmt, brauchen wir eigentlich nicht“ (137). Dies lässt sich als Konklusion im Sinne der Argumentationsstruktur von Toulmin (1975) begreifen, die nachfolgend gezeigt werden soll. Die Notwendigkeit, Gruppenetiketten für die zwei neuen Gruppen zu vergeben, wird dabei als Datum betrachtet. Zur Begründung führen Rebecca und Selina an, dass durch die neue Zuteilung eine „Fischölperson mit einer Normalölperson“ (138) in einer neuen Gruppe landen könne. Dies wird gestützt durch die Unabhängigkeit vom Öl (141) und durch die Nullhypothese (144) und durch ein Beispiel (148). All diesen Begründungen liegt nun die Stützung zugrunde, dass die neue Zuteilung der Werte auf die Gruppen zufällig geschieht. Dies wird von den Teilnehmerinnen nicht explizit ausgesprochen, stellt aber die notwendige Stützung an dieser Stelle dar. Es wird davon ausgegangen, dass dies den beiden Teilnehmerinnen bewusst ist, da nur dadurch die von ihnen hervorgebrachten Stützungen sinnvoll sind.

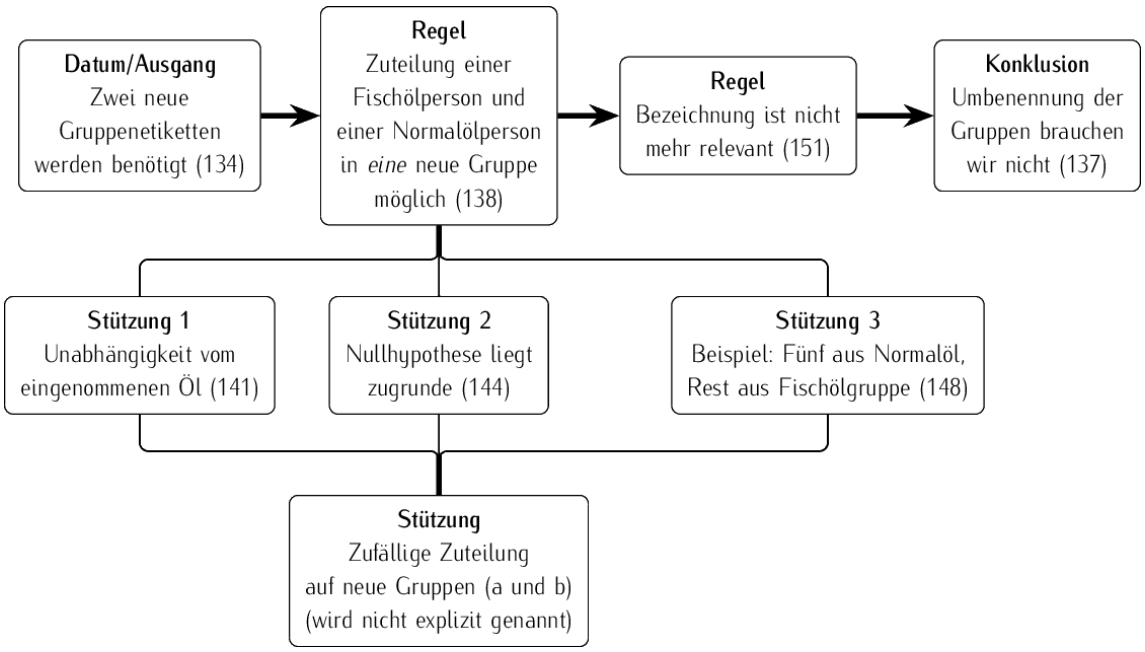


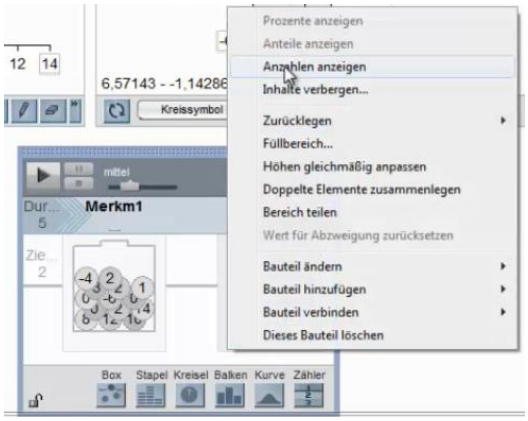
Abbildung 203: Rekonstruierte und ergänzte Argumentation von Rebecca und Selina zur Nichtumbenennung der beiden neuen Gruppen

Die vorgeschlagene Umbenennung der Merkmale war ein Argumentationsanlass für Selina und Rebecca. Dass die beiden solange über die Umbenennung sprechen, zeigt, dass sie an dieser Stelle eine Notwendigkeit empfinden, die Bedeutung der Gruppenetiketten auszuhandeln und sich durch diesen Anlass das Verfahren klar machen und miteinander aushandeln. Für den Verstehensprozess lässt sich schließen, dass dies eine der bedeutendsten Stellen des Gesprächs ist, denn hier machen sich die beiden klar, dass während des Simulierens eine zufällige Neuzuordnung der Personen mit den Blutdruckwerten unabhängig vom eingenommenen Öl geschieht.

Episode 8-5: Reproduktion der Gruppengrößen (Z. 154-181)

154 I: Ist in Ordnung. Ihr müsst jetzt nur noch das entsprechend so füllen, dass da auch jeweils sieben sind. Das geht über diesen kleinen Pfeil, (.) der da unten drunter ist. Da gibt es irgendwo Anzahl, ganz oben.

155 *Die Studentin wählt über den kleinen Pfeil unterhalb der Zufallsmaschine den Eintrag "Anzahl anzeigen"*



156 I: Ja. Da kannst du jetzt aus der Eins ne Sieben jeweils machen.

157 *Die Studentin ändert die Anzahl von 1 auf 7 in beiden Spalten des Bauteils.*

158 I: Genau. Ist klar warum?

159 R:Mhm (bejahend).

160 S: Ja klar, weil ja sieben Personen in jeder Gruppe sind.

161 I: Genau. Das soll reproduziert werden. (.) Dann kommt immer die Frage, die man sich beim Simulieren stellen muss: Soll mit oder ohne Zurücklegen gezogen werden?

162 S: Äh ja ohne.

163 I: Warum?

164 S: (lacht) Die Frage ist so simpel, weil ja jeder Wert nur einmal benutzt wird. Also/.

165 R: /Ja.

166 S: (lacht) Ich weiß nicht, wie man das jetzt schöner ausdrücken kann, aber/.

167 R: /Ja es könnte theoretisch sein, dass man zum Beispiel auch sieben Mal den Wert acht/

168 S: /Oder null hat. Genau.

169 R: Hätte.

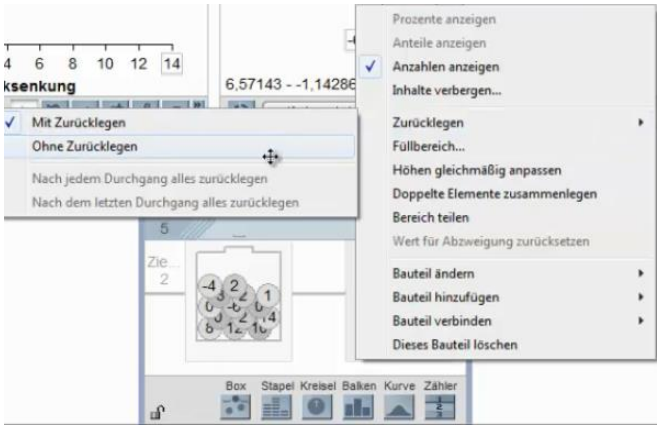
170 S: Und es müssen ja alle Werte vorkommen.

171 R: Ja genau.

172 S: Ja.

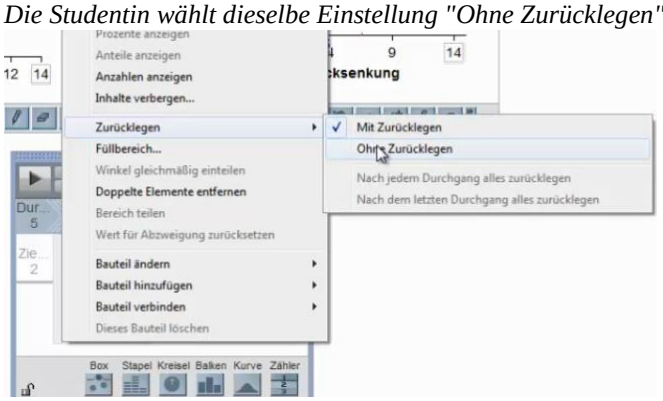
173 I: Ganz genau. Gut das heißt das müsst ihr jeweils noch einstellen. Das geht auch wieder mit dem kleinen Pfeil. Da gibt es Zurücklegen und da könnt ihr Einstellen ohne Zurücklegen. (.)

174 *Die Studentin klickt erneut den kleinen Pfeil an. Es erscheint der Eintrag "Zurücklegen". Dieser wird angeklickt und es erscheint die Einstellung "Ohne Zurücklegen". Diese wählt die Studentin aus.*



175 I: Und bei dem anderen dann auch noch.

176 *Die Studentin wählt dieselbe Einstellung "Ohne Zurücklegen" für das erste Bauteil aus.*



177 S: Ich lerne hier heute noch richtig was (lacht).

178 I&R&S: (lachend).

179 I: Ja für so was ist ein Interview auch da, dass ihr anschließend mit mehr Wissen rausgeht als ihr vorher hattet.

180 R: Ja wenn man nochmal einmal wiederholt, das ist echt gut.

181 S: Ja.

Die fünfte Episode schließt mit einer Pause, in der die beiden Studierenden nichts tun, weshalb die Interviewerin den technischen Hinweis gibt, dass nun noch die Gruppen so gefüllt werden müssen, dass jeweils sieben Elemente in den Stapeln a und b vorhanden sind. Anschließend erfolgt eine technische Anweisung hierzu und die Umsetzung durch die beiden Studierenden (154-157). Auf die Nachfrage, warum dies gemacht werden müsse, folgt der Dialog (159-172). Im Nachhinein betrachtet wäre hier ein Bezug der Interviewerin auf die in der Veranstaltung eingeführte Fabrikmetapher (Kapitel 8.1) sinnvoll gewesen, um den Vorgang zu erläutern.

Selina erklärt kurz auf Kontextebene, dass das daran läge, dass sieben Personen in jeder Gruppe seien (160). Hierauf übernimmt wieder die Interviewerin den Turn und formuliert als Regel hierzu, dass dies reproduziert werden solle (161). Auf die anschließende Frage, ob mit oder ohne Zurücklegen gezogen werden müsse, antwortet erneut Selina mit „Die Frage ist so simpel, weil ja jeder Wert nur einmal benutzt wird“ (164), bezugnehmend auf die erste Box der Zufallsmaschine. Rebecca bestätigt in (165) mit „Ja“,

wobei sich dies entweder darauf bezieht, dass die Frage so leicht ist oder inhaltlich darauf, dass jeder Wert nur einmal benutzt werden kann. Selina scheint sich in der Pflicht zu fühlen, dies noch weiter auszuführen, erklärt jedoch, dass sie dies nicht schöner auszudrücken wisse (166). Daraufhin übernimmt wieder Rebecca den Turn und fügt eine Erklärung wiederum in Form eines Beispiels an, die im Folgenden von beiden in satzergänzender Weise in den nächsten Turns ausgeführt wird (167-172).

Nachdem geklärt ist, dass es sich hier um ein Zufallsexperiment ohne Zurücklegen handelt, werden auf Softwareebene wieder technische Anweisungen durch die Interviewerin benötigt, um die Einstellungen vorzunehmen. Die Bedienung der Software durch die Studierenden funktioniert flüssig, nachdem die Interviewerin erklärt hat, wo die Einstellung zu finden ist (173-181). Die flüssige Bedienung lässt sich interpretieren, dass hier *instrumented action schemes* wieder aktiviert werden, die zuvor bereits vorhanden waren.

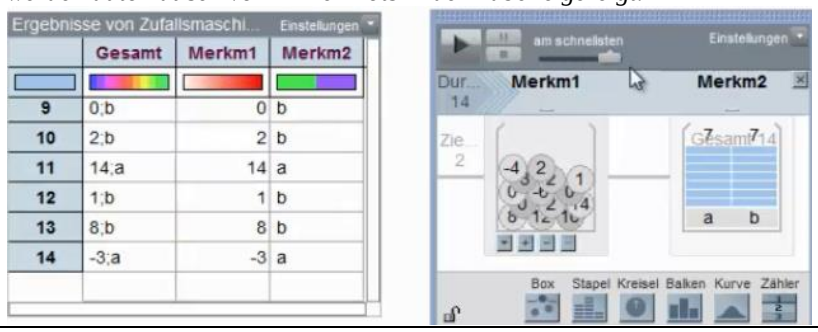
Episode 8-6: Durchgangsanzahl bestimmen und Zufallsmaschine laufen lassen (Z. 182-203)

182 R: So müssen wir noch was einstellen? Die Durchgangsanzahl/ (..)
183 S: Äh ja. Aus jedem Bauteil einmal.
184 R: Das ist sieben Mal.
185 I: Fast. (lacht)
186 S: Nee, vierzehn.
187 R: Ja, ja, ja klar.
188 S: Es sind ja vierzehn Teilnehmer. Nur ich hatte ja überlegt, wir müssen es ja mehrmals machen. Also erst mal vierzehn Mal, aber dann/.
189 I: Das ist das, was ihr vorhin gesagt habt.
190 S: Genau.
191 I: Da kommen dann die Messgrößen ins Spiel.
192 S: Genau.
193 R: Ja. Genau. Nur dass es hier nicht zehntausend Mal war, was man einstellen muss.
194 I: Richtig, denn nach vierzehn Mal ist die Urne leer.
195 R: Ja.
196 I: Dann werden nur noch keine Werte gezogen.
197 R: Ja.
198 S: Äh.
199 I: Äh einfach hierhin klicken, wo die fünf steht.
200 *Die Studentin klickt auf "Durchgänge", wo die 5 steht.*

201 S: Achso und jetzt vierzehn.
202 *Die Studentin ändert die Zahl 5 auf 14.*

203 I: Genau.
204 R: Ja klar. Dann nimmt der sich immer einen von diesen blauen Balken.
205 S: Genau, einen von den blauen.
206 I: Jetzt könnt ihrs einfach mal laufen lassen.
207 *Die Studentin startet die Zufallsmaschine.*

208 S: Achso dann könnten wir es ja ganz schnell machen. Wie der das macht wissen wir ja.
209 S: So.
210 *Die Geschwindigkeit der Zufallsmaschine wird auf „am schnellsten“ gestellt und die Ergebnisse werden automatisch von TinkerPlots in der Tabelle gezeigt.*



Diese Episode wird eingeleitet durch eine Frage von Rebecca, die zunächst die Software betrifft, nämlich, ob sie noch etwas einstellen müssen (182). Im direkten Anschluss gibt sie selbst die Antwort mit „Die Durchgangsanzahl“. Nach kurzer Unsicherheit wird mit Bezug auf den Kontext geklärt, dass hier die Anzahl 14 benötigt wird (Selina: „Es sind ja 14 Teilnehmer“ (188)), und dass weder sieben noch zehntausend die richtige Anzahl ist.

Nachdem inhaltlich geklärt ist, dass eine Durchgangsanzahl von 14 benötigt wird, wird nach einem Einwurf der Interviewerin auf Softwareebene „Einfach hierhin klicken, wo die fünf steht“ (199) die Durchgangsanzahl 14 eingestellt. Möglicherweise liegt hier ein semantischer Konflikt vor, weil die Studierenden intuitiv bei der Anzahl der Ziehungen in TinkerPlots die Stichprobengröße n = 14 realisieren möchten und über die Durchgangsanzahl dann die Wiederholung der Simulation. Durch die Programmierung von TinkerPlots (Anzahl der Bauteile legt die Anzahl der Ziehungen fest, so dass Messgrößen zur Wiederholung der Simulation verwendet werden müssen) ist dieses Vorgehen jedoch nicht möglich (vgl. Kapitel 4). Bevor die Zufallsmaschine laufen gelassen wird, klärt Rebecca für sich das Funktionsprinzip des Bauteils Stapel mit „Dann nimmt der sich immer einen von diesen blauen Balken“ (204). Wieder ist es so, dass sie ein Beispiel heranzieht, hier für einen Durchgang der Zufallsmaschine, um sich etwas zu erklären. Generell scheint ihnen die Funktionsweise der Zufallsmaschine klar zu sein, denn Selina formuliert in (208): „Achso dann könnten wir es ja ganz schnell machen. Wie der das macht wissen wir ja.“ Daraufhin stellen sie die Geschwindigkeit auf „am schnellsten“, so dass keine Animation mehr zu sehen ist, sondern die Tabelle mit den Ergebnissen der Simulation direkt von TinkerPlots angezeigt wird. An dieser Stelle zeigt die flüssige und selbständige Bedienung der Software, dass sie sich nun wieder sicherer fühlen im Umgang mit ihr. Es werden nicht mehr so viele Nachfragen und Anweisungen benötigt wie noch in den vorherigen Episoden. Das Gespräch findet auf Softwareebene statt.

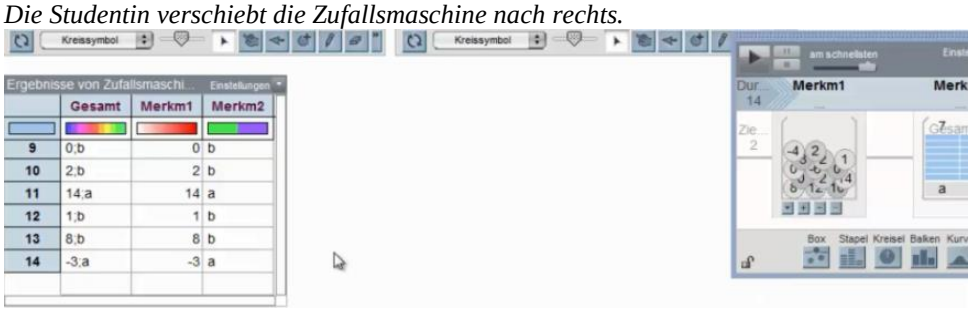
Interaktionseinheit 9: Teststatistik definieren und Referenzverteilung erstellen (Z. 212-430)

Die Konzepte Teststatistik und Referenzverteilung werden in der neunten Interaktionseinheit thematisiert. Diese Interaktionseinheit wird in neun Episoden geteilt. In der ersten Episode klären die beiden Teilnehmerinnen, was nun als nächstes zu tun ist. Die Darstellung der simulierten Ergebnisse in einer Graphik gestaltet sich als etwas schwierig und ist beinhaltet in Episode 2, die sich größtenteils auf der Ebene der Software abspielt. Am Ende dieser Episode werden die durchgeführten Aktivitäten erläutert. In Episode 3 wird die Differenz der arithmetischen Mittel mit Hilfe des Lineals in TinkerPlots berechnet, wobei sich einige Schwierigkeiten bei der Benutzung des Lineals ergeben, wie beispielsweise das korrekte Ziehen der Endpunkte über die beiden Mittelwerte. Die Teststatistik wird hier festgelegt, ohne dass sie von den Teilnehmerinnen so benannt wird. Anschließend folgt Episode 4, in der Messgrößen gesammelt werden. In der fünften Episode werden die gesammelten Messgrößen als Referenzverteilung dargestellt, wiederum wird der Begriff nicht von den Teilnehmerinnen verwendet. Hieran schließt sich die sechste Episode an, in der die Erwartungshaltung bezüglich der gerade erstellten Referenzverteilung thematisiert wird. In der siebten Episode wird das Randomisierungstestschema weiter ausgefüllt und in der achten Episode noch einmal ein sehr kur-

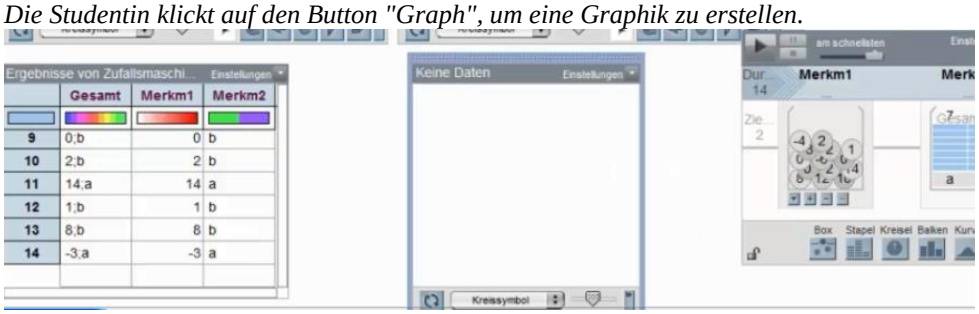
zer Rückbezug auf die Erwartungshaltung zur Referenzverteilung vorgenommen. In der neunten Episode wird das Schema bezüglich der Referenzverteilung ausgefüllt.

Episode 9-1: Planen der nächsten Schritte (211-222)

211 I: Jetzt habt ihr einmal randomisiert. Was macht ihr jetzt damit?
212 S: Jetzt müssen wir das ja nur/.
213 R: /Jetzt haben wir ja im Prinzip einmal die zufällige Zuordnung erreicht, die wir haben wollen. Aber wir brauchen das ja ziemlich häufig, um dann letztendlich ja wirklich diese Referenzverteilung zu haben, wo wir ganz viele (.) von diesen Mediandifferenzen haben. Ich weiß jetzt aber nicht mehr genau, ob wir jetzt hier zuerst eine Differenz vom Median oder vom arithmetischen Mittel in dem Fall einstellen müssen oder das Ganze erst über die Messgrößen (.).
214 I: Also das müsst ihr erst mal darstellen. Hier habt ihr jetzt so eine Tabelle erzeugt, wie sie das ursprüngliche Experiment hatte.
215 *Die Studentin verschiebt die Zufallsmaschine nach rechts.*



216 R&S: Ja.
217 I: Das heißt das müsst ihr jetzt erst mal darstellen, um überhaupt die Mittelwerte rauszubekommen.
218 S: Mhm (bejahend)
219 I: Ihr braucht also eine Graphik.
220 *Die Studentin klickt auf den Button "Graph", um eine Graphik zu erstellen.*



221 R: Ja.
222 S: Ah ok. (.) äh/

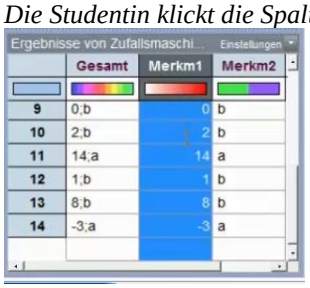
Nach der einleitenden Frage der Interviewerin (211) und einem abgebrochenen Turn von Selina (212) übernimmt Rebecca den Turn (213). Sie konstatiert zunächst, was sie bisher erreicht haben: „Jetzt haben wir ja im Prinzip einmal die zufällige Zuordnung erreicht, die wir haben wollen.“ Die nächsten Sätze von ihr zeigen zum Einen vorausschauendes Denken, wo die Aktivitäten hingehen sollen, auf der anderen Seite Unsicherheit, wie genau die nächsten Schritte auszuführen sind. Hier wird nicht klar, ob es sich um Unsicherheiten auf Ebene der Statistik oder auf der Ebene der Software handelt. Die letztere Deutung ist aufgrund des Wortes „einstellen“ eher wahrscheinlich. Als Ziel formuliert sie zunächst, dass sie die Referenzverteilung mit ganz vielen „von diesen

Mediandifferenzen“ anstrebt. Hier wird deutlich, dass ihr bewusst ist, dass der Prozess des Randomisierens häufig durchgeführt werden muss und dass das Ergebnis der Mittelwertdifferenzen in einer Referenzverteilung dargestellt wird. Anscheinend erinnert sie sich an den Baustein der Lehrveranstaltung, in welchem in zwei Beispielen (in der ersten Vorlesung und im Seminar) mit der Differenz der Mediane als Teststatistik gearbeitet wurde. Da bisher im gesamten Prozess weder von Rebecca und Selina noch auf dem Aufgabenblatt von Medianen die Rede war, kann dies nur so erklärt werden. Rebecca formuliert hierzu ihre Unsicherheit als „Ich weiß jetzt aber nicht mehr genau, ob wir jetzt hier zuerst eine Differenz vom Median oder vom arithmetischen Mittel in dem Fall einstellen müssen (...)“ und fügt sogar eine weitere Unsicherheit hinzu: „oder das Ganze erst über die Messgrößen“. Das letztere lässt sich interpretieren als Unsicherheit, wie das weitere Vorgehen in TinkerPlots gestaltet werden muss. Erneut erinnert sie sich vermutlich an einzelne Aspekte aus dem Baustein zur Inferenzstatistik, nämlich dass Messgrößen benötigt werden. Wie sie zu den Messgrößen kommt, kann ihr an dieser Stelle jedoch nicht klar sein, sonst hätte sie diese an dieser Stelle nicht erwähnt, da sie erst später verwendet werden können, wenn die entsprechende Differenz definiert ist. Als Anregung erwidert die Interviewerin, dass sie nun so eine Tabelle hätten, wie sie das ursprüngliche Experiment gehabt habe (214) und dass die Ergebnisse nun erst mal dargestellt werden müssten, um überhaupt Mittelwerte zu bekommen (216). Dabei wird von ihr bewusst vermieden, Median oder arithmetisches Mittel zu erwähnen. Nachdem Rebecca und Selina dies bestätigen, aber nicht tätig werden, weist die Interviewerin darauf hin, dass nun eine Graphik benötigt würde (219). Daraufhin wird eine Graphik erstellt und mit einem dem Tonfall nach erleichterten „Ah ok“ von Selina (222) begleitet. Hier lässt sich interpretieren, dass den beiden tatsächlich nicht klar war, was nach dem Laufenlassen der Zufallsmaschine der nächste Schritt im Prozess des Simulierens war und sie den Hinweis auf die Graphik benötigten, um weiterarbeiten zu können.

Es ergeben sich in dieser Episode große Schwierigkeiten, den weiteren Prozess in TinkerPlots durchzuführen, da den Teilnehmerinnen nicht klar ist, wie sie mit den simulierten Ergebnissen eines Durchgangs weiterarbeiten können. Das Gespräch findet hier auf Kontext-, Statistik- und Softwareebene statt.

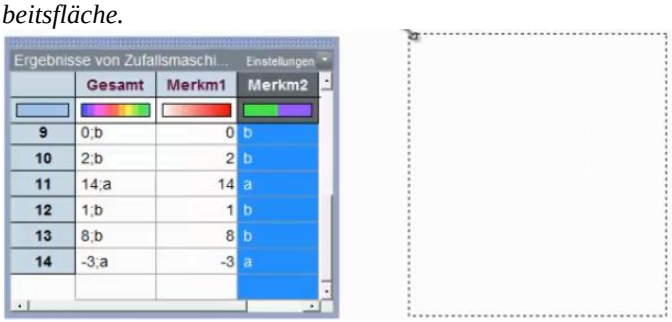
Episode 9-2: Darstellen der simulierten Ergebnisse (223-249)

223 *Die Studentin klickt die Spalte "Merkmal 1" an.*

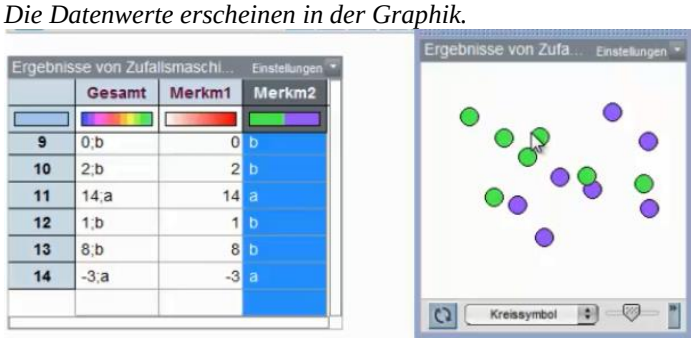


224 I: Wenn ihr Fragen habt, fragt. Sonst lass ich euch einen Moment machen.
225 S: Ja ich hab so lange nichts mit TinkerPlots gemacht.
226 I: Ist klar, ist lange her. (lacht)
227 S: Äh wie füge ich das denn nochmal in den? Ah/.
228 I: /Einfach oben auf den Merkmalsnamen klicken und rüber ziehen.
229 *Die Studentin klickt das Merkmal an und versucht es in die leere Graphik zu ziehen.*

230 S: Mach du das mal (lacht).
231 R: Gott. (..) hmm.
232 *Auch die zweite Studentin versucht es ohne dass etwas passiert.*
233 I: Warte mal. Mach mal den Graphen nochmal weg. Die Alternative wäre (..) ihr habt einfach das Merkmal markiert in der Tabelle und zieht euch dann den Graphen rein.
234 *Die Studentin markiert die Spalte "Merkmal 2" und zieht anschließend den Graphen in die Arbeitsfläche.*



235 R: Ja.
236 S: Achso und dann zeigt/.
237 I: /Genau und dann hat der automatisch die Datenwerte schon drin.
238 *Die Datenwerte erscheinen in der Graphik.*

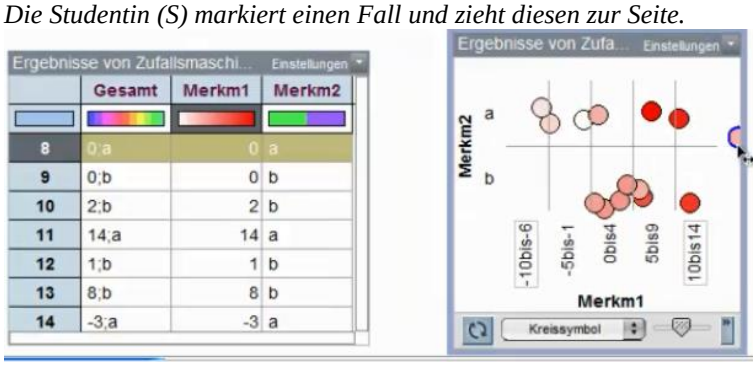


239 R: Ja.
240 I: Jetzt müsst ihr nur noch gucken, was gerade ausgewählt ist. Welches Merkmal, was ihr da habt, was ihr wie darstellen möchtet.
241 *Die Studentin klickt das "Merkmal 1" an. Die Farben der Punkte in der Graphik verändern sich.*

242 I: Kannst du ein bisschen erklären, was du da machst, während du das machst, also nur dass ich das nachher nachvollziehen kann.

243 S: Ja. Also ich hab jetzt gerade ähm (.) diese Kügelchen auseinandergezogen, damit die beiden Gruppen separat dargestellt werden, a und b. Und dann hab ich ähm das andere Merkmal eins, was dann ja die ähm (..) Werte angibt, äh reingezogen, damit ich die jetzt auch (..) auseinanderziehen kann. (..) Ist mir das jetzt geglückt?

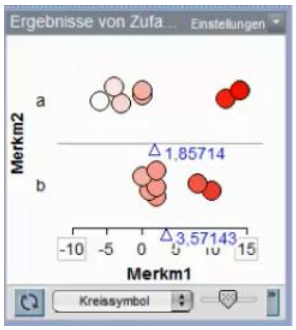
244 *Die Studentin (S) markiert einen Fall und zieht diesen zur Seite.*



245 S: (lacht) Ja genau. Das wird dann auf einer Achse dargestellt. Und mit dem Mittelwert äh/ nee beziehungsweise mit dem arithmetischen Mittel wurde ja hier gearbeitet. Dann kann ich das arithmetische Mittel der jeweiligen Gruppe äh (.) anzeigen lassen. Also das ist in TinkerPlots auch denkbar einfach (lacht).

246 I: (lacht)

247 *Das arithmetische Mittel wird eingeblendet.*



248 S: Man muss einfach nur auf den Button für arithmetisches Mittel klicken und Werte anzeigen und dann ja/ nee ja quasi sofort das arithmetische Mittel für Gruppe a. Das ist jetzt in diesem Fall 1,85714 und das arithmetische Mittel für Gruppe b ist 3,57143 (.) Und jetzt müssten wir eben die Differenz/.

249 R: /Genau.

Die Episode wird eingeleitet mit dem Turn der Interviewerin (224), in welchem sie zu verstehen gibt, dass die Studierenden nun erst einmal alleine arbeiten sollen, bei Fragen aber nachfragen können. Daraufhin übernimmt Selina direkt den Turn und führt auf

Softwareebene an, dass sie so lange nichts mit TinkerPlots gemacht habe (225). Dies lässt sich interpretieren zum einen als Erklärung, weshalb sie teilweise Schwierigkeiten mit der Bedienung der Software hat, zum anderen als Vorausschau, dass sie auch weiterhin auf Hilfe bei der Bedienung angewiesen ist. Die letzte Interpretation wird gestützt durch den übernächsten Turn (227), in welchem sie nachfragt, wie etwas einzufügen sei. Hiermit ist sicherlich gemeint, wie sie das Merkmal in den Graph einfügt, denn die zuvor erstellte Graphik beinhaltet keine Fälle, da zum Zeitpunkt der Aktivierung der Graphik weder die Zufallsmaschine noch die Tabelle markiert waren, so dass von TinkerPlots automatisch eine leere Graphik erstellt wurde. Dies führt nun zu Schwierigkeiten, da die technische Bedienung, wie Fälle in eine Graphik eingefügt werden können, ein Hindernis darstellt. Erläuternd lässt sich hierzu sagen, dass dies eine selten vorkommende Schwierigkeit ist, da meistens nach dem Simulieren die Tabelle markiert ist und somit die Fälle aus der Tabelle automatisch in einer neu erzeugten Graphik angezeigt werden. Mit Anweisungen der Interviewerin auf Softwareebene (233) gelingt es schließlich, das Merkmal 2 in die Graphik zu ziehen, so dass die Fälle dort angezeigt werden (230-239).

Nachdem die Fälle in der Graphik angezeigt werden, ist die Bedienung der Software wieder flüssig. Die Fälle sind nach Merkmal 2 eingefärbt und werden horizontal nach den Ausprägungen a und b getrennt (241). Da dies kommentarlos geschieht, bittet die Interviewerin darum, zu erläutern, was geschieht (242). Daraufhin erklärt Selina ihr Vorgehen. Dabei bindet sie sprachlich die Ebene der Software „diese Kügelchen auseinandergezogen“ (243) in die Ebene der Statistik „damit die beiden Gruppen separat dargestellt werden“. Auch für das erste Merkmal, in welchem die Blutdruckreduktionen stehen, zieht sie eine Verbindung zum Kontext: „Merkmal eins, was dann ja die ähm (..) Werte angibt“. Sie stockt kurz währenddessen, was zeigt, dass sie einen Moment nachdenkt, bevor sie angibt, was Merkmal 1 darstellt. Die Werte werden vollständig separiert (244) und in der nachfolgenden Erklärung wird ihr selbst bewusst, ob sie den Median oder das arithmetische Mittel darstellen soll, was zu Beginn der Interaktionseinheit nicht geklärt wurde. Sie argumentiert „mit dem arithmetischen Mittel wurde ja hier gearbeitet. Dann kann ich das arithmetische Mittel der jeweiligen Gruppe anzeigen lassen. Also das ist in TinkerPlots auch denkbar einfach.“ (245). Das „hier“ kann sich nur auf die Darstellung der Daten des ursprünglichen Experiments beziehen, denn an keiner anderen Stelle ist ein arithmetisches Mittel zu sehen. Nachdem sie die Darstellung erzeugt hat, die strukturell ähnlich zu der Darstellung der Daten des ursprünglichen Experiments ist, plant sie das Anzeigen der arithmetischen Mittel der beiden Gruppen und bewertet dies als „das ist in TinkerPlots auch denkbar einfach“. Dies wird hier daraufhin gedeutet, dass das Bedienen von TinkerPlots im Bereich der Datenanalyse wenig Schwierigkeiten darstellt und auch nicht als schwierig empfunden wird. In (248) wird

das technische Vorgehen von Selina hierzu genannt und am Ende des Turns wird die Differenz genannt, die als nächstes benötigt wird.

Episode 9-3: Berechnen der Differenz der arithmetischen Mittel (Z. 250-291)

Dieser Transkriptausschnitt beinhaltet viele technische Schwierigkeiten um die Bedienung der Linealfunktion in TinkerPlots. Es wird aus Platzgründen darauf verzichtet, das komplette Transkript darzustellen, die wesentlichen Inhalte werden zusammengefasst.

Als nächstes möchten die beiden auf Softwareebene die Differenz der arithmetischen Mittel in TinkerPlots berechnen (250). Hierzu bekommen sie den Hinweis von der Interviewerin, dass dies mit dem Lineal in TinkerPlots gemacht werden könne (251). Erneut zeigen sich Unsicherheiten in der Bedienung mit TinkerPlots, die zögerlich von Selina (253) und direkt von Rebecca „Boah, ich weiß das auch nicht mehr“ (254) formuliert werden. Sie beginnen trotz der ausgedrückten Schwierigkeit mit dem Einblenden des Lineals (255). Mit verbaler Unterstützung durch die Interviewerin werden die Enden des Lineals auf die arithmetischen Mittel zu ziehen versucht. Da die Graphik sehr klein ist, ist dies schwierig. Zunächst wird die Graphik vergrößert und mit weiteren Anweisungen die Enden des Lineals über den arithmetischen Mitteln platziert (264-289). Allerdings ist kein Ergebnis ablesbar, weil die durch das Lineal gemessene Differenz nicht zu sehen ist. Auf Anweisung der Interviewerin wird das Lineal erneut aus- und wieder eingeblendet. Bei der Bedienung der Linealfunktion ergeben sich hier viele Schwierigkeiten für die beiden Teilnehmerinnen.

Während der ganzen Episode findet die Kommunikation auf Ebene der Software statt. Erst am Ende, nachdem die Differenz der arithmetischen Mittel sichtbar ist, wechselt Rebecca kurz auf die Ebene der Statistik und interpretiert das Ergebnis als „Genau das ist dann hier die Differenz der arithmetischen Mittel, wenn wir jetzt (.) das erste Mal eben diese neue Gruppenzuordnung gemacht haben“ (291). Es kann interpretiert werden, dass sie durch diesen expliziten Bezug für sich selbst das Vorgehen zusammenfasst bzw. ihr Verständnis über die durchgeführte Aktivität ausdrückt. Hieran lässt sich auch erkennen, dass ihr der statistische Hintergrund der Aktivitäten in TinkerPlots klar ist und die Schwierigkeiten auf der Softwareebene und nicht auf der Statistikebene liegen.

Nahezu die gesamte Episode spielt sich auf der Softwareebene ab und beinhaltet Kommunikation um die Bedienung und deren Schwierigkeiten in TinkerPlots. Die Studierenden berechnen in dem Datensatz die Mittelwertdifferenz, für die die Software eine Auswertungsfunktion definiert, die eine funktionale Abhängigkeit von drei Zellen konstruiert. Durch diese „Hintergrundarbeit“ der Software wird die Teststatistik erzeugt, die jedoch als Fachbegriff bisher von den beiden Teilnehmerinnen nicht erwähnt wurde. Im Baustein Inferenzstatistik wurde gelernt, dass sich die Mittelwerte bei erneuter Simulation automatisch aktualisieren, und die Messgrößenfunktion übernimmt daraufhin das

Sammeln der unterschiedlichen Differenzen für die Referenzverteilung. Inwieweit sich die Studierenden von diesem Prozess ein Bild machen, kann aus der Kommunikation nicht genau rekonstruiert werden.

Episode 9-4: Messgrößen sammeln (Z. 292-346)

Diese Episode wird eingeleitet auf Softwareebene von Selina mit „Und das wollen wir jetzt häufig durchführen, das heißt das [Mittelwertdifferenzen] wollen wir sammeln“ (292), bezugnehmend auf die gerade ermittelte Differenz der arithmetischen Mittel der beiden neuen Gruppen. Wieder ist das Vorgehen klar, die auszuführenden Schritte in TinkerPlots werden jedoch nicht durchgeführt, so dass die Interviewerin eingreift und eine Anleitung gibt, wie die Messgrößenfunktion zu verwenden ist (294-299). Nachdem die Differenz der arithmetischen Mittelwerte als Messgröße gesammelt wurde und in einer neuen Tabelle von TinkerPlots automatisch protokolliert wurde, ist die Funktionsweise auf Softwareebene jedoch nicht klar, wie der folgende Transkriptausschnitt zeigt.

300	S: Und jetzt müssen wir das halt noch ein paar Mal durchführen, ne?
301	I: Ja. Aber bevor du es durchführst. Erst mal wie häufig wollt ihrs durchführen?
302	R: Zehntausend Mal?
303	S: Aber müssen wir (unv.) also händisch den äh/ also nicht händisch, aber wir müssen danach ja immer den einzelnen Wert dann wieder eintragen.
304	R: Ich glaube der mach das auto/, kann man das nicht hier einstellen?
305	I: Ja genau.
306	R: Dass du da dann hast, wie oft du das sammelst und dann sammelt der das für sich selbst und erstellt dir dann nachher eine Referenzverteilung wo die einzelnen Punkte schon immer die Differenz (.) des arithmetischen Mittels sind.
307	S: Dann mach mal. Ja super, perfekt. Dann äh stell das mal ein (lacht).

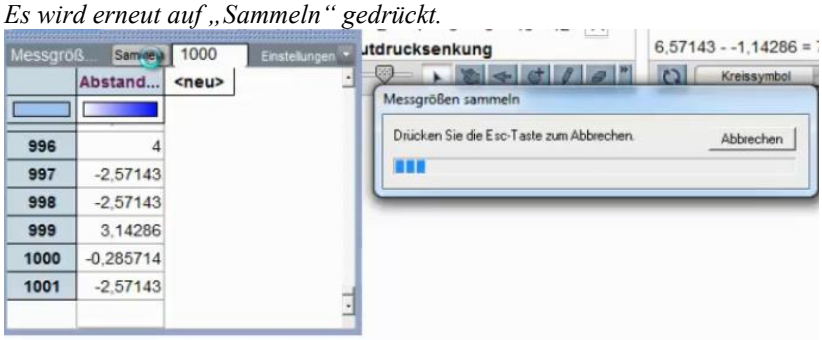
Selinas Turns in (300) und (303) lassen die Vermutung zu, dass sie denkt, dass für jede einzelne Durchführung der Simulation der Wert der mit dem Lineal gemessenen Differenz der arithmetischen Mittel von Hand in die Tabelle eingetragen oder vielleicht auch als Messgröße neu definiert werden muss. Rebecca erläutert daraufhin in (304), dass dies „auto“ (vermutlich automatisch) von TinkerPlots gemacht werde. In (306) führt sie das aus als „dann sammelt der das für sich selbst“. Ohne eine solche vor Augen zu haben, erläutert sie auch direkt die Bedeutung der einzelnen Punkte in der Referenzverteilung (306), was erneut ihr gutes Verständnis bezüglich des Prozesses und des Bezugs zur statistischen Ebene zeigt.

Als Wiederholungsanzahl für die Simulation wird „Zehntausend Mal?“ (302) von Rebecca fragend in den Raum geworfen, ohne dass dies an dieser Stelle weiter geführt wird.

Bevor es zum Sammelvorgang kommt, greift die Interviewerin erneut ein mit dem Hinweis darauf, dass der Rechner mit zehntausend Mal überfordert sei, weil er gleichzeitig aufnähme und schlägt stattdessen eine Wiederholungszahl von eintausend vor, was häu-

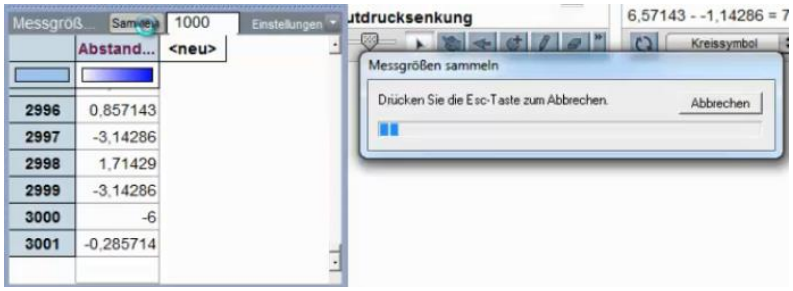
figer wiederholt werden könne, um auf die gewünschten zehntausend zu kommen (309). Nach der Eingabe von eintausend im entsprechenden Feld der Messgrößentabelle greift sie erneut ein, um die Aktionen durchzuführen, die zur Beschleunigung des Sammelprozesses hilfreich sind. Von ihr werden die Elemente Zufallsmaschine, Tabelle zur Zufallsmaschine und Graphik zur Zufallsmaschine minimiert und im Einstellungsmenü zu den Messgrößen die Animation ausgeschaltet (312-325). Die Studierenden klicken auf Sammeln und der Vorgang ist in Sekundenschnelle abgeschlossen. Daraufhin schließt sich der folgende Transkriptausschnitt an.

326 I: Das dauert jetzt einen Moment. Ja ok wir sehen das geht relativ schnell, also du kannst das zehn Mal machen, um auf die gewünschten zehntausend zu kommen, wenn ihr es gerne zehntausend Mal machen möchtet.
327 S: Nein, ist schon ok.
328 I: Also ich wollte euch nicht davon abhalten es zehntausend Mal/ wir haben ja jetzt gesehen, eintausend geht ganz schnell/ also wenn du jetzt noch neun Mal draufklickst, dann haben wir es zehntausend Mal.
329 R: Soll ich nochmal einmal?
330 I: Es ist eure Simulation.
331 S: Ach es ist ja eigentlich, also/.
332 R: /Ich würd es noch ein paar Mal mehr als tausend vielleicht machen damit man so ein bisschen/.
333 S: /Ja dann.
334 R: Einfach dann nochmal drauf?.
335 I: Nochmal auf Sammeln genau.
336 *Es wird erneut auf „Sammeln“ gedrückt.*

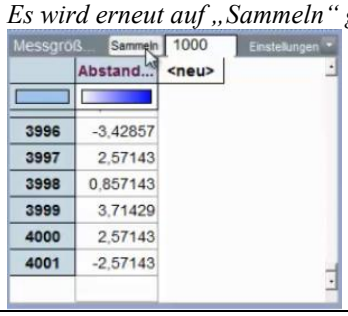


337 S: Ich meine, so wird der Wert präziser ja schon.
338 R: Ja genau.
339 *Es wird erneut auf „Sammeln“ gedrückt.*

340 S: Und nochmal. Wir treffen uns bei fünf, ne?
341 R: Ok. Machen wir.
342 I& R& S: (lachen)
343 *Es wird erneut auf „Sammeln“ gedrückt.*



344 S: Und einmal noch, ne?
345 R: Ja.
346 *Es wird erneut auf „Sammeln“ gedrückt.*



In (329) fragt Rebecca „Soll ich nochmal?“ und meint damit wohl, auf den Sammel-Button klicken. Selina scheint etwas unentschlossen zu sein und das nicht für nötig zu halten (331), woraufhin Rebecca entgegnet, dass sie es noch „ein paar Mal mehr als tausend“ machen würde „damit man so ein bisschen“ (332). Leider führt sie den Satz nicht zu Ende. Selinas Tonfall daraufhin ist immer noch leicht unentschlossen, aber sie lenkt ein mit „Ja dann“ (333). Nachdem eintausend weitere Messgrößen gesammelt wurden, bemerkt Selina korrekt, aber ungenau: „Ich meine, so wird der Wert präziser ja schon.“ (337), was im nächsten Turn von Rebecca bestätigt wird. Hier kann sie entweder das Konzept der Genauigkeit von Simulationen im Kopf haben, welches im dritten Baustein der Lehrveranstaltung thematisiert wurde und dazu die typische Vorstellung haben, dass je häufiger simuliert wird, desto genauer die Ergebnisse sind. Es ist auch nicht klar, was sie mit „der Wert“ meint, denn es wird ja nicht ein einzelner Wert gesucht, sondern eine Verteilung. Aber hier geht keine von Beiden weiter darauf ein. Nach den nächsten eintausend Messgrößen schlägt Selina vor „Wir treffen uns bei fünf, ne?“ (340), womit sich Rebecca einverstanden erklärt: „Ok. Machen wir“ (341). Es scheint an dieser Stelle kein genaues Konzept vorzuliegen, an welchem sie sich orientieren, um die Anzahl der zu sammelnden Messgrößen festzulegen. In den Vorlesungen des Bausteins Inferenzstatistik wurde mit n = 10000 Wiederholungen gearbeitet, im Seminar mit explizitem Hinweis auf die unterschiedlichen Rechnerleistungen mit n = 1000 mit Hinweis auf die daraus möglicherweise resultierende Abweichung von ±3% (vgl. Kapitel 8.1). Letztlich werden 4001 Messgrößen gesammelt.

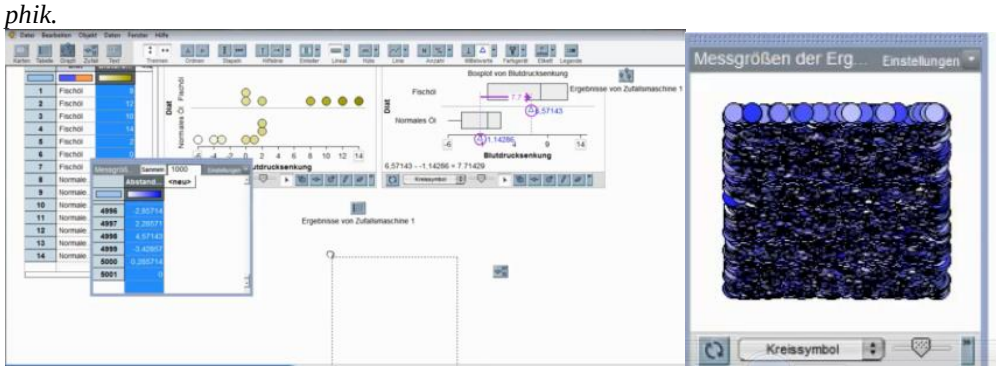
In Zeile (347–361) findet ein Exkurs statt über den allgemeinen Einsatz von Software, in welchem beide mitteilen, dass sie bisher nur mit den Programmen Word und Power-

Point gearbeitet haben. Da dieser irrelevant für die weitere Bearbeitung ist, wird dieser nicht analysiert.

Episode 9-5: Darstellen der Referenzverteilung (Z. 362-373)

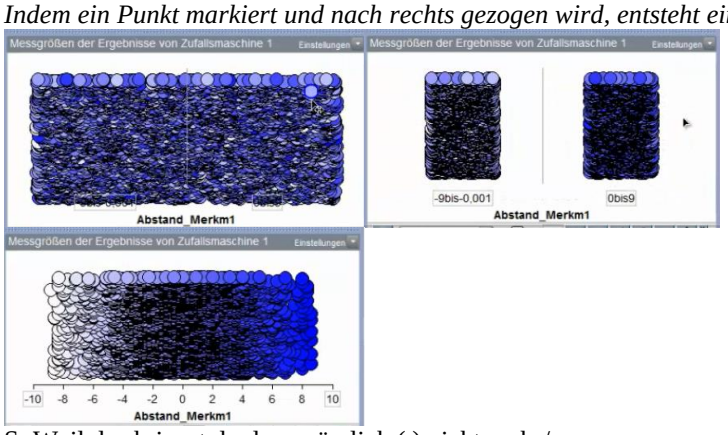
362 S: Echt, man lernt da einiges. Ok und jetzt müssen wir das Ganze ja wieder in eine.

363 *Mit Hilfe der Markierung und der Betätigung des Buttons „Graph“, erscheint eine neue Graphik.*



364 S: Machen wir das mal gleich so. (lacht) Genau jetzt weiß ich auch wieder, wieso ich mit einer kleinen Wiederholungszahl angefangen habe, weil ich war immer total überfordert, wenn da so viele/.

365 *Indem ein Punkt markiert und nach rechts gezogen wird, entsteht eine genauere Einteilung.*



366 S: Weil das kriegst du dann nämlich (.) nicht mehr/.

367 R: Genau jetzt musst du/.

368 S: /(unv.) müsste schon.

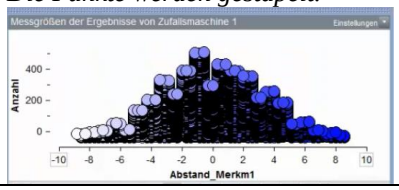
369 I: Jetzt musst du noch stapeln, damit du auch was erkennen kannst.

370 *Anschließend wird der Button „Stapeln“ betätigt.*

371

372 S: Ah guck mal (.) ah schön.

373 *Die Punkte werden gestapelt.*



Aus den gesammelten Messgrößen erstellen die beiden selbständig eine Referenzverteilung, die nach einem Hinweis auf Softwareebene der Interviewerin auf die Funktion des Stapelns gestapelt wird, so dass die Verteilung aus Abbildung 204 erzeugt wird.

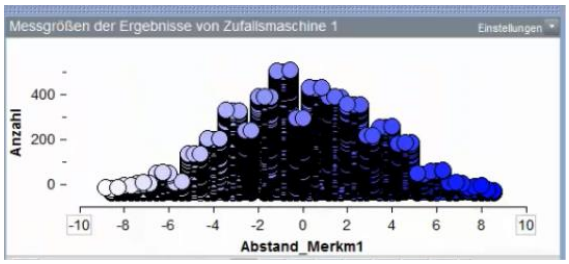


Abbildung 204: Von Rebecca und Selina erstellte Referenzverteilung

Die Episode endet mit „Ah guck mal (.) ach schön“ (372) von Selina. Entweder ist sie erfreut, dass die Referenzverteilung ohne Schwierigkeiten erstellt wurde, oder ihr gefällt, wie die Referenzverteilung aussieht.

Weder auf Statistik- noch auf Kontextebene findet in dieser Episode Kommunikation statt.

Episode 9-6: Erwartungshaltung zur Referenzverteilung (Z. 374-389)

374 R& S: (lachen)
375 I: Ah danke, gut dass ich diese Begeisterung aufnehme (lacht).
376 S: Ja, wenn man dann sieht, dass es ungefähr so aussieht, wie es aussehen soll. Weil spätestens an dem Punkt sieht man dann auch, wenn man was falsch gemacht hat, weil wenn man zumindest so ein bisschen (.).
377 I: Warum?
378 S: Wenn man so ein bisschen den Grundsatz verstanden hat und äh da stellt sich jetzt irgendwas dar, was irgendwie überhaupt nicht logisch sein kann, dann weiß man, dass man entweder die Zufallsmaschine falsch eingestellt hat. Und dass ich quasi was ganz anderes simuliert habe, wie ich wollte. Weil so ein bisschen kann man das Ergebnis ja erahnen und dann sieht man spätestens an dem Graphen, dass hier jetzt erst mal falls was falsch gemacht wurde.
379 R: Mhm (bejahend)
380 I: Ok da würde ich gerade gerne nochmal einhaken. Das heißt, dass du eine gewisse Erwartung gehabt hast, bevor du quasi diese Verteilung irgendwie gesehen hast?
381 R: Ja man rechnet ebenso damit. Also natürlich hat man die Forschungshypothese und die (..) Nullhypothese, weil ich finde schon man denkt schon eher immer so ein bisschen eher in dieser Forscherebene und möchte das dann doch irgendwie für richtig beweisen, dass eben ähm diese Blutdrucksenkung ähm/ also dass dieses Fischöl irgendwo wirksam ist. Und wenn man so was dann irgendwo sieht, dann freut man sich, weil man nachher weiß, wenn ich diesen ähm Wert von 7,7 dazu in Bezug setze, dann weiß ich, dass nicht mehr viele Werte gleich oder extremer sind. Und ich finde (.) ja man denkt irgendwo schon eher in diesem Forscherdenken und möchte eben die Nullhypothese gerne ablehnen.
382 S: Ja und wenn sich das dann komplett anders dargestellt hätte, dann hätte man sich erst mal gefragt, ob man erst mal irgendwas bei einer Einstellung falsch gemacht hat. Ne also dass man äh eventuell (.) dann noch auf Spurensuche geht, weil das/.
383 R: /Ja stimmt, obwohl es ist ja interessant, es könnte ja auch theoretisch sein, dass die Nullhypothese absolut wahr ist. Und man fragt sich irgendwie doch im ersten Moment, falls das mit der Forschungshypothese nicht ganz passt, ob man irgendwie was falsch eingestellt hat.
384 S: Ja, ja könnte.

385 I: Und was sind jetzt Kriterien, an denen ihr festmacht, dass ihr sagt: (.) das sieht so ungefähr aus wie das was ich erwartet habe, das passt? An was macht ihr das konkret fest?

386 R: Wir haben dann ja immer/ im Prinzip geht es jetzt ja auch darum, wir haben ja jetzt hier diese ganzen ähm (.) ja fünftausend Mal eben dieses arithmetische Mittel. Also die Differenz von den beiden aus diesen neuen Gruppen, die wir eben haben und dazu (.) müssen wir ja jetzt im Grunde unsere Ursprungsgruppe, wo es wirklich die Fischölgruppe und die normales Öl Gruppe gab (.). Eben da diese Differenz des arithmetischen Mittels, diese 7,7 in Bezug zu dem Ganzen setzen und ähm um dann eben diesen P-Wert/.

387 S: /Genau.

388 R: Als Maßstab nehmen, wie viele Ergebnisse. Denn wenn man in der Nullhyp/ wenn man von der Nullhypothese ausgeht und schaut, welche genauso sind, wie der Ursprungswert oder eben noch extremer. Und dann will man ja einen relativ kleinen p-Wert haben und man kann ja hier (zeigt auf die Referenzverteilung) sich schon vorstellen, wenn ich hier die 7,7 einblende, dann ist da nicht mehr viel was gleich oder extremer ist. Also scheint wohl (.) meine Forschungshypothese wahrscheinlich angenommen werden zu können.

389 S: Ja.

An dieser Stelle fügt Selina an „Ja, wenn man dann sieht, dass es ungefähr so aussieht, wie es aussehen soll. Weil spätestens an dem Punkt sieht man dann auch, wenn man was falsch gemacht hat, weil wenn man zumindest so ein bisschen (.)“ (376). Hier lässt sich interpretieren, dass sie eine Erwartungshaltung zur Referenzverteilung hat, die sie nun durch die erzeugte Referenzverteilung (373) bestätigt bekommen hat. Gleichzeitig führt sie hier aus, dass sie über eine Evaluierungsstrategie verfügt, denn scheinbar nutzt sie die entstandene Referenzverteilung im Abgleich mit ihrer Erwartungshaltung dazu, ihr Vorgehen auf Fehler zu analysieren. Als mögliche Fehlerquelle führt sie in (378) die Zufallsmaschine an. Sie führt jedoch bis hierhin nicht aus, an welchen Kriterien sie festmacht, ob die Referenzverteilung zu ihren Erwartungen passt. Sie lässt einen weiteren Einblick zu durch die Aussage „Weil so ein bisschen kann man das Ergebnis ja erahnen“ (378). Auf eine Nachfrage hierzu von der Interviewerin führt nun Rebecca ihre Erwartungshaltung genauer aus. Offensichtlich hat sie ebenso eine bewusste Erwartungshaltung zur Referenzverteilung wie Selina. Folgende Punkte führt sie dabei aus. Sie beginnt mit dem Rückgriff auf die Begriffe Forschungs- und Nullhypothese und positioniert sich auf Seite der Forscher. Hierzu drückt sie aus, dass man diese Perspektive einnähme, womit der Wunsch einhergehe, dass man „für richtig beweisen“ möchte, dass die durchgeführte Maßnahme, hier das Fischöl, wirksam sei (381). Ohne dass der P-Wert bisher erwähnt, geschweige denn geschätzt wurde, führt sie aus, dass nicht viele Werte gleich oder extremer als die beobachteten 7,7 sein werden. Als Ziel nennt sie schließlich, dass man die Nullhypothese gerne ablehnen möchte. Dies lässt einen tiefen Blick in ihre kognitiven Prozesse zu. Sie zeigt hier auf hohem Niveau ein Verständnis über den Sinn des Randomisierungstests. Während Selina in (382) erneut darauf hinweist, dass man bei einer abweichenden Darstellung auf Ebene der Software nach Fehlern suchen würde, übernimmt Rebecca erneut den Turn und bringt wieder die Ebene der Statistik hinein. Sie zeigt, dass sie weiß, dass es „ja theoretisch auch sein [könnte], dass die Nullhypothese absolut wahr ist“ führt dies jedoch wie Selina zuvor daran aus,

dass man in diesem Fall noch einmal prüfen würde, „ob man irgendwie was falsch eingestellt hat“ (383). In (385) fragt die Interviewerin explizit nach Kriterien, an denen die Passung der Referenzverteilung mit den eigenen Erwartungen beurteilt werde. Rebecca antwortet hierauf in (386) erst mit der Erklärung, dass in der Referenzverteilung die Differenzen der arithmetischen Mittel der beiden neuen Gruppen zu sehen sei und diese mit den 7,7 in Bezug gesetzt werden müssten. Als Maßstab zieht sie den P-Wert heran. Auf statistischer Ebene sehr gut bezieht sie sich auf die Nullhypothese als Ausgangspunkt dafür, zu schauen, wie viele Werte gleich oder extremer als der beobachtete Wert seien. Anschließend formuliert sie als Ziel, dass man einen „relativ kleinen P-Wert“ haben möchte. Auf die Referenzverteilung deutend, erklärt sie nun, dass „da nicht mehr viel [ist] was gleich oder extremer ist. Also scheint die Forschungshypothese wahrscheinlich angenommen werden zu können“ (388). Diese Aussage lässt rückblickend in der Turn-by-Turn Analyse den Schluss zu, dass sie dies von Beginn an im Blick gehabt haben könnte und somit ein hohes Maß an Verständnis für den statistischen Teil des Randomisierungstests zeigt.

In einer vertiefend durchgeführten Argumentationsanalyse ergibt sich die Struktur aus Abbildung 205 für die Argumentation zur Erwartungshaltung bei einer Referenzverteilung von Rebecca und Selina.

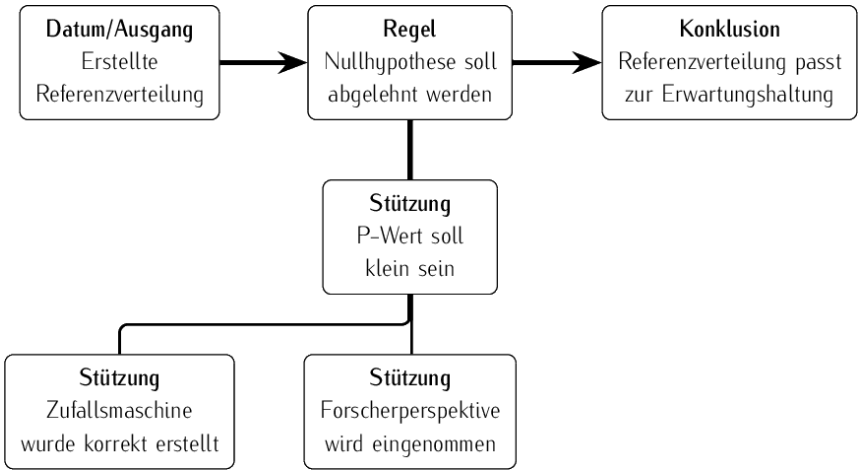


Abbildung 205: Rekonstruierte Argumentation von Rebecca und Selina zur Passung zwischen ihrer Vorstellung von der Referenzverteilung und der tatsächlich dargestellten Referenzverteilung

Als Ausgangspunkt wird die erstellte Referenzverteilung genommen. Über die Regel „die Nullhypothese soll abgelehnt werden“ mit den Stützungen, dass ein kleiner P-Wert erreicht werden soll und dass die Zufallsmaschine korrekt erstellt wurde, schließen die Beiden darauf, dass die Referenzverteilung zu ihrer Erwartungshaltung passe. Alternativ lässt sich diese Episode deuten, dass sie nicht die gesamte Verteilung in den Blick nehmen, sondern nur auf die Lage von 7,7 schauen. Sie argumentieren hierzu, dass wenn es

so aussieht, als würde die Nullhypothese nicht verworfen werden können, das Vorgehen überprüft werde mit der Begründung, dass ein Interesse bestehe, die Nullhypothese zu verwerfen. Diese Interpretation wird gestützt dadurch, dass sie nicht darüber argumentieren, dass die realen Daten dazu führen, dass die Forschungshypothese vermutlich richtig ist. Kommunikation findet in dieser Episode auf Kontext-, Statistik- und Softwareebene statt.

Episode 9-7: Ausfüllen des Schemas zur Simulation (Z. 390-395)

Im Folgenden wird das Schema zur Durchführung eines Randomisierungstests weiter ausgefüllt. Ohne weitere Diskussion wird zu Punkt 3 (Simulation) notiert: „5.000 mal zufällige Zuordnung zu den zwei neuen Gruppen A & B“.

Episode 9-8: Rückbezug zur Erwartungshaltung zur Referenzverteilung (Z. 396-401)

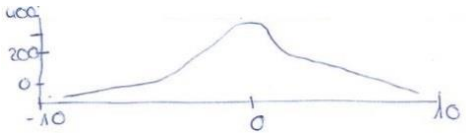
396	S: Also ich bin gerade froh, dass ich nicht die Einzige bin, die so denkt/ also die das dann so erklärte.
397	I: (lacht)
398	S: Denn ich dachte, ich wäre die einzige, die so denkt. Naja, wenn der Wert nicht so ist wie er sein soll.
399	R: Obwohl es ja theoretisch auch sein könnte, dass die Nullhypothese nicht/.
400	S: /Ja.
401	R: Aber man hat dieses Denken irgendwie so.

Im Anschluss an das Ausfüllen des Schemas schließt sich vor dem strukturierten Weiterarbeiten eine kurze Episode auf Statistikebene an, in der von Selina noch einmal ein Rückbezug zur Erwartungshaltung bezüglich der Referenzverteilung vorgenommen wird (396), womit sie die Turns aus der sechsten Episode von Rebecca im Nachhinein zusätzlich bestätigt und das somit vorgebrachte Wissen als gemeinsames Wissen kategorisiert. Interessant ist hier noch Rebeccas letzte Aussage in (401), dass man dieses Denken irgendwie so „habe“, womit sie vermutlich meint, dass man den Wunsch verspürt, die Nullhypothese nach einem durchgeführten Test ablehnen zu können. Dies lässt sich aufgrund ihrer vorigen Aussage in (399) interpretieren und stützt die in Abbildung 205 ausgeführte Argumentation.

Episode 9-10: Ausfüllen des Schemas zur Teststatistik

402	R: Ja und bei: Welcher Wert soll überprüft werden, da soll ja im Prinzip wieder die Differenz des/ der arithmetischen Mittel bei den neuen Gruppen a und b genommen werden. Und dazu soll dann der/ die ursprüngliche Differenz in Bezug (..) gesetzt werden.
403	S: Ja.
404	*Es wird notiert: „-Differenz der ar. Mittel der neuen Gruppen A & B => dazu dann den Ursprungswert (7,7) in Bezug setzen.“*
405	S: Und wie sieht die Referenzverteilung aus?
406	R: Das fand ich da schon schwierig zu beschreiben.
407	I: Geht ja einfach darum eine grobe Skizze zu machen.
408	R: Ja.
409	I: Einfach nur quasi von wo bis wo oder so. Also dass in einer groben Skizze einmal abmalen, dass wenn man anschließend nur noch diesen Zettel vorliegen hätte, dass man trotzdem nach-

vollziehen kann.
410 S: Achso man kann einfach eine Skizze machen?
411 I: Genau.
412 S: Oder man muss das gar nicht mit Worten beschreiben.
413 I: Nein.
414 S: Achso ja dann kann es ja/.
415 R: /Soll denn dann auch die Skalen mit?/
416 S: Ja mach einfach minus zehn bis plus zehn.
417 R: Ja genau.
418 S: Und also (.) und dann würde ich da einfach äh/.
419 S: /Genau und dann machst du einfach so (unv.).
420 R: Gut, in so was bin ich nicht gut.
421 S: Ja soll ja nicht schön sein (lacht), soll ja nur eine Skizze sein.
422 R: Ja nicht so krass aber/.
423 S: /Ja doch das trifft es ja schon ganz gut.
424 R: Ungefähr (lacht).
425 S: Ja also Minimum (unv.) Maximum.
426 S: Ich kann ja hier nochmal, ne/.
427 R: Null.
428 S: Genau. Damit man das so ein bisschen einfacher erkennen kann.
429 R: Ja. Das ist dann hier ungefähr die Vierhundert.
430 *Die folgende Verteilung wird skizziert:*



Das Gespräch wendet Rebecca nun wieder dem Ausfüllen des Schemas zu. Sie formuliert auf Statistikebene zunächst die Teststatistik korrekt als „Differenz der arithmetischen Mittel bei den neuen Gruppen a und b“ und zeigt erneut vorausschauendes Denken mit dem Hinweis, dass dazu „der/die ursprüngliche Differenz in Bezug (.) gesetzt werden [soll]“ (402). Die Teststatistik wird auf dem Schema notiert und Selina stellt nun die Frage des Schemas „Und wie sieht die Referenzverteilung aus?“ (405). Erneut formuliert hierauf Rebecca ihre Schwierigkeiten in (406) mit „Das fand ich da schon schwierig zu beschreiben“. Mit dem Hinweis der Interviewerin, dass eine grobe Skizze gemacht werden solle (407), was in (409) etwas präzisiert wird, wird nun Selina in den folgenden Turns erst deutlich, welche Anforderung in dem Schema hier gestellt wird. Es klärt sich, dass sie dachte, dass die Referenzverteilung in Worten beschrieben werden solle (412). Von (414) bis (430) handeln sie auf technischer Ebene aus, von wo bis wo die Skala der Skizze genommen werden soll und produzieren schließlich die Skizze in (430). Das Randomisierungstestschema enthält keine explizite Aufforderung, den beobachteten Wert einzuzeichnen, um dessen Lage qualitativ zu dokumentieren, sondern nur die Frage *Wie sieht die Referenzverteilung aus?*

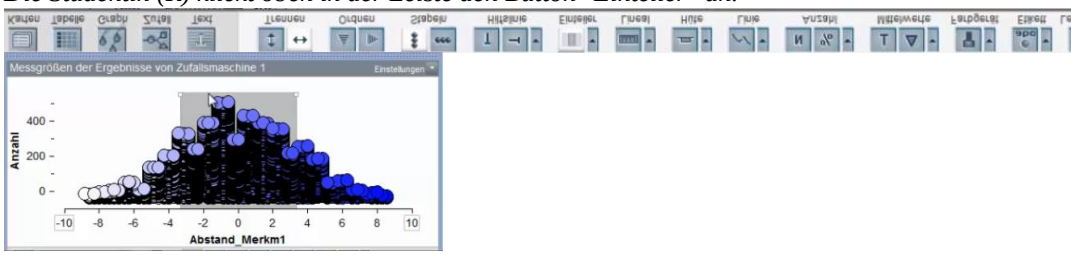
Interaktionseinheit 10: Ermitteln des P-Werts (Z. 431-454)

Das Konzept des P-Werts ist Inhalt der zehnten Interaktionseinheit.

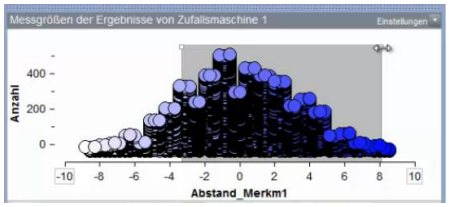
431 S: Wie groß ist der P-Wert? (seufzt) Weißt du noch wie man den berechnet?

432 R: Ich meine man könnte das über den Einteiler machen, dass man den auf 7,7 stellt (.). Dann kann man sich ja die Prozente einblenden lassen, dann sieht man ja im Prinzip wie viele Werte gleich oder eben extremer sind.

433 *Die Studentin (R) klickt oben in der Leiste den Button "Einteiler" an.*



434 *Dieser erscheint und die Studentin zieht das rechte Ende weiter nach rechts.*



435 R: Also den bis ans Ende, den anderen auf 7,7.

436 S: /Nee, wenn du das/.

437 R: Achso könnten das/ kann ich das irgendwo eingeben, ne?

438 I: Doppelklick auf das Ende.

439 S: Genau.

440 R: Also ich hätte jetzt gesagt den hier. (zeigt auf das linke Ende des Einteilers)

441 S: Bis minus zehn.

442 R: Oder den, den bis sieben. Also ich hätte jetzt gesagt, den bis 7,7 und den dann einfach bis ans Ende ziehen und dann hab ich ja eigentlich alle Werte mit drin, die bei genauem Hinsehen (unv.).

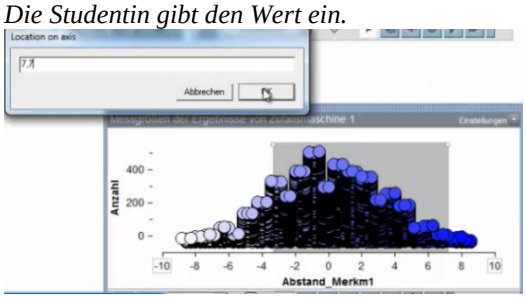
443 S: Achso ok alles klar.

444 R: Oder die.

445 S: Genau.

446 R: Eben noch größer sind. Eben noch extremer die Differenz. Ich kann es ja mal einmal machen, wir können es ja sonst immer nochmal/.

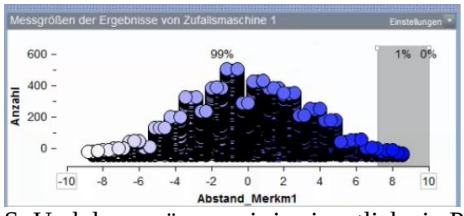
447 *Die Studentin gibt den Wert ein.*



448 S: /Ja aber ich hab ja jetzt verstanden was du, ich hätte es halt genau anders gemacht und da äh/ aber dann hätte ja der Einteiler nicht/.

449 R: /Also ich hätt es jetzt so gemacht.

450 *Anschließend wird der Einteiler verändert und der Button „%“ gedrückt. Dadurch erscheinen prozentuelle Werte.*



451 S: Und dann müssen wir ja eigentlich ein Prozent
452 R: Ja, das heißt im Prinzip haben wir den P-Wert/.
453 S: /Den P-Wert von eins.
454 R: Ein Prozent.

Selina wendet sich nun wieder auf Softwareebene dem weiteren Prozess im Randomisierungstest zu und liest zunächst die Frage des Schemas in (431) vor und wendet sich dann an Rebecca „Wie groß ist der P-Wert? Weißt du noch wie man den berechnet?“. Sie meint wohl eher ermitteln oder bestimmen statt berechnen, denn eine formale Berechnung des P-Werts war nicht Lerngegenstand des Bausteins Inferenzstatistik. Ihre Frage an Rebecca in Verbindung mit dem Seufzen lässt als Interpretationsmöglichkeit zu, dass sie selbst nicht (mehr?) weiß, wie der P-Wert in TinkerPlots ermittelt werden kann. Rebecca antwortet daraufhin und leitet vorsichtig mit „Ich meine“ ein und erläutert im Folgenden das Vorgehen technisch korrekt (432). Daraufhin wird der Einteiler eingeblendet. Es wird kurz ausgehandelt, welches Ende des Einteilers wohin gezogen wird, schließlich überdeckt nach Rebeccas Anregung der graue Bereich des Einteilers alle Werte ab 7,7. Selina wollte alle Werte kleiner als 7,7 mit dem Einteiler abdecken, was sich im Nachhinein in (448) klärt. Auf der Ebene der Software ist Rebeccas Vorgehen richtig, denn der graue Bereich des Einteilers beinhaltet alle Werte innerhalb der Grenzen dieses Bereichs, also alle Werte ab 7,7. An dieser Stelle zeigen die Beiden eine gute Beherrschung der Software.

Gleich zu Beginn macht Rebecca deutlich, dass sie die Bedeutung des P-Werts auf Statistikebene parat hat. Nach ihrer technischen Erläuterung zum Vorgehen, wie der P-Wert in TinkerPlots ermittelt werden kann, fügt sie an „dann sieht man ja im Prinzip wie viele Werte gleich oder eben extremer sind“ (432). Zur vollständigen Definition würde hier noch fehlen: Unter Voraussetzung, dass die Nullhypothese wahr ist. Aber der bisherige Gesprächsverlauf lässt die Deutung zu, dass dies Rebecca bewusst ist, und hier eine sprachliche Verkürzung vorliegt und dies nicht auf Nichtwissen hindeutet.

Nach der Verschiebung des Einteilers werden Prozente eingeblendet und Selina formuliert zunächst „Und dann müssen wir ja eigentlich ein Prozent“ (451) als wäre sie nicht sicher, ob sie den richtigen Prozentwert als P-Wert interpretiert. Das Wort „eigentlich“ deutet hier auf eine geringe Unsicherheit hin und fordert gleichzeitig Rebecca zu einer Reaktion auf. Rebecca übernimmt den Turn und bestätigt „Ja, das heißt im Prinzip haben wir den P-Wert.“ (452). Hierauf quantifiziert nun Selina noch einmal „Den P-Wert

von eins.“ (453), was von Rebecca präzisiert wird durch die Angabe von „Ein Prozent“ (454).

Wieder wird sehr kooperativ und einander ergänzend statistische Bedeutung ausgehandelt, in diesem Fall der P-Wert ermittelt. Hierbei treten keinerlei Schwierigkeiten auf, abgesehen davon, dass nicht explizit kommuniziert wird, dass es sich um eine Schätzung des P-Werts handelt.

Kommunikation in Verbindung mit der Kontextebene findet nicht statt.

Interaktionseinheit 11: Schlussfolgerungen ziehen

Schlussfolgerungen werden in der elften Interaktionseinheit thematisiert. Diese Einheit wird in fünf Episoden gegliedert. In der ersten Episode interpretieren Rebecca und Selina den P-Wert auf der statistischen Ebene. In der zweiten Episode formulieren sie in einem Exkurs ihre Schwierigkeiten, die sie beim Formulieren während der Lehrveranstaltung hatten. Episode drei beinhaltet ein Gespräch zwischen den drei Gesprächsteilnehmern ebenfalls als Exkurs über kritisches Denken, das durch den Baustein Inferenzstatistik gefördert wurde. In Episode vier wird das Schema ausgefüllt. In der fünften Episode findet erneut ein Exkurs statt.

Episode 11-1: Statistische Interpretation des P-Werts (Z. 455-468)

455	S: Ja und damit ist das ein signifikantes Ergebnis und die Nullhypothese kann abgelehnt werden.
456	R: Genau.
457	S: Ja.
458	R: Halt immer mit dem im Hinterkopf, dass man nie einen absoluten Beweis hat.
459	S: Genau.
460	R: Was wir eben ja gesagt haben mit diesen Unsicherheiten. Sie könnte theoretisch/ ich glaube wir hatten mal gesagt, es bleibt eine Restunsicherheit von einem Prozent, dass die eventuell doch noch/.
461	S: /Also man kann damit nicht sagen, dass sie falsch ist, aber die Wahrscheinlichkeit, dass sie richtig ist, ist ja dann um den p-Wert. Oder das war doch so. (.). Man kann also nicht zu hundert Prozent sagen, dass die Nullhypothese nicht gilt. Nur/.
462	R: /Ja, nur ich glaub man konnte keine genaue Wahrscheinlichkeit angeben, also man konnte jetzt nicht sagen, die Wahrscheinlichkeit für die Forschungshypothese ist so und so viel Prozent, die für die Nullhypothese ist so und so viel. Ich glaube man konnte immer nur mit so Restunsicherheiten/.
463	S: /Ja genau.
464	R: Von so und so viel Prozent/.
465	S: /Ja.
466	R: Argumentieren. So hatte ich das zumindest verstanden.
467	S: Ja, ja, so hatte ich es auch verstanden.
468	R: Früher und dann.

Direkt an die letzte Episode anschließend interpretiert Selina auf Statistikebene hier „Ja und damit ist das ein signifikantes Ergebnis und die Nullhypothese kann abgelehnt werden“ (455). Mit dem im Baustein Inferenzstatistik Gelernten ist dies richtig und sprachlich korrekt wiedergegeben. Beide bestätigen einander hiernach und Rebecca ergänzt in

(458), dass das keinen „absoluten Beweis“ darstelle. Auch dies ist positiv zu bewerten, da sie hiermit ausdrückt, dass ihr bewusst ist, dass die Forschungshypothese nicht bewiesen werden kann. Selina bestätigt das. Eine genaue Interpretation, was der P-Wert von einem Prozent bedeute, versucht Rebecca in (460) „… es bleibt eine Restunsicherheit von einem Prozent, dass die eventuell doch noch“, allerdings führt sie diesen Satz nicht zu Ende. Eine richtige Interpretation würde lauten: In einem Prozent der Fälle tritt ein Ergebnis wie das beobachtete oder ein noch extremeres auf, wenn die Nullhypothese gilt. Den Satz von Rebecca in diesem Sinne korrekt zu Ende zu führen, scheint jedoch schwierig. Es läuft eher darauf hinaus, dass sie auf eine Hypothesenwahrscheinlichkeit anspielt, die jedoch durch den P-Wert nicht gegeben wäre. Auf eine Wahrscheinlichkeit für die Nullhypothese versucht Selina sich in (461) zu beziehen: „Also man kann damit nicht sagen, dass sie falsch ist, aber die Wahrscheinlichkeit, dass sie richtig ist, ist ja dann um den P-Wert. Oder das war doch so. (.) Man kann also nicht zu hundert Prozent sagen, dass die Nullhypothese nicht gilt.“ In diesem Turn von Selina stecken richtige und falsche Aussagen. Sie wiederholt noch einmal in eigenen Worten, was Rebecca in (458) gesagt hat, dass man die Nullhypothese nicht sicher ablehnen könne. Jedoch versucht sie, mit Hilfe des P-Werts eine Wahrscheinlichkeitsangabe dafür zu machen, dass die Nullhypothese richtig ist. Dies entspricht einer typischen Fehlvorstellung zum P-Wert. Diese falsche Aussage wird direkt im nächsten Turn von Rebecca korrigiert, die sagt: „Ja, nur ich glaub man konnte keine genaue Wahrscheinlichkeit angeben, also man konnte jetzt nicht sagen, die Wahrscheinlichkeit für die Forschungshypothese ist so und so viel Prozent, die für die Nullhypothese ist so und so viel. Ich glaube man konnte immer nur mit so Restunsicherheiten“ (462). Durch das zweifache „Ich glaube“ in diesem Turn lässt sich einerseits interpretieren, dass sie selbst nicht ganz sicher ist über ihre eigene Aussage, andererseits lässt sich interpretieren, dass ihr bewusst ist, dass sie nun gegen Selinas vorige Aussage argumentiert und dies sprachlich höflich einleitet. Dies wird jedoch von Selina im nächsten Turn direkt bestätigt, so dass keine weiteren Argumentationen nötig sind. Wieder einmal gilt das von Rebecca hervorgebrachte Wissen als gemeinsam geteilt durch die Zustimmung in den nächsten Turns (463-468).

Letztlich beschäftigen sich Rebecca und Selina in dieser Episode mit der Frage, wie sich der P-Wert im Hinblick auf das Nichtverwerfen der Nullhypothese interpretieren lässt. Hier geraten sie in leichte Schwierigkeiten und können dies nicht statistisch korrekt formulieren. Jedoch revidieren sie Selinas falsche Aussage über eine Wahrscheinlichkeit für die Nullhypothese und geben auch an, dass bei einem P-Wert von einem Prozent die Nullhypothese nicht sicher abgelehnt werden kann. Sie neigen hier zu der natürlichen Fehlinterpretation des P-Werts, erinnern sich aber daran, dass man dies nicht interpretieren darf, ohne sich an Gründe hierfür zu erinnern. Im Sinne des Lehrkonzepts entspricht das dem, was man erwarten kann.

Es findet in dieser Episode kein Gespräch auf Softwareebene statt, dies ist auch nicht nötig. Eine Interpretation des P-Werts im Kontext findet nicht statt. Die beiden Teilnehmerinnen bleiben auf der statistischen Ebene der Interpretation.

Episode 11-2: Exkurs zu Schwierigkeiten beim Formulieren (Z. 469-479)

In den folgenden elf Turns sprechen die beiden über die Schwierigkeit, den gerade behandelten Sachverhalt korrekt zu formulieren. Ihnen ist bewusst, dass „… die Formulierungen extrem wichtig sind…“ (473, Selina), worauf Rebecca antwortet: „Das war aber wirklich mühselig sich das anzueignen“ (474). Für den Lösungsprozess spielt diese und auch die nächste Episode nur eine untergeordnete Rolle, deshalb werden diese beiden Episoden nur kurz zusammengefasst.

Episode 11-3: Exkurs zum kritischen Denken (Z. 480-501)

In dieser Episode wird thematisiert, wie der Baustein Inferenzstatistik die Vorstellung über publizierte Studienergebnisse der Beiden beeinflusst hat, was von der Interviewerin mit „kritischem Denken“ bezeichnet wird. Selina beschreibt dies in (489): „Das ist die Vorlesung. Ja also ich fand halt auch man hat dadurch einen generell anderen Blick auf Statistiken bekommen. Also ich hab mir bestimmt niemals zuhause vorher Gedanken gemacht, ob das eventuell jetzt doch nicht wahr sein könnte. Also ne, dass da immer so eine Restunsicherheit. Sondern ganz im Gegenteil, wenn jetzt eine Studie von (.) einer zuverlässigen Quelle/ da muss man ja auch nochmal hinterher gucken. Das weiß man ja durch wissenschaftliches/ durch Hausarbeiten äh schon ausreichend. Aber dass selbst da noch so eine Restunsicherheit ist. Also man hat es einfach nicht hinterfragt, sondern meistens als so gegeben hingenommen. Und jetzt auf einmal denkt man so, naja (.) vielleicht.“ Beide berufen sich hier immer wieder auf die „Restunsicherheit“, die ein statistischer Test mit sich bringe und von der in den Medien selten berichtet würde. Dies ist ein sehr erfreuliches Ergebnis im Hinblick auf das Lehrkonzept.

Episode 11-4: Ausfüllen des Schemas bezüglich Schlussfolgerungen (Z. 502-523)

502 R: Man würd das dann hier auch einfach so aufschreiben. Forschungshypothese kann angenommen, Nullhypothese kann abgelehnt werden. Allerdings bleibt da eine gewisse Restunsicherheit, wenn wir die Schlussfolgerung/.
503 S: /Ja und äh.
504 R: /Ziehen.
505 *Es wird notiert: „-Forschungshypothese kann angenommen & Nullhypothese abgelehnt werden, allerdings unter Restunsicherheit“*
506 S: Also, Population ist ja auch noch. Es ist ja natürlich auch die Frage, inwiefern das repräsentativ ist, wenn vierzehn Personen/.
507 R: /Genau.
508 S: Also, ob man das auf die Gesamtheit der Population übertragen kann.
509 R: Ja das muss man/.
510 S: /Ja.
511 R: /Auch noch miteinbeziehen, dass man da (.)
512 S: Also das.

513 R: Also dass da auch Vorsicht geboten ist, dass es jetzt eine Stichprobe an Personen ist.
514 S: Ja vor allem, von halt nur vierzehn Teilnehmern/.
515 R: /Ja.
516 S: Und, ob man daraus dann ne Aussage für (.) alle anderen ziehen kann.
517 R: Ja.
518 S: Ist ja auch fraglich.
519 R: Ja.
520 R: Soll ich das mal?
521 S: Ja, schreib das mal kurz und prägnant.
522 I& R& S: (lachen)
523 *Es wird notiert: „-Lässt sich nicht verallgemeinern (nur 14 Teilnehmer)“*

In der vierten Episode dieser Interaktionseinheit wenden sich die beiden nun dem Ausfüllen des Schemas und somit dem Verschriftlichen ihrer Schlussfolgerungen zu. Rebecca fasst ihre Schlussfolgerungen in (502) auf statistischer Ebene zusammen: „… Forschungshypothese kann angenommen, Nullhypothese kann abgelehnt werden. Allerdings bleibt da eine gewisse Restunsicherheit, wenn wir die Schlussfolgerung [ziehen].“ Dies wird im Folgenden so von den beiden notiert. Rebecca betont hier noch einmal die „Restunsicherheit“, wobei unklar bleibt, ob sie dieses Konzept wirklich verstanden hat. Anschließend bemerkt Selina „Population ist ja auch noch. Es ist ja natürlich auch die Frage, inwiefern das repräsentativ ist, wenn vierzehn Personen.“ (506). Dies ist vermutlich angeregt durch das Schema, das sie im Moment ausfüllen und dass den Punkt *Population?* enthält, da zu keinem Zeitpunkt vorher jemals die Reichweite der Schlussfolgerungen in Bezug auf die Population von den Beiden thematisiert wurde. Selinas Aussage hierzu, über mehrere Turns hinweg ist „Also ob man das auf die Gesamtheit der Population übertragen kann“ (508) „Ja vor allem von halt nur vierzehn Teilnehmern“ (514) „Und ob man daraus dann ne Aussage für (.) alle anderen ziehen kann“ (516) „Ist ja auch fraglich“ (518). Rebecca wirft kurze bestätigende Turns zwischendrin ein und fügt hinzu „Also dass da auch Vorsicht geboten ist, dass es jetzt eine Stichprobe an Personen ist“ (513). Wie im bisherigen Gesprächsverlauf ergänzen die beiden Teilnehmerinnen im Dialog ihre Sätze und scheinen ihre Gedankengänge wechselseitig zu vervollständigen. Am Ende der Episode wird notiert „Lässt sich nicht verallgemeinern (nur 14 Teilnehmer)“ (523). Wünschenswert wäre gewesen, dass noch eine Aussage getroffen wird über die Wirksamkeit für diese 14 Teilnehmer, dies geschieht jedoch nicht.

Episode 11-5: Exkurs Interesse am Thema (Z. 524-542)

In der letzten Episode dieser Einheit wendet sich das Gespräch einem anderen Thema zu. Selina formuliert, dass sie bedauert, dass Inferenzstatistik nur in zwei Vorlesungen des Kurses thematisiert wurde (524). Im Weiteren beschreiben die beiden Teilnehmerinnen ihre Vorerfahrungen im Bereich der Stochastik (vgl. Kapitel 8.1) und betonen die Wichtigkeit, dieses „Kritische“ zu verinnerlichen (540). Für den Lösungsprozess ist diese Episode von untergeordneter Rolle und wird deshalb nicht analysiert.

Hiernach ist die selbständige Durchführung des Randomisierungstests durch die beiden Teilnehmerinnen abgeschlossen und es wird im Folgenden den Interviewfragen des Leitfadens (Kapitel 9.4) nachgegangen. Da viele Inhalte bereits während des ersten Teils besprochen wurden, wurden einige Fragen nur kurz gestreift und werden hier nur insoweit wiedergegeben, als dass sie neue Erkenntnisse liefern.

Interaktionseinheit 12: Zusammenhang zwischen den formulierten Hypothesen und den möglichen Erklärungen (Z. 544-548)

544	I: Mhm (bejahend) (.) Ok. Das heißt ihr seid jetzt soweit eigentlich durch. Jetzt würde ich gerne noch so ein paar Nachfragen stellen. (.) Inhaltlicher Art oder auch statistischer Art. Also, ihr habt/ ich würd gerne mal eure Forschungs- und eure Nullhypothese nochmal sehen oder vielleicht kann sie nochmal jemand vorlesen von euch.
545	R: Genau. Als Forschungshypothese hatten wir eben, dass Fischöl hilfreich oder wirksam für Blutdrucksenkung ist. Und als Nullhypothese dass eben die zufällige Gruppenzuordnung zu dieser Fischöldiät und zu dieser normale Öl Diät Gruppe die Ursache für diese beobachteten Unterschiede ist. Also dass es einfach an der Gruppenzuordnung liegt und gar nicht daran, ob man jetzt mit Fischöl oder mit dem normalen Öl eben (.) ja behandelt wurde.
546	I: Mhm (bejahend) Genau dann noch einmal kurz drauf eingegangen auf die Nullhypothese. Das heißt, wir hatten ja am Anfang über diese Erklärung gesprochen, was es für mögliche Erklärungen gibt für die Ergebnisse, die beobachtet wurden. Welche Erklärung steckt quasi hinter der Nullhypothese?
547	R: Hmm also hinter der Nullhypothese ja im Grunde, eigentlich die Erklärung. (.) Blickt man auf die Unterschiede zwischen den beiden Gruppen, die ja auf den ersten Blick vielleicht darauf schließen lassen, dass Fischöl wirksam ist. Dass das im Grunde ja einfach daran liegt, dass zufällig ähm (.) die Gruppen so eingeteilt wurden, dass eben Leute in der einen und der anderen Gruppe waren, die eben einfach durch ihre personengebundenen Eigenschaften zu diesen Ergebnissen (.) geführt haben, aber dass es eben gar nicht an dem Öl wirklich liegt.
548	S: Ja ich/ und dass es halt null Zusammenhang zwischen den entsprechenden Ölen und der Blutdrucksenkung gibt. Also davon geht die Nullhypothese ja aus.

In der zwölften Interaktionseinheit wird der Leitfrage A1 des Leitfadens nachgegangen. Wie in Kapitel 9.4 beschrieben soll hier geklärt werden, ob die Teilnehmerinnen einen Bezug zwischen den von ihnen formulierten Hypothesen und den vorher genannten Erklärungen für die beobachteten Unterschiede herstellen können. Rebecca formuliert, dass hinter der Nullhypothese die Erklärung stehe, dass die Gruppen zufällig eingeteilt wurden und dass es „eben gar nicht an dem Öl wirklich liegt“ (547). Selina formuliert noch deutlicher im darauffolgenden Turn „… dass es halt null Zusammenhang zwischen den entsprechenden Ölen und der Blutdrucksenkung gibt. Also davon geht die Nullhypothese ja aus“ (548).

Wie sich bereits im Lösungsprozess dieses Paars angedeutet hat, können sie die möglichen Erklärungen in Verbindung bringen mit der Nullhypothese. Da sich dies bereits während des vorigen Interviewteils in der selbständigen Bearbeitung der Beiden angedeutet hat, wurde an dieser Stelle von der Interviewerin darauf verzichtet, zum wiederholten Mal hierzu Nachfragen zu stellen. Die Forschungshypothese wird nicht in Ver-

bindung mit einer möglichen Erklärung thematisiert, da diese bereits in der Formulierung eine Erklärung beinhaltete.

Interaktionseinheit 13: Beschreiben der Zufallsmaschine (Z. 550-570)

Da während des Lösungsprozesses (Interaktionseinheit 8) viel Hilfe von Seiten der Interviewerin auf Softwareebene nötig war und bereits dort viele Erläuterungen von den Studierenden eingefordert wurden, wurde dieser Teil des Interviews am Ende sehr knapp gehalten. Es wurde vor allem aufgrund der fortgeschrittenen Interviewdauer nur eine spezifische Nachfrage von der Interviewerin zur nicht erfolgten Umbenennung der Merkmale gestellt. Der gesamte erste Teil der Einheit beschäftigt sich mit dem technischen Auffinden der Zufallsmaschine unter all den minimierten Objekten in der Arbeitsfläche von TinkerPlots (550-564). Die folgenden Transkriptzeilen beziehen sich auf die inhaltlichen Aspekte dieser Einheit.

565	I: Dann von mir noch eine Frage. Ihr habt die zwei Merkmale, Merkmal eins und zwei/ die habt ihr nicht umbenannt. Ähm (..) warum nicht?
566	S: Weils jetzt nicht relevant war, oder?
567	R: Ja. (.) Ja im Prinzip.
568	S: Also war ja keine Relevanz. Wir wollten ja einfach zwei neue zugeordnete Gruppen haben und welches Merkmal da dargestellt wird, ist ja eigentlich irrelevant, wir wollen ja nur gucken, wie sich das neu verteilt.
569	R: Ja.
570	S: Ja. Genau wir hatten ja unten die Gruppen A und B benannt, die wir haben wollen und das reicht uns ja im Grunde nachher für die Zuordnung, die wir brauchen.

Die von TinkerPlots automatisch vergebenen Merkmale Merkm1 und Merkm2 wurden nicht umbenannt, „Weils jetzt nicht relevant war, oder?“ (566, Selina), was von Rebecca im nachfolgenden Turn bestätigt wird. Als Erklärung fügt Selina abschließend an, dass sie nur daran interessiert waren, die zwei neuen Gruppen A und B zu erhalten und dass dies ausreichend war (568). Es lässt sich interpretieren, dass durch die Struktur der Daten eine Umbenennung der Merkmale den Teilnehmerinnen nicht nötig erschien. Schwierigkeiten aus der Nichtumbenennung der Merkmale, wie sie von Noll und Kirin (2017) bei einer ähnlichen Aufgabe berichtet wurden, haben sich für Selina und Rebecca nicht ergeben.

Interaktionseinheit 14: Verständnis über einzelne Punkte der Referenzverteilung (Z. 571-588)

In dieser Interaktionseinheit wird der Frage A3 des Leitfadens zur Referenzverteilung nachgegangen. In diesem Abschnitt erklärt zunächst Rebecca, was ein bestimmter Punkt der Referenzverteilung bedeutet: „Ja das ist ja im Grunde von einer dieser ganz vielen zufälligen neuen Zuordnungen zu Gruppe A und B. Da hat man eben die beiden arithmetischen Mittel genommen und dann eben die Differenz davon gebildet.“ (574). Diese Erklärung ist zwar sprachlich etwas holprig, aber das mag der Interviewsituation geschuldet sein. Inhaltlich ist die Erklärung korrekt. In (578) fragt die Interviewerin wei-

terführend, was ein Punkt mit dem Wert 9,6 in Bezug auf die Nullhypothese bedeute. Als Antwort formuliert Rebecca „… wenn es jetzt gerade so einen ganz häufig gäbe, dann (.) wäre es ja wahrscheinlich so, dass wir hier nicht bei diesem einen Prozent bei dem P-Wert liegen würden, sondern einen viel höheren hätten. Und dann letztendlich eben eventuell die Nullhypothese annehmen könnten…“ (579). Sie interpretiert korrekt, dass der P-Wert höher wäre, wenn es einen solchen Wert häufiger gäbe. Bei der weiteren Schlussfolgerung zeigt sich, dass die Studentinnen die Veränderung des P-Werts korrekt vermerken, ebenso die daraus resultierende Konsequenz für die Nullhypothese, die nun nicht abgelehnt werden kann, auch wenn dies sprachlich nicht richtig formuliert wird. Weiterhin zeigt sich die Schwierigkeit bezüglich der Nullhypothese, dass sie die Nullhypothese bei einem großen P-Wert annehmen würde, was ein statistisch nicht korrekter Schluss ist. Auch Selina formuliert dies ähnlich in ihrem nächsten Turn und zeigt damit die gleiche Schwierigkeit.

Interaktionseinheit 15: Schlussfolgerungen ziehen (Z. 589-626)

Den Fragen A5 bis A9 des Leitfadens wird in der fünfzehnten und letzten Interaktionseinheit des Interviews nachgegangen. Diese Einheit wird in drei Episoden gegliedert. In der ersten Episode wird den Fragen A5 und A6 in leicht abgewandelter Form nachgegangen. Die zweite Episode thematisiert die Frage A7 des Leitfadens und am Ende wird in der letzten Episode erst der Frage A9 und dann der Frage A8 nachgegangen. Die Frage A4, was sich aus einem P-Wert von einem Prozent schließen lässt, wurde nicht explizit behandelt, weil dies bereits ausführlich thematisiert wurde.

Episode 15-1: Schlussfolgerungen im Kontext (Z. 589-598)

589 I: Gut. (.) Das haben wir auch. Signifikantes Ergebnis habt ihr aufgeschrieben. Jetzt nochmal auf den P-Wert. Ich glaube ihr hattet vorhin so ein bisschen (.) ähm darüber nachgedacht/ nochmal die genaue Formulierung was der P-Wert bedeutet (.) Ich gebe das auch gerne vor. Du hattest glaube ich vorhin sehr gut formuliert, der P-Wert ist quasi die Wahrscheinlichkeit so ein Ergebnis, wie das was wir beobachtet haben, sprich die 7,7 oder was noch extremeres zu bekommen. (..) Und wir haben jetzt hier einen P-Wert von einem Prozent, halt ungefähr. Was sagt ihr denn jetzt? Sagt ihr jetzt, Fischöl hat eine blutdrucksenkende Wirkung? Also die Forschungsfrage, die ihr am Anfang formuliert habt, wie beantwortet ihr die?

590 R: Also ich würde sie schon mit JA beantworten, aber eben auf jeden Fall die Formulierung vermeiden: Es ist bewiesen, dass Fischöl eine blutdrucksenkende Wirkung hat, weil man da eben an dieser Stelle vorsichtig sein muss, dass man eben nicht von einem Beweis spricht, weil wir eben/ was wir eben auch schon gesagt haben, immer noch diese Restunsicherheiten haben. Und einfach da nicht von hundertprozentiger Sicherheit ausgehen können.

591 I: Mhm (bejahend).

592 S: Ja genau, also dem würde ich zusammen. Ich/ Man kann jetzt annehmen, dass Fischöl eine blutdrucksenkende Wirkung hat, aber zu hundert Prozent äh (.) sagen kann man es halt nicht.

593 I: Mhm (bejahend).

594 S: Man kann auch nicht sagen zu neunundneunzig prozentiger Wahrscheinlichkeit/ also das war ja auch so ein Thema: Ist Fischöl wirksam? Also das sagt der P-Wert halt eben nicht aus. Sondern man kann jetzt halt nur sagen, tendenziell ist es wahrscheinlicher, dass Fischöl eine Wirkung hat als dass es nicht so ist. Aber eine hundertprozentige oder überhaupt eine prozentuelle Aussage kann ich jetzt nicht treffen.

595 I: Gut. Aber ihr würdet beide sagen, Fischöl hat eine Wirkung? Lässt sich das hieraus schließen?
596 S: Zumindest für diese Gruppe.
597 I: Ok wichtige Einschränkung. Genau für diese Gruppe.
598 S: Genau, also ich kann das ja jetzt nicht, wie eben schon gesagt, für die Gesamtpopulation beschreiben, weil die Stichprobe mit vierzehn Personen (.) da müsste man sich fragen, wie aussagekräftig das ist.

Zum ersten Mal formuliert Rebecca eine Schlussfolgerung im Kontext in (590): „Also ich würde sie [die Forschungsfrage] schon mit JA beantworten, aber eben auf jeden Fall die Formulierung vermeiden: Es ist bewiesen, dass Fischöl eine blutdrucksenkende Wirkung hat, weil man da eben an dieser Stelle vorsichtig sein muss…“. Die von Rebecca und Selina formulierte Forschungsfrage war: „Ist Fischöl wirksam, um eine Blutdrucksenkung herbeizuführen?“. Wie nun bereits mehrfach geschehen, weist Rebecca noch einmal daraufhin, dass es sich um keinen Beweis handele. Wenn man es genau nimmt, so hat Rebecca hier nur ausgedrückt, was sie nicht formulieren würde. Eine tatsächliche Formulierung liefert Selina in ihrem nächsten Turn mit „… Man kann jetzt annehmen, dass Fischöl eine blutdrucksenkende Wirkung hat, aber zu hundert Prozent äh (.) sagen kann man es halt nicht.“ (592). Diese Aussage ist richtig, wenn man von der fehlenden Einschränkung auf die Personen, die am Experiment teilgenommen haben, absieht. Direkt im Anschluss führt sie eine mögliche Schlussfolgerung noch weiter aus, diesmal thematisiert sie von sich aus die Angabe einer Wahrscheinlichkeit für die Hypothese „Man kann auch nicht sagen zu neunundneunzig prozentiger Wahrscheinlichkeit/ also das war ja auch so ein Thema: Ist Fischöl wirksam? Also das sagt der P-Wert halt eben nicht aus. Sondern man kann jetzt halt nur sagen, tendenziell ist es wahrscheinlicher, dass Fischöl eine Wirkung hat als dass es nicht so ist. Aber eine hundertprozentige oder überhaupt eine prozentuelle Aussage kann ich jetzt nicht treffen“ (594). Zunächst einmal stellte Selina hier richtig fest, dass mit dem P-Wert keine Wahrscheinlichkeitsangabe für die Nullhypothese gemacht werden kann, im Gegensatz zu früheren Formulierungen. Dann drückt sie mit Hilfe des Wortes „tendenziell“ dies vorsichtiger aus. Das Wort „tendenziell“ wurde nicht während des Bausteins Inferenzstatistik eingeführt und auch nicht wissentlich in diesem Zusammenhang verwendet, wie Selina es hier tut. Es stammt aus dem Baustein zur Datenanalyse, als Gruppenvergleiche thematisiert wurden. Wenn beispielsweise der Frage nachgegangen wurde, ob Jungen größer sind als Mädchen, so wurde gelehrt, dass eine Allaussage wie „Jungen sind größer als Mädchen“ selten richtig ist (eben nur, wenn tatsächlich jeder Junge größer ist als jedes Mädchen eines Datensatzes), dass aber durch das Wort „tendenziell“ die Aussage positiv korrigiert werden kann.[48] Eventuell in diesem Sinne des vorsichtigeren Ausdrückens verwendet Selina hier „tendenziell“. Hierauf fragt die Interviewerin erneut, ob nun Fischöl eine Wirkung habe und ob sich das nun schließen lasse (595). Daraufhin schränkt Selina

[48] Es ließ sich in vielen Hausaufgaben und Seminarsitzungen beobachten, dass die Studierenden allgemein eine Vorliebe für das Wort „tendenziell“ entwickelten und es sehr häufig einsetzten.

antwortend ein: „Zumindest für diese Gruppe“ (596), was von der Interviewerin im nächsten Turn honoriert wird. Selina ergreift erneut das Wort und führt aus, dass es nicht für die Gesamtpopulation beschrieben werden könne (598).

Episode 15-2: Sicherheit bezüglich der Einschätzungen (Z. 599-616)

599	I: Dann aber meine Frage, wie sicher seid ihr euch bei der Einschätzung? Das ist jetzt nichts was ihr gelernt habt im Kurs. Wie sicher seid ihr euch denn, wenn ihr jetzt sagt: Für diese vierzehn Personen gehen wir davon aus, dass Fischöl wirksam ist.
600	R: Das ist schwierig, ne?
601	S: Das kann man halt immer nur wieder relativierend beantworten. Also es ist ja einfach/ wie gesagt dadurch, dass diese ganzen anderen Faktoren, die ich gar nicht ausschließen kann, das sie sich unbewusst doch anders verhalten haben, etc. pp, dass ich das jetzt nicht überprüfen kann. (.) Kann ich annehmen, dass Fischöl vielleicht einen Beitrag dazu geleistet hat, aber ob das jetzt die ausschlaggebende Komponente war (...) vermag ich jetzt nicht zu sagen. Also ich würde dann vielleicht einfach noch weitergehende Studien brauchen, um äh die Aussagekraft zu überprüfen. Also anhand einer Studie würde ich doch keine sichere Aussage treffen können. Das wäre mir zu unsicher, ob da nicht vielleicht doch irgendwas anderes, (unv.) unbewusstes Verhalten oder was wir da eben alles hatten, ja zutrifft.
602	R: Ja und es sind ja eben auch nur vierzehn.
603	I: Ja aber wenn wir bei den vierzehn bleiben.
604	S: Also da würde ich einfach das auch nochmal wiederholen und langfristiger/ vor allem wie lange war die äh / wie lange war äh?
605	I: Vier Wochen.
606	S: Vier Wochen. Also auch der Zeitraum scheint mir zu gering/ also das müsste längerfristig/ meines Erachtens äh überprüft werden, um zu gucken, ob man dann wirklich noch eindeutigere Ergebnisse erzielt.
607	R: Ich bin mir jetzt auch nicht sicher, aber da steht ja auch, dass am Anfang und am Ende der Studie der Blutdruck gemessen wurde. Vielleicht kann man da auch zwischendurch auch immer noch mal Messungen anstellen. Einfach dass man mehrere Werte hat. Vielleicht hat sich jemand an diesem Tag von diesen vierzehn Leuten besonders stark aufgeregt oder so als das dann gemessen wurde. Es kann ja immer einfach aufgrund irgendwelcher Faktoren (.) doch was sein. Man kann ja gar nicht alles irgendwo ausschließen. Dass man vielleicht da nochmal für diese vierzehn Leute nochmal genauer guckt und mehr Messwerte vielleicht mit einbezieht.
608	I: Mhm (bejahend) (.) Das heißt ich höre bei euch beiden schon eine Unsicherheit raus (...) für diese vierzehn Leute das so richtig zu bejahen, dass Fischöl wirksam ist. Ist das richtig?
609	R: Ja also ein absolutes Ja würde ich jetzt nicht.
610	I: Könnt ihr das irgendwie quantifizieren? Seid ihr sehr sicher? Seid ihr sehr unsicher? Könnt ihr es in Prozenten ausdrücken? Würdet ihr sagen naja so mit achtzig Prozent bin ich sicher oder naja eigentlich nur mit zwanzig Prozent. Also (.) versteht ihr worauf ich hinaus möchte?
611	R& S: Mhm (bejahend).
612	S: Also (lacht) das ist ja (...) also (unv.) (..) es sind in Prozent jetzt (...) wie viel sagen wir (unv.).
613	I: Wenn ihr es nicht könnt, könnt ihr es nicht.
614	S: Also ich würde es nicht kategorisch ablehnen und würde sagen, nee auf keinen Fall, also ich glaube jetzt nicht, dass das Fischöl nicht gewirkt hat, weil ähm (...) man kann ja sehen, dass da äh (.) also (...) ich gehe davon aus, dass es einen Beitrag geleistet hat.
615	R: Wir haben ja auch diesen kleinen P-Wert von einem Prozent. Nur kategorisch, dass man sich jetzt absolut unsicher ist und sagen würde, ok eigentlich passt es gar nicht, ist bei mir auch nicht so. Aber ich finde es schwierig, das in Prozenten auszudrücken, wie man das so gewichtet. Man hat Punkte, die dafür sprechen. Man hat aber auf der anderen Seite eben auch Punkte, wo man sagt ja das könnte aber doch irgendwo sein, dass es nicht ganz passt. Also ich könnte es jetzt nicht in genauen Prozentzahlen ausdrücken.
616	S: Also, wenn ich jetzt Bluthochdruck hätte würde ich vielleicht nicht losrennen, um sofort Fischöl zu kaufen, um das mal äh so auszudrücken, ne. Ok es ist wahrscheinlich nicht schädlich,

schadet nicht, hilft vielleicht, dann würde ich es vielleicht doch machen. Aber ich würde jetzt nicht sagen JA das ist eine super Idee, ich lasse blutdrucksenkende Mittel weg und nehme alternativ Fischöl. Also so sicher wäre ich mir da nicht, dass es zum Erfolg führt. (lacht) Also, ne. Sonst könnte man ja so einiges kaufen, was statistisch belegt ist. Was nicht alles gegen irgendwas hilft (lachend).

Zur Frage A7 des Leitfadens (Wie sicher seid ihr euch bei eurer Einschätzung?) wird nur ein kurzer Eindruck wiedergegeben. Selina antwortet sehr gut damit, dass sie auf weitere Unsicherheitsquellen hinweist: „…Kann ich annehmen, dass Fischöl vielleicht einen Beitrag dazu geleistet hat, aber ob das jetzt die ausschlaggebende Komponente war (...) vermag ich jetzt nicht zu sagen. Also ich würde dann vielleicht einfach noch weitergehende Studien brauchen, um äh die Aussagekraft zu überprüfen. Also anhand einer Studie würde ich doch keine sichere Aussage treffen können. Das wäre mir zu unsicher, ob da nicht vielleicht doch irgendwas anderes, (unv.) unbewusstes Verhalten oder was wir da eben alles hatten, ja zutrifft“ (601). Neben weiteren Studien nennt sie in den nächsten Turns noch zeitliche längere Studien, die durchgeführt werden sollten, um ihr mehr Sicherheit bezüglich ihrer Einschätzung zu geben. Rebecca fügt hinzu, dass weitere Messungen innerhalb einer Studie durchgeführt werden sollen, um die Verlässlichkeit zu erhöhen und um weitere Faktoren (Stress, Aufregung) bei den Versuchspersonen auszuschließen. Rebecca ist unentschieden: „…Man hat Punkte, die dafür sprechen. Man hat aber auf der anderen Seite eben auch Punkte, wo man sagt ja das könnte aber doch irgendwo sein, dass es nicht ganz passt…“ (615) und Selina ist noch skeptischer und formuliert „Also wenn ich jetzt Bluthochdruck hätte, würde ich vielleicht nicht losrennen, um sofort Fischöl zu kaufen, um das mal so auszudrücken, ne. Ok, es ist wahrscheinlich nicht schädlich, schadet nicht, hilft vielleicht, dann würde ich es vielleicht doch machen…“ (616). Es lässt sich interpretieren, dass die beiden auf statistischer Ebene zwar die Nullhypothese ablehnen und die Forschungshypothese, dass Fischöl wirksam ist, um den Blutdruck zu senken, annehmen, wie sie es in der Interaktionseinheit 11 formuliert haben, aber diese Episode hier zeigt, dass eine Diskrepanz zwischen den statistischen Schlussfolgerungen (die Nullhypothese wird abgelehnt) und den Schlussfolgerungen im Kontext (ich würde nicht sofort losrennen, um Fischöl zu kaufen) vorliegt. Dies ist im Kontext dieser kleinen medizinischen Studie sehr positiv zu bewerten. Der Frageninhalt dieser Episode und die Rolle der Interviewerin hier scheinen beiden für den weiteren Fortgang des Interviews nicht zu benötigen.

Episode 15-3: Interpretation eines P-Werts von 30 Prozent (Z. 617-626)

617 I: Ok. Ähm dann als letzte Frage des Interviews. Das hattet ihr eben schon gesagt, wenn wir jetzt mal annehmen, dass der P-Wert nicht ein Prozent ist, sondern zum Beispiel dreißig Prozent wäre. (..) Was würdet ihr dann schließen?

618 S: Ja, dass die (..) Forschungshypothese ähm abgelehnt werden muss, also dass Fischöl nicht wirksam ist. Was nicht/ hatten wir ja eben schon darüber gesprochen, was nicht bedeutete, dass die Nullhypothese gilt, weil es ja dann noch immer nicht bewiesen ist, dass es an dieser zufälligen Zuordnung liegt. Ja, nur äh tendenziell/ dann haben wir eben kein signifikantes Ergebnis.

	Und dann muss man davon ausgehen, dass es doch nicht daran lag, dass Fischöl wirksam ist, sondern an irgendeinem anderen Faktor, den wir jetzt noch nicht genau benennen können.
619	I: Mhm (bejahend) Es tut mir Leid ich habe doch noch eine Frage. Und zwar: Was ist denn eine alternative Interpretation von einem Prozent? Von einem einprozentigen P-Wert. Ihr habt gesagt: Der lässt uns darauf schließen, dass wir die Nullhypothese ablehnen und dass wir deswegen die Forschungshypothese annehmen. Was könnte man denn alternativ schließen? Also was ist denn die andere Interpretation für diesen kleinen P-Wert?
620	S: Ja also (...) dass es natürlich auch eine geringe Wahrscheinlichkeit gibt, dass die Nullhypothese gilt. Also (...) oder sagt man, dass die Nullhypothese gilt? Also zumindest, dass es eine geringe Wahrscheinlichkeit gibt, dass andere Faktoren dafür ausschlaggebend sind, außer Fischöl. (..)/.
621	I: /Also formulieren würde man in der Tat (..) In einem Prozent der Fälle ist es aufgetreten, dass unter dieser zufälligen Zuordnung ein solcher Wert wie der, den wir beobachtet haben oder ein noch extremerer aufgetreten ist.
622	R: Ja das steht dann ja auch für diese Restunsicherheit, die man hat.
623	I: Genau.
624	R: (unv.) Ich glaube das hatten wir auch so in der Vorlesung. Wenn dann immer diese Restunsicherheit von einem Prozent bleibt, weil es eben ein Prozent der Fälle/ weil die eben dann doch aufgetreten sind.
625	I: Genau in einem Prozent der Fälle würde man die Nullhypothese sozusagen irrtümlich verwerfen. Dann lehnt man die irrtümlich ab, denn in der Tat kann das passieren und dann hätte man halt die Nullhypothese fälschlicherweise abgelehnt und auf die Wirkung von Fischöl geschlossen.
626	S: Ja.
627	I: Ok. Gut, dann sind wir inhaltlich erst mal durch. Ich danke euch sehr, ihr habt super konzentriert gearbeitet.

Auf die Frage A9 des Leitfadens, was aus einem hypothetischen P-Wert von 30 Prozent geschlossen werden könne (617, Interviewerin), antwortet Selina auf Kontext- und Statistikebene: „Ja, dass die (..) Forschungshypothese ähm abgelehnt werden muss, also dass Fischöl nicht wirksam ist. Was nicht/ hatten wir ja eben schon darüber gesprochen, was nicht bedeutet, dass die Nullhypothese gilt, weil es ja dann noch immer nicht bewiesen ist, dass es an dieser zufälligen Zuordnung liegt. Ja, nur äh tendenziell/ dann haben wir eben kein signifikantes Ergebnis. Und dann muss man davon ausgehen, dass es doch nicht daran lag, dass Fischöl wirksam ist, sondern an irgendeinem anderen Faktor, den wir jetzt noch nicht genau benennen können.“ (618). Sie interpretiert korrekt, dass in diesem Fall die Forschungshypothese abgelehnt werden kann und setzt damit gleich, dass dies nicht bedeute, dass die Nullhypothese gelte. Sie begibt sich auf die Suche nach weiteren Erklärungen für den beobachteten Unterschied und nennt „irgendeinen anderen Faktor“, der den beobachteten Unterschied verursacht haben könnte, der nicht benannt werden könne, sie bezieht sich also auf konfundierende Variablen. Die Formulierungen in (620) zeigen einen Rückfall in Hypothesenwahrscheinlichkeiten, was sich dadurch als fragiles Wissen herausstellt. Trotzdem bezieht sie sich korrekt auf die „Restunsicherheit“ (624), die bei einer Entscheidung bleibt, zeigt also, dass sie dies grundsätzlich verstanden hat.

Hiermit endet das Interview.

11.2.2 TinkerPlots-Benutzung zur Simulation von Rebecca und Selina

Aus den vorherigen Analysen wird in diesem Abschnitt entlang den in Kapitel 10.4.4 vorgestellten Kategorien analysiert, welche der notwendigen Schritte in TinkerPlots in den Interaktionseinheiten 8, 9 und 10 von Rebecca und Selina selbständig, mit leichter Intervention oder mit starker Intervention durch die Interviewerin durchgeführt werden.

Tabelle 30: Notwendige Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Rebecca und Selina

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen der Zufallsmaschine	-	-	Ja
Vorhandene Kugeln der Zufallsmaschine löschen	-	-	Ja
Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren	-	-	Ja
Ein zweites Bauteil der Zufallsmaschine hinzufügen	Ja	-	-
Zweites Bauteil füllen	-	-	Ja
Die Einstellung ohne Zurücklegen bei beiden Bauteilen auswählen	-	-	Ja
Durchgangsanzahl festlegen	-	Ja	-

Es lässt sich erkennen, dass Rebecca und Selina bezüglich der Arbeit mit der Zufallsmaschine bis auf zwei Ausnahmen immer starke Interventionen durch die Interviewerin benötigen. Sie benötigen zusätzlich zu den genauen Anweisungen, wie sie die Zufallsmaschine bedienen müssen auch Hinweise, welche Einstellungen überhaupt vorgenommen werden müssen. Lediglich das Hinzufügen eines zweiten Bauteils, um die neuen Gruppen zu modellieren, können sie selbständig durchführen. Und für die festzulegende Durchgangsanzahl benötigen sie nur den Hinweis, dass sie direkt in das Feld die gewünschte Anzahl eingeben können. Beide hatten jedoch mehrfach formuliert, dass sie Schwierigkeiten beim Bedienen der Software erwarten würden, insofern ist dies kein erstaunliches Ergebnis.

Tabelle 31: Optionale Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Rebecca und Selina

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einstellen der Geschwindigkeit	Ja	-	-
Umbenennen der Merkmale Merkm1 und Merkm2	Ja	-	-

Die Geschwindigkeit wird selbständig von Rebecca und Selina erhöht, sie benötigen hierzu keinen Hinweis. Über das Umbenennen der Merkmale haben sie diskutiert und dieses für nicht nötig befunden.

Tabelle 32: Notwendige Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Rebecca und Selina

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention

Erstellen einer Graphik zur Simulation	-	-	Ja
Das erste Merkmal auswählen und darstellen	-	-	Ja
Das zweite Merkmal auswählen und in der Graphik passend darstellen	Ja	-	-
Die Mittelwerte einblenden	Ja	-	-
Differenz der Mittelwerte mit Hilfe der Linealfunktion messen	-	-	Ja
Messgröße definieren	-	-	Ja
Anzahl der zu sammelnden Messgrößen eintragen	-	Ja	-
Messgrößen sammeln	Ja	-	-
Erstellen einer Graphik zu den gesammelten Messgrößen	Ja	-	-
Referenzverteilung darstellen	Ja	-	-

Beim Arbeiten mit der Graphik können Rebecca und Selina einige Schritte ohne Intervention durchführen. Allerdings haben sie große Schwierigkeiten, einen Start zu finden. In dem Moment, als die simulierten Ergebnisse in einer Tabelle vorliegen, wissen sie nicht, wie sie damit weiterarbeiten müssen und benötigen die starke Intervention, dass sie die Ergebnisse in einer Graphik darstellen müssen. Auch der Anfang der Gestaltung der Graphik gestaltet sich als schwierig, was auch daran liegt, dass ihnen in der Graphik zunächst keine Fälle angezeigt wurden, weil die Ergebnistabelle nicht markiert war. Nachdem das erste Merkmal (die neuen Gruppen) dargestellt sind, fällt ihnen das Arbeiten mit der Graphik in TinkerPlots leichter und es gelingt ihnen selbständig, das zweite Merkmal (Blutdruckwerte) passend darzustellen. Für die Benutzung des Lineals benötigen sie erneut starke Intervention, hier ergibt sich wieder eine Schwierigkeit durch die Software, weil die gemessene Differenz durch Verschieben der einzelnen Linealelemente nicht sichtbar ist. Dass die Differenz als Messgröße gesammelt werden muss, geschieht ebenfalls erst nach einer starken Intervention. Das Erstellen der Referenzverteilung geschieht jedoch wieder selbständig von den Beiden. Hier haben sie keine Schwierigkeiten. Zuvor hat sich gezeigt, dass sie eine bestimmte Erwartungshaltung bezüglich der Referenzverteilung hatten, möglicherweise ist das der Grund, wieso sie an dieser Stelle genau wissen, was sie tun müssen, um die Referenzverteilung zu erzeugen.

Aufgrund der vielen Schwierigkeiten, die dieses Paar hat, benötigen sie viel Zeit für die Umsetzung ihrer statistischen Ideen in TinkerPlots.

Tabelle 33: Optionale Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Rebecca und Selina

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Fälle in der Graphik stapeln	-	Ja	-
Beschleunigen des Messgrößensammelns	-	-	Ja

Bezüglich des Stapelns geschieht eine leichte Intervention durch die Interviewerin und wie erwartet wird starke Hilfe benötigt, um den Messgrößensammelvorgang zu beschleunigen.

Tabelle 34: Notwendige Schritte in TinkerPlots zur Ermittlung des P-Werts von Rebecca und Selina

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einteiler verwenden	Ja	-	-
Prozente einblenden	Ja	-	-

Das Ermitteln des P-Werts gelingt Rebecca und Selina selbständig.

11.2.3 Fallzusammenfassung Rebecca und Selina

In den vorigen beiden Abschnitten wurde ausführlich der Bearbeitungsprozess von Rebecca und Selina dargelegt und interpretiert. In diesem Abschnitt geschieht eine knappe Fallzusammenfassung entlang besonderer Auffälligkeiten (Voigt 1984).

Rebecca und Selina präsentieren sich auf statistischer Ebene stark, was sich unter anderem an ihren Argumentationen zum Konzept der zufälligen Aufteilung und dessen Bezug zur Nullhypothese zeigt. Unter Bezugnahme auf konfundierende Variablen wie Alter oder Geschlecht erklären die Beiden in der zweiten Interaktionseinheit, wieso eine zufällige Aufteilung der Versuchspersonen auf die Gruppen sinnvoll ist und versuchen auch einen möglichen Prozess zu dieser zufälligen Aufteilung aufzuzeigen, indem sie auf das Werfen einer Münze verweisen. Hieran (und an weiteren Stellen im Laufe des Interviews) zeigt sich, dass sich Rebecca und Selina in die Situation hineinversetzen und aus der Perspektive der Forscher die Aufgabe bearbeiten („… ob sie jetzt eine Münze geworfen haben…“ (66), „Wir sind Vertreter der Nullhypothese.“ (144)). Gleichzeitig üben sie Kritik am Randomisierungstest, indem sie die hypothetische Gleichverteilung bezüglich der relevanten Merkmale, die aus der zufälligen Aufteilung entstehen soll, zumindest implizit in Frage stellen (16). Das Konzept der zufälligen Aufteilung wird von ihnen im Laufe der Bearbeitung immer wieder adressiert und korrekt als Stützung einer Argumentation verwendet, z. B. als eine der möglichen Erklärungen für die beobachteten Unterschiede in den empirischen Daten, bei der Formulierung der Nullhypothese (85), bei der Erinnerung an die händische Simulation (100) und bei der Zusammenfassung der Simulation im Randomisierungstestschema (395). Allerdings wird die zufällige Zuordnung von den Beiden nicht als Argument bei den Schlussfolgerungen verwendet, obwohl es an dieser Stelle die entscheidende Begründung wäre. Im Gegensatz zu den Analysen von Pfannkuch et al. (2015) (Kapitel 6.1) können Rebecca und Selina die zufällige Aufteilung nicht nur als mögliche Erklärung nennen und erläutern, sondern beziehen sich in ihrem Bearbeitungsprozess auch immer wieder darauf.

Das Aufstellen der Nullhypothese, das häufig als schwierig angesehen wird (z. B. Biehler et al. 2015, DelMas et al. 2007) bereitet Rebecca und Selina keine Schwierigkeiten und innerhalb weniger Turns haben sie eine korrekte Nullhypothese aufgestellt (85), die sie im Folgenden testen wollen und für die sie vorab das Ziel formulieren, die-

se verwerfen zu wollen. Ihnen ist permanent bewusst, dass sie unter der in der Nullhypothese angenommenen Situation simulieren und diese ihrer Bearbeitung als wahr zugrunde legen.

Im Vergleich zur restlichen Bearbeitung werden mögliche Schlussfolgerungen aus dem gefundenen P-Wert von einem Prozent von Rebecca und Selina nur knapp formuliert. Auf die Stichworte des Randomisierungstestschemas gehen sie an dieser Stelle zum ersten Mal nicht vollständig ein. Auf statistischer Ebene verwerfen sie die Nullhypothese mit Verweis auf einen signifikanten P-Wert, halten gleichzeitig jedoch fest, dass dies kein Beweis für die Richtigkeit der Forschungshypothese sei. Sie diskutieren den P-Wert als Angabe einer Wahrscheinlichkeit für die Gültigkeit der Nullhypothese, verwerfen dies jedoch, was zunächst so aussieht, als ob das Wissen gefestigt ist, dass keine Wahrscheinlichkeitsangabe über eine Hypothese gemacht werden kann, was sich später in der letzten Interaktionseinheit jedoch wieder als fragiles Wissen herausstellt. Während der selbständigen Bearbeitung des Randomisierungstests ziehen die Beiden keine Schlussfolgerungen im Kontext, erst im Interviewteil wird der Bezug zum Kontext explizit auf Nachfrage formuliert. Hier zeigt sich ebenfalls das gute Verständnis dieses Paares. Denn sie formulieren zwar, dass somit gezeigt ist, dass Fischöl für die vierzehn Versuchsteilnehmer eine blutdrucksenkende Wirkung zu haben scheint, aber auch, dass dies nicht verallgemeinert werden könne und dass die Wirkung von Fischöl somit nicht bewiesen sei. Sie nehmen eine kritische Haltung gegenüber dem Ergebnis ein und fordern beispielsweise weitere Studien, um die Wirksamkeit von Fischöl weiter zu untersuchen.

Eine Episode, die sich im Bearbeitungsprozess von Rebecca und Selina als besonders kritisch herausgestellt hat, ist die Diskussion um das Umbenennen der Merkmale (Episode 8-4). Zunächst machen sie einen Anlauf, eine der beiden neuen Gruppen in „Fischöl“ umzubenennen, machen dies jedoch wieder rückgängig und diskutieren ausführlich, ob, und wenn ja wie, die neuen Gruppen benannt werden sollten. Hieran wird das Konzept der zufälligen Aufteilung der Versuchspersonen auf die Gruppen noch einmal ausführlich thematisiert und mit Stützungen ausgeführt, was dies bedeutet (vgl. Abbildung 200). Bei Noll und Kirin (2017) hat sich bei einer strukturell ähnlichen Aufgabe das Nichtumbenennen von Merkmalen als kritisch herausgestellt in dem Sinne, wenn es nicht erfolgte, dass daraus Schwierigkeiten in der weiteren Bearbeitung resultierten. Die vorliegende Fallstudie zeigt im Gegensatz hierzu, dass das Nichtumbenennen der Merkmale zu keinen Schwierigkeiten geführt hat, dass allerdings die Diskussion hierüber wichtige Erkenntnisse für die beiden Teilnehmerinnen geliefert hat.

Bezüglich des Umgangs mit der Software ist Rebecca und Selina bewusst, dass die Bedienung von TinkerPlots eine Schwierigkeit darstellt, aber scheinbar empfinden sie dies nicht als Hürde, um die Aufgabe zu bewältigen und auch nicht als Anlass, um zunächst

eine händische Simulation durchzuführen. Sie können die Anweisungen der Interviewerin sehr kompetent und flüssig umsetzen, so dass der Schluss nahe liegt, dass ihnen der Umgang mit der Software prinzipiell vertraut ist und ihnen lediglich Übung im Kontext von Randomisierungstests gefehlt hat. Es entsteht der Eindruck, dass Rebecca und Selina das Bedienen der Software an den Stellen, an denen sie zu einem statistischen Konzept explizit eine Erwartungshaltung formuliert haben, wie bei der Referenzverteilung und dem P-Wert, leichter fällt.

Über alle Interaktionseinheiten hinweg zieht sich ein sehr ausgewogenes Interaktionsverhältnis zwischen den Beiden. Im Dialog werden viele Konzepte von den beiden Teilnehmerinnen gemeinsam erarbeitet und Inhalte vertieft. Argumentationsstrukturen werden wechselseitig aufgegriffen und ergänzt, so dass im Dialog gemeinsame Deutung hervorgebracht wird.

Während ihres Lösungsprozesses orientieren sich Rebecca und Selina eng an dem Randomisierungstestschema und gehen nicht über die dort vorgeschlagenen Stichworte hinaus. Jedoch zeigen die sprachlichen Exkurse, dass sie stark an dem Thema Inferenzstatistik interessiert sind und dies für grundlegend wichtig halten.

11.3 Fallstudie Fabia und Laura

In diesem Abschnitt werden die zusammenfassenden Interpretationen der 15 Interaktionseinheiten von Fabia und Laura präsentiert. Das Vorgehen ist dabei analog zu dem der Analyse von Rebecca und Selina.

11.3.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Fabia und Laura

Interaktionseinheit 1: Vorstellung und Vorlesen des Aufgabenblatts (Z. 1-7)

In der ersten Interaktionseinheit geschieht wenig Interaktion. Es wird festgestellt, dass die beiden Teilnehmerinnen, Fabia und Laura, bereits im Vorfeld miteinander gearbeitet haben und „eingespielt“ sind. Daraufhin liest die Interviewerin das Aufgabenblatt komplett vor und erkundigt sich, ob es Fragen gebe. Dies wird direkt verneint, was nach der a priori Analyse in Kapitel 9.2.2 zu erwarten war.

Interaktionseinheit 2: Begründung für die zufällige Aufteilung auf die Gruppen (Z. 8-23)

In dieser Interaktionseinheit wird der Frage N1 des Leitfadens, weshalb die Versuchspersonen zufällig auf die Gruppen aufgeteilt wurden, nachgegangen.

8 I: Gut. Ähm (.) dann sehen wir jetzt hier, wenn man die Mittelwerte dieser beiden Gruppen betrachtet, einen Unterschied von 7,7 zwischen den Mittelwerten der beiden Gruppen. (..) Jetzt dazu meine erste Frage: Wieso wurden die Versuchspersonen, die hieran teilgenommen haben zufällig auf diese zwei Gruppen aufgeteilt?

9	L: Also ich kann mir vorstellen, wenn man jetzt weiß, man nimmt an so einer Studie teil und hat diese Vorahnung, das Fischöl hilft den Blutdruck zu senken (.) dann ähm möchte man auch lieber in dieser einen Gruppe sein. Weil man ja möchte, dass es was bringt und das ist dann vielleicht psychisch die Beeinflussung, wenn man sich das aussuchen könnte quasi.
10	I: Ah ok. Mhm (bejahend) Um quasi die Vorlieben der Teilnehmer auszugleichen?
11	L: Ja.
12	I: Ok.
13	F: Ja, also beide haben ja vorher auch an der Diät gar nicht teilgenommen. Die wurden dann (.) wirklich zufällig ausgelost. Ja DU nimmst jetzt an dieser Diät teil und DU an der. Sie sind vorher mit der/ mit dem Öl gar nicht in Verbindung gekommen deswegen (.). Ja ist das für die dann neu.
14	I: Mhm (bejahend).
15	F: Schwer zu erklären.
16	I: Was wäre denn gewesen, wenn jetzt zum Beispiel alle, die gerne Sport machen in ihrer Freizeit in die eine Gruppe gesollt hätten und ich sag mal alle Sportmuffel in die andere?
17	F: Dann wären die Merkmale ja nicht so verteilt gewesen, wie es normalerweise in der Gesellschaft ist, also wenn man/.
18	I: /Welche Merkmale?
19	F: Quasi jetzt in diesem Fall Sport. Also das hat dann bei sportlichen Leuten, hat das vielleicht andere Auswirkungen und dann kann man das nicht mehr sehen, das Ergebnis daraus. (..) Und das wäre ja auch keine zufällige Einteilung. Weil die ja schon gruppiert worden sind.
20	I: Genau deswegen die Frage: Warum wurde das denn zufällig eingeteilt?
21	F: Sonst müsste man das Ergebnis ja auf sportliche Teilnehmer schließen quasi.
22	I: Mhm (bejahend). Ok. (..) Noch irgendwas dazu?
23	F& L& I: (lachen).

In (8) beginnt die Interaktionseinheit mit der von der Interviewerin gestellten Frage „Wieso wurden die Versuchspersonen, die hieran teilgenommen haben, zufällig auf diese zwei Gruppen aufgeteilt?“. Daraufhin übernimmt Laura den Turn (9), und nimmt mit den Worten „Also ich kann mir vorstellen…“ auf Kontextebene die Perspektive einer Person ein, die an dem Experiment teilnimmt. Für eine solche Person formuliert sie eine Erwartungshaltung bezüglich der Wirkung, nämlich „… wenn man jetzt weiß, man nimmt an so einer Studie teil und hat diese Vorahnung, das Fischöl hilft den Blutdruck zu senken (.) dann ähm möchte man auch lieber in dieser einen Gruppe sein…“ Als Begründung liefert sie „weil man ja möchte, dass es was bringt“, woraus sie schließt, dass dies dann „psychisch die Beeinflussung [wäre], wenn man sich das aussuchen könnte quasi.“ Hiermit spielt sie auf das an, was in der Psychologie als „Hawthorne“-Effekt bekannt ist. Dieser wird definiert als

> „Der Hawthorne-Effekt beschreibt den verzerrenden Einfluss bei experimentellen Untersuchungen, bei denen nicht die durchgeführte experimentelle Manipulation sich auf die abhängigen Variablen auswirkt, sondern allein die Tatsache, dass eine Untersuchung durchgeführt wird.“ (Stangl 2018, Stichwort „Hawthorne-Effekt“)

Dabei wird allerdings nicht davon ausgegangen, dass Laura der Hawthorne-Effekt namentlich bekannt ist, sie beschreibt lediglich umgangssprachlich das Phänomen. Laura scheint davon auszugehen, dass die Versuchspersonen sich selbst eine Gruppe aussuchen könnten. Dies wird im nächsten Turn von der Interviewerin aufgegriffen, die das als „Vorlieben der Teilnehmer“ interpretiert, die durch die zufällige Zuteilung ausgegli-

chen werden sollen (10). Diese Interpretation wird von Laura bestätigt (11) und die Bestätigung wird von der Interviewerin wiederum bestätigt (12), so dass nun Einigkeit hierüber herrscht. Hierauf ergreift Fabia den Turn und versucht ebenfalls eine Erklärung zu geben, die jedoch schwer nachvollziehbar ist. Zunächst spricht sie von „beiden“, die „vorher an der Diät gar nicht teilgenommen [haben]“ und „Die wurden dann (.) wirklich zufällig ausgelost…“ (13). Bei ihr mag die Fehlvorstellung vorliegen, dass die Versuchspersonen zufällig ausgewählt wurden und nicht, dass sich die Zufälligkeit auf die Gruppenzuteilung bezieht. Im übernächsten Turn (15) gibt sie zu, dass dies „schwer zu erklären“ sei. Daraufhin wird von der Interviewerin die Frage gestellt, was gewesen wäre, wenn alle, die gerne Sport machten, in eine Gruppe eingeteilt worden wären und alle „Sportmuffel“ in die andere (16), gemäß der Nachfrage N1O des Leitfadens. Diesmal übernimmt Fabia den Turn und antwortet: „Dann wären die Merkmale ja nicht so verteilt gewesen, wie es normalerweise in der Gesellschaft ist, also wenn man“ (17) und bricht ab, weil sie von der Interviewerin unterbrochen wird. Interpretiert werden kann, dass sie den Umstand, dass jemand gerne Sport macht oder nicht, als Merkmal einer Person auffasst und eine bestimmte Vorstellung dazu hat, wie dieses Merkmal, zum Beispiel mit den beiden Ausprägungen ‚macht gerne Sport‘ und ‚macht nicht gerne Sport‘ in der Gesellschaft verteilt ist. Es kann interpretiert werden, dass hinter diesem „normalerweise in der Gesellschaft“ verteilt die Vorstellung einer zufälligen Verteilung steht. Auf die unterbrechende Nachfrage „Welche Merkmale?“ (18) der Interviewerin erklärt sie: „Quasi jetzt in diesem Fall Sport. Also das hat dann bei sportlichen Leuten, hat das vielleicht andere Auswirkungen…“ (19). Sie beantwortet also zunächst die Frage und dann lässt sich interpretieren, dass sie die Vorstellung hat, dass Sportverhalten einen Einfluss auf die Blutdrucksenkung hat, was sie mit dem „das“ andeutet. Sie stellt nun fest, dass die von der Interviewerin vorgeschlagene Einteilung (16) keine zufällige Einteilung mehr sei, sondern eine Gruppierung. Dies ist eine richtige Feststellung, die jedoch für die Frage, warum die Versuchspersonen zufällig eingeteilt wurden, keine Erklärung liefert, weshalb die Interviewerin im nächsten Turn (20) diese Frage noch einmal stellt. Jetzt liefert Fabia die Erklärung: „Sonst müsste man das Ergebnis ja auf sportliche Teilnehmer schließen quasi.“ (21). Aus dieser etwas holprigen Formulierung lässt sich interpretieren, dass Fabia meint, dass das Ergebnis des Experiments, was auch immer sie ggf. darunter versteht, auf die Sportlichkeit der Teilnehmer zurückzuführen wäre und nicht auf die Wirksamkeit eines bestimmten Öls. Es wird geklärt, dass kein weiterer Gesprächsbedarf hierzu vorliegt (22-23).

Fabia baut hier eine Argumentation auf, die wie in Abbildung 206 schematisch dargestellt werden kann.

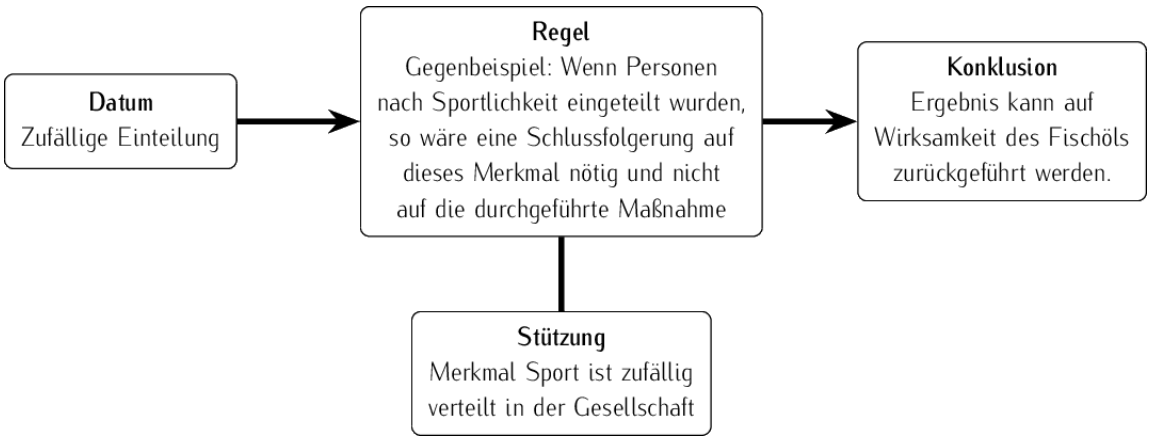


Abbildung 206: Rekonstruierte Argumentation zur Wirksamkeit von Fischöl anhand eines Gegenbeispiels, aufbauend auf Fabias Aussagen

Fabia geht davon aus, dass die zufällige Einteilung der Versuchspersonen zu den Gruppen erfolgt ist. Zur Konklusion, dass dadurch auf die Wirksamkeit des Fischöls geschlossen werden kann, was von ihr jedoch nicht so formuliert, sondern durch die Analyse an dieser Stelle interpretiert wird, kommt sie über die Regel eines Gegenbeispiels. Das Gegenbeispiel wird dadurch gestützt, dass sie zugrunde legt, dass das Merkmal dieses Gegenbeispiels zufällig verteilt ist, wobei Fabia hier von „wie es normalerweise in der Gesellschaft ist“ (17) spricht, also nicht ganz klar ist, ob sie damit tatsächlich die zufällige Verteilung meint. Abbildung 206 stellt insofern eine mögliche Argumentation dar, die auf Teilen von Fabias Aussagen basiert.

Interaktionseinheit 3: Mögliche Erklärungen für die beobachteten Unterschiede (Z. 24-35)

Die dritte Interaktionseinheit wird durch die Frage N2 des Leitfadens zu möglichen Erklärungen für die beobachteten Unterschiede bestimmt.

24	I: In Ordnung. Hmm dann, wenn ihr euch die Unterschiede anguckt: Welche möglichen Erklärungen gibt es denn überhaupt? (..) Für diese Unterschiede, die beobachtet werden zwischen den beiden Gruppen.
25	L: Also man sieht ja, dass es nur wirklich sieben in jeder Gruppe sind. Da kann man ja einmal davon sprechen. Bei zweien in der Fischöl gibt es ja gar kein Ergebnis quasi, also da kann man gar nichts davon sagen, dass es überhaupt eine Senkung gab.
26	I: Du meinst gar keine Veränderung.
27	L: Gar keine Veränderung, ja. Und ähm bei den vielen kann es halt entweder daran liegen, dass wirklich das Fischöl was bringt, also dass es wirklich dazu führt, dass der Blutdruck gesenkt wird. (.) Oder es war Zufall. Oder sie haben irgendwas in ihrem Leben anders gemacht in den vier Wochen. Das weiß man ja nicht.
28	F: Ja und dann wieder diese psychischen Faktoren, dass man daran denkt Fischöl ist wirksamer als normales Öl, irgendwas Besonderes.
29	I: Mhm (bejahend) das heißt wir hätten einmal die Erklärung als Möglichkeit: Fischöl ist wirksam, was ihr gerade gesagt habt. Und die andere Erklärung wäre sozusagen: Faktoren, die wir gar nicht beeinflussen können, wie psychische Faktoren oder was, was nicht abgefragt wurde: Machen die mehr Sport in der einen Gruppe, und so weiter. (..) Ok und (.) es gibt noch eine Er-

	klärung. Wenn ihr nochmal an die Vorlesung denkt ähm (.) Da hatten wir immer noch eine mögliche Erklärung, für quasi so Unterschiede, die man beobachtet. Fällt euch das noch ein?
30	F: Ich würde sagen die zufällige Einteilung. Weil irgendein Satz war da immer. Grund war/ sind nicht die Unterschiede, sondern zufällige Einteilungen der Gruppenmitglieder. Irgendwie so was (lacht).
31	L: Ich kann mich gerade gar nicht erinnern.
32	I: Genau das ist richtig erinnert. Und wenn wir da nochmal drüber nachdenken. Wir hatten ganz am Anfang dieser zwei Vorlesungen hatten wir was über Stichprobenschwankungen gemacht. (.) Da ging es darum, wenn ich mir quasi eine Stichprobe ziehe aus einer bestimmten Population/ hier in dem Fall ist die Stichprobe nicht zufällig, sondern es sind einfach vierzehn Freiwillige gewesen. Deren Werte sind zufällig verteilt. Und wenn ich die jetzt anders genommen hätte/ wenn ich eine andere Stichprobe genommen hätte (.) dann hätte das zum Beispiel anders aussehen können.
33	F: Mhm (bejahend) ja.
34	I: Ja, das ist eine weitere Erklärung. Also, wenn ich die Personen zufällig auf die Gruppen einteile. Gehen wir mal davon aus, Fischöl hat gar keine Wirkung. Es ist egal in welcher Gruppe ich bin und (.) dann könnte es ja sein, dass einfach nur durch die Einteilung/ wie die Leute in die Gruppe eingeteilt wurden, bestimmte Unterschiede aufgetreten sind und das wäre dann sozusagen eine weitere Erklärung.
35	F: Ja.

Die Einheit wird eingeleitet durch die Frage, die die Interviewerin stellt (24). Wieder ergreift Laura den Turn, um zu antworten. Sie hebt zunächst hervor, dass sieben Personen in jeder Gruppe seien (25). Aus welchem Grund sie das betont, wird nicht klar. Eventuell hat sie dies in diesem Moment erst bewusst wahrgenommen und erwähnt es deshalb. Eine alternative Deutung ist, dass ihr eine Gruppengröße von jeweils sieben Personen als sehr klein erscheint, das lässt sich durch das „nur“ stützen. Dann weist sie im gleichen Turn auf Kontextebene auf zwei Personen aus der Fischölgruppe hin, bei denen keine Senkung vorläge. Damit bezieht sie sich auf die zwei Personen, die als Ergebnis nach der vierwöchigen Diät eine Blutdruckreduktion von 0 mmHg aufweisen. Sie beantwortet also die Frage nicht, sondern schaut noch einmal genauer die Daten an und formuliert hierzu Erkenntnisse. Dann führt sie nach einem Einwurf durch die Interviewerin (26) ihre Erläuterung auf Kontextebene in (27) weiter. Sie stellt nun fest, dass „… bei den vielen kann es halt entweder daran liegen, dass wirklich das Fischöl was bringt, dass es wirklich dazu führt, dass der Blutdruck gesenkt wird. (.) Oder es war Zufall. Oder sie haben irgendwas in ihrem Leben anders gemacht in den vier Wochen. Das weiß man ja nicht.“ (27). Mit „den vielen“ sind vermutlich die fünf Personen gemeint, bei denen eine Blutdruckreduktion echt größer als 0 mmHg verzeichnet wurde. Mit Blick auf den Erwartungshorizont, der in Kapitel 9.2.2 beschrieben ist, ist das eine sehr gute Antwort, da sie alle drei möglichen Erklärungen aufführt, die in Frage kommen. Ihre erste Erklärung ist, dass Fischöl wirksam ist, die zweite Erklärung ist, dass es am Zufall liegt – womit sie hoffentlich die zufällige Einteilung meint, und ihre dritte Erklärung lässt sich interpretieren, dass es an weiteren Einflussfaktoren liegt. Sie hat also alle drei möglichen Erklärungen im Blick und formuliert diese knapp. Fabia übernimmt anschließend den Turn und führt aus ihrer Sicht eine Erklärung an, die sie als Ergänzung

zu Lauras Erklärung sieht: „Ja und dann wieder diese psychischen Faktoren…“ (28). Mit dem „Ja“ deutet sie Einverständnis mit den von Fabia genannten Erklärungen an, das „und dann wieder“ zeigt auf eine ergänzende Erklärung hin, die sie nun formuliert. Hier nennt sie nun die „psychischen Faktoren“, die sie bereits in der Einheit zuvor genannt hat als Beispiel, wieso eine zufällige Einteilung der Personen auf die Gruppen geschehen solle (9). Dies ist jedoch keine weitere Erklärung zu den von Fabia genannten drei, sondern ein mögliches Beispiel zu Fabias dritter Erklärung, also eine mögliche konfundierende Variable. Dies wird allerdings von beiden nicht erkannt, oder vielleicht auch nur nicht ausgesprochen. Deshalb fasst die Interviewerin dies im nächsten Turn zusammen und bringt die von Laura genannten „psychischen Faktoren“ mit der dritten Erklärung von Fabia zusammen (29). In ihrer Zusammenfassung fehlt eine weitere Erklärung, nach der sie am Ende ihres Turns fragt. Vermutlich hat sie die weitere Erklärung von Fabia mitten im Turn (27) vorher „… Oder es war Zufall…“ überhört, da sie nun in der Zusammenfassung diese nicht aufführt. Fabia greift dies jedoch sofort auf und führt die zufällige Einteilung an mit dem Hinweis, dass da immer so ein Satz gewesen sei (30), womit nur die Lerneinheit gemeint sein kann, in welcher dies formuliert wurde als *Die zufällige Einteilung der Personen zu den Gruppen ist eine mögliche Erklärung für die beobachteten Unterschiede* (vgl. Kapitel 8.2.2.3). Am Ende ihres Turns sagt sie „Irgendwie so was“, womit die Formulierung gemeint ist, die von ihr nicht sauber gegeben wird und lacht auffordernd. Hierauf übernimmt Laura den Turn (31) und beschreibt ihr Unwissen hierzu, dass sie sich nicht erinnern könne. Einerseits kann dies bedeuten, dass sie sich ähnlich wie Fabia nicht an die genaue Formulierung erinnern kann, wie die zufällige Einteilung der Versuchspersonen die beobachteten Unterschiede erklären kann, andererseits kann sie aber auch meinen, dass sie sich an die gesamte Erklärung, in der die zufällige Einteilung der Personen auf die Gruppen eine Rolle spielt, nicht mehr erinnert. Die erste Deutung ist jedoch wahrscheinlicher, da sie bereits im Turn (27) den Zufall als mögliche Erklärung angeführt hat. In den nachfolgenden Turns führt die Interviewerin die Erklärung über die zufällige Einteilung aus und nennt auch den Aspekt, dass es sich um keine Zufallsstichprobe handele, sondern um 14 Freiwillige. Dieses Wissen wird also von ihr und nicht von den Teilnehmerinnen formuliert. Fabia wirft kurze, bestätigende Turns ein, wodurch interpretiert wird, dass sie den Erläuterungen der Interviewerin folgt und diese versteht. Während der gesamten Einheit werden die Kontextebene und die Statistikebene miteinander in Bezug gesetzt.

Interaktionseinheit 4: Überleitung zum Randomisierungstest (Z. 36-43)

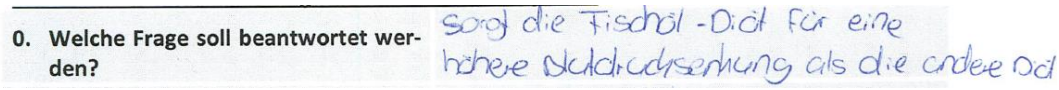
Dies ist eine kurze Überleitungseinheit, die nicht analysiert wird, sondern kurz zusammengefasst wird. In der vierten Interaktionseinheit leitet die Interviewerin zum Randomisierungstest über, den die Studierenden nun durchführen sollen. Sie weist auf die Hilfsmittel hin, die benutzt werden können (Randomisierungstestschema, leere Zettel, TinkerPlots), woraufhin Laura nachfragt, ob sie sich aussuchen können, „ob wir das mit

den Zetteln machen oder mit äh TinkerPlots“ (36). Dies wird bestätigt mit dem Hinweis, dass auch beides benutzt werden dürfe.

Interaktionseinheit 5: Forschungsfrage notieren (Z. 44-53)

In der fünften Interaktionseinheit wird die Forschungsfrage von Laura und Fabia formuliert.

44 L: Hmm. Gibt es bei Personen, die an der Fischöldiät teilnehmen, wird da der Blutdruck mehr gesenkt, als bei denen, die bei einer normalen Diät teilnehmen?
45 F: Ja. Könnte man das denn knapp aufschreiben?
46 L: Ist ja auch eigentlich quasi das gleiche wie die./ Obwohl die Forschungshypothese ist dann dass ähm (.) Fischöl eine Auswirkung auf die Senkung ist (unv.) (..) Willst du schreiben, soll ich schreiben?
47 F: Sollen wir da anfangen? (*zeigt auf das Feld „Nullmodell“*)
48 L: Ja, wir können ja auch oben anfangen.
49 F: Ok, welche Frage sollen wir nehmen?
50 L: Gibt es bei Personen, die bei der Fischöldiät teilnehmen (...) eine höhere Senkung. (..) Sorgt die Fischöldiät für eine höhere Blutdrucksenkung/.
51 F: /Blutdruckreduktion.
52 L: Als die andere Diät?
53 *Es wird notiert:*



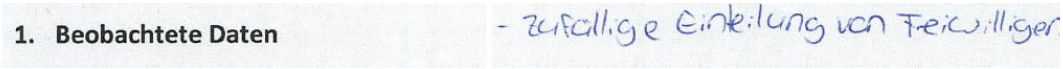
Laura steigt direkt mit der Formulierung der Forschungsfrage im Kontext in Turn (44) ein. Die von ihr genannte Frage scheint Fabia zu lang zu sein, denn sie fragt daraufhin, ob man dies knapp aufschreiben könne (45), was zeigt, dass sie inhaltlich offenbar einverstanden ist. Daraufhin zieht Laura einen Vergleich auf statistischer Ebene zur Forschungshypothese: „Ist ja auch eigentlich quasi das gleiche wie die./ Obwohl die Forschungshypothese…“ (46). Sie erkennt hier richtig, dass inhaltlich die Forschungsfrage und die Forschungshypothese die gleiche Perspektive im Blick haben, nämlich dass Fischöl wirksam ist. Dies sollte jedoch in der Forschungsfrage als Frage formuliert werden und in der Forschungshypothese als Aussage bzw. Behauptung. Laura versucht an dieser Stelle bereits eine Formulierung der Forschungshypothese mit „… dass ähm Fischöl eine Auswirkung auf die Senkung ist…“ (46), was mit sehr leiser Stimme hervorgebracht wird und in undeutlicher Sprechweise endet. Dies wird als Indiz darauf gedeutet, dass sie sich über die Formulierung der Forschungshypothese nicht sicher ist, obwohl sie diese hier richtig ausdrückt. Sie bricht ab, mit der Frage an Fabia, ob diese schreiben wolle, womit sie zeigt, dass sie nun eine Reaktion von ihr einfordert, nicht nur auf die konkrete Frage, wer schreiben solle, sondern auch bezüglich der Formulierungen. Fabia nimmt auf, dass Laura bereits die Forschungshypothese zu formulieren versucht hat, indem sie auf das Feld Nullmodell des Randomisierungstestschemas zeigt und fragt, ob sie dort anfangen wollen (47). Dies wird von Laura mit dem Gegenvorschlag bedacht, dass sie „oben“ anfangen können (48), womit das Feld der Forschungs-

frage gemeint sein muss, die zu formulieren das eigentliche Ziel der begonnen Interaktionseinheit war. Somit wird von ihr das Schema als Strukturierung des Prozesses genutzt. Fabia geht darauf erneut mit einer Gegenfrage ein, welche Frage denn nun genommen werden solle (49). Wie bereits zu Beginn der Einheit liefert wieder Laura einen Formulierungsvorschlag für die Forschungsfrage mit „Gibt es bei Personen, die bei der Fischöldiät teilnehmen (…) eine höhere Senkung. (..) Sorgt die Fischöldiät für eine höhere Blutdrucksenkung.“ (50). Von Fabia wird das Wort „Blutdruckreduktion“ (51) genannt, wahrscheinlich als Alternative zu Lauras „Blutdrucksenkung“. Laura ergänzt fragend im nächsten Turn „Als die andere Diät?“, was nicht weiter kommentiert wird, aber im nächsten Schritt notiert wird als „Sorgt die Fischöl-Diät für eine höhere Blutdrucksenkung als die andere Diät?“ (53). Somit ist im gemeinsamen Gespräch der Beiden eine zufriedenstellende Formulierung der Forschungsfrage gefunden. Dies geschah ohne weitere Schwierigkeiten mit Formulierungen im Kontext, zu denen die Ebene der Statistik, beispielsweise über den Verweis auf die Forschungshypothese einbezogen wurde.

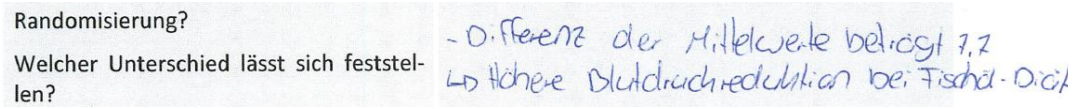
Interaktionseinheit 6: Beobachtete Daten (Z. 54-74)

Die beobachteten Daten zu untersuchen und Unterschiede zwischen den Gruppen herauszuarbeiten ist Bestandteil der sechsten Interaktionseinheit.

54 L: Randomisierung ist ja zufällige Einteilung/.
55 F: /Mhm (bejahend)
56 L: In zwei Gruppen. Zufällige Einteilung von Freiwilligen vielleicht.
57 *Es wird notiert:*



58 F: Mhm (bejahend) Welcher Unterschied lässt sich feststellen?
59 L: Ähm Differenz des Mittelwertes von 7,7.
60 F: Also, müssen wir den Teil (unv.).
61 L: Ja.
62 F: Ich weiß noch bei dem anderen hatten wir hingeschrieben: Differenz der Mediane.
63 L: Ja und hier haben wir jetzt den Mittelwert, der ist hier ja auch abgebildet. Vielleicht nehmen wir dann den/.
64 F: /Also Differenz/ Differenz der Mittelwerte. Der Blutdrucksenkung.
65 L: Welcher Wert soll überprüft werden, weil da müssen wir ja die Differenz der Mittelwerte. Hier müssen wir nur erst mal nur sagen welche/.
66 F: /Soll ich schreiben: Differenz?
67 L: Unterschiede/welcher Unterschied wirklich feststellbar ist. Und es ist ja festlegbar, dass die Mittelwer/ äh der Mittelwert der Fischöl (.) Studie viel höher ist. Also dass da wirklich eine höhere Blutdruckreduktion war, oder? (..) Weil dann ist ja/ das haben wir beobachtet quasi. (...) Differenz der Mittelwerte beträgt 7,7. Und dann Fischöldiät. (..) Höhere Blutdrucksenkung.
68 F: Soll ich einen Folgepfeil machen?
69 L: Ja (unv.)
70 F: Oder müssen wir es ausformulieren? (*Blickt zur Interviewerin*)
71 I: Nein.
72 F: Ok. (...) Hmm Fischöl/ was hattest du gesagt? (..) Blutdrucksenkung der Fischöldiät beträgt.
73 L: Höhere Blutdruckreduktion durch Fischöldiät, bei Fischöldiät.
74 *Es wird notiert:*


Randomisierung?
Welcher Unterschied lässt sich feststellen?
- Differenz der Mittelwerte beträgt 7,7
↳ höhere Blutdruckreduktion bei Fischöl-Diät

Nachdem die Forschungsfrage in der vorigen Interaktionseinheit notiert wurde, nimmt Laura direkt das Stichwort *Randomisierung* des Randomisierungstestschemas auf und erläutert es als „Randomisierung ist ja zufällige Einteilung“ (54). Eine Überleitung zwischen den Einheiten fünf und sechs findet somit nicht statt. Zu Beginn der Einheit ruft sie sich ins Gedächtnis, was das Wort Randomisierung bedeutet, dies hat sie kennengelernt während des Bausteins Inferenzstatistik als zufällige Einteilung, was so von ihr hier auf statistischer Ebene wiedergegeben wird. Fabia bestätigt dies (55), woraufhin Laura ihre Aussage in (56) fortführt mit Bezug zum Kontext: „In zwei Gruppen. Zufällige Einteilung von Freiwilligen vielleicht.“ Den Aspekt, dass Freiwillige am Experiment teilnahmen, ergänzt sie hier mit einem „vielleicht“, was interpretiert werden kann als Vorschlag, diesen Aspekt in die schriftliche Formulierung aufzunehmen. Andererseits könnte es auch bedeuten, dass sie unsicher ist, ob dies entweder richtig ist, oder ob dies überhaupt aufgenommen werden soll. Fabia nimmt es schriftlich auf als „Zufällige Einteilung von Freiwilligen“ (57) und sieht dies als ausreichend an, da sie sofort mit der nächsten Frage des Schemas fortfährt: „Welcher Unterschied lässt sich feststellen?“ (58). Laura beantwortet mit „Differenz des Mittelwerts von 7,7“ (59), welches direkt aus der Graphik, die auf dem Arbeitsblatt dargestellt ist, abgelesen werden konnte. Bei den nächsten beiden Turns (60) und (61) bezieht sich Fabia auf etwas, das interpretativ nicht rekonstruiert werden kann, das jedoch auf einer Metaebene zum Vorgehen zwischen den Beiden angesiedelt zu sein scheint. Im nächsten Turn (62) bezieht sich Fabia darauf, dass sie „bei dem anderen“ hingeschrieben hätten „Differenz der Mediane“. „Bei dem anderen“ ist vermutlich ein anderes Schema, das sie ausgefüllt haben, dies kann sich nur auf die letzte Seminarsitzung beziehen, in welcher die Studierenden das Randomisierungstestschema am Beispiel der Gedächtnisleistung zum ersten Mal benutzt hatten und mit der Differenz der Mediane arbeiten sollten (vgl. Kapitel 8.2.2.2). Laura erkennt richtig, dass hier die arithmetischen Mittel und nicht die Mediane dargestellt sind und schlägt vor, diese zu benutzen (63). In den folgenden Turns handeln sie präziser aus, was überprüft werden soll (64-66). In (67) wird von Laura präzisiert, dass „… die Mittelwert/äh der Mittelwert der Fischölstudie viel höher ist…“, was sie zur Konklusion bringt, dass „… da wirklich eine höhere Blutdruckreduktion war, oder?...“ und führt als Begründung an, dass die „… Differenz der Mittelwerte 7,7 [beträgt]…“. Dies lässt sich als Argumentation im Sinne Toulmins darstellen wie in Abbildung 207.

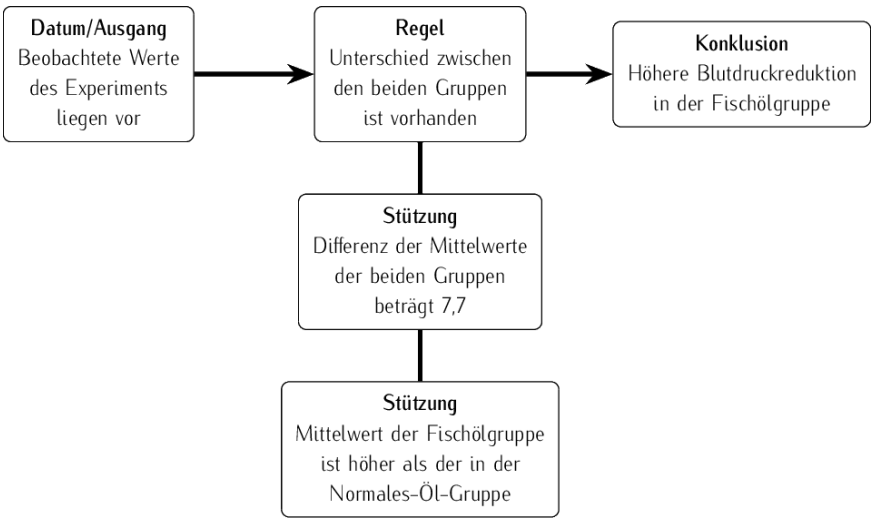


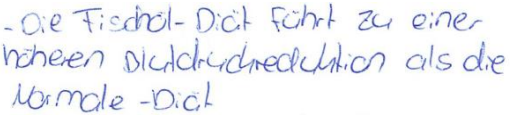
Abbildung 207: Argumentation von Fabia und Laura zu den beobachteten Daten

In den weiteren Turns handeln die beiden die schriftliche Formulierung aus, die in „Differenz der Mittelwerte beträgt 7,7 → Höhere Blutdruckreduktion bei Fischöl-Diät“ (74) resultiert. Fabia und Laura nennen hier zunächst die Differenz von 7,7 zwischen den beiden Gruppen, die sie im zweiten Satz präzisieren durch die Angabe, dass in der Fischölgruppe die höheren Werte vorkommen.

Interaktionseinheit 7: Nullmodell (Z. 75-93)

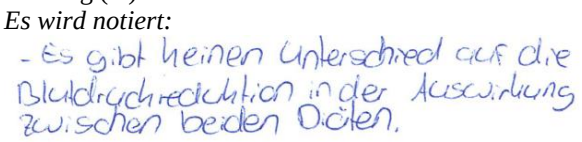
Null- und Forschungshypothese werden in der siebten Interaktionseinheit aufgestellt.

75 L: Ja (...) Forschungshypothese ist ja dann eigentlich, dass die Fischöldiät zu einer höheren/ man muss ja eigentlich eher von den Ergebnissen sprechen, oder?
76 F: Von den/ nee ich glaube es reicht, wenn wir schreiben: Es muss ja eine Frage (...) bei Forschern einfach so die Fisch/ die Fischöldiät führt zu einer/ ist ja eigentlich so was ähnliches wie hier (*zeigt auf die Frage, die zuvor formuliert wurde*) wieder.
77 L: Ja nur als Feststellung, als Hypothese dann quasi.
78 F: Ja.
79 L: Das äh/ die Fischöldiät führt zu einer höheren Reduktion als die "Normale Öl Diät"
80 *Es wird notiert:*



81 F: Gut, dass ich die Reflexion erst vor kurzem geschrieben hatte.
82 L: Einmal nochmal angeguckt.
83 F: Ja irgendwann musste ich es ja machen. Es gibt keinen Unterschied (...) zwischen den beiden Diäten. Und die beiden Diäten führen zu keinem unterschiedlichen Blutdruck.
84 L: Mhm (bejahend) Die haben keine Auswirkung auf die Blutdruckreduktion.
85 F: Und da ging es ja glaube ich um die beiden Diäten. Dass da kein Unterschied zwischen ist/ zwischen denen ist kein Unterschied.
86 L: Es gibt keinen Unterschied zwischen beiden Diäten (...) in Hinblick auf die/ es muss ja auf die Blutdruckreduktion (...) zielen.
87 F: Es gibt keinen Unterschied zwischen beiden Diäten in Hinblick auf die Blutdruckreduktion.
88 L: Auf die Auswirkung (unv.) (...) Ich hasse so was.

89 F: Ähm, es lässt sich kein Unterschied feststellen zwischen den Auswirkungen (...) der beiden Diäten auf die Blutdruckreduktion.
90 L: Denk dir was aus. (lacht)
91 F: Man weiß ja, was man sagen will, aber/ (...) Es gibt keinen Unterschied auf (...) die/ (...) Es gibt keinen Unterschied auf.
92 L: Zwischen den beiden Diäten. Das passt aber irgendwie auch nicht wirklich. (...) In der Auswirkung (...) zwischen beiden Diäten.
93 *Es wird notiert:*



Sowohl der Forschungshypothese als auch der Nullhypothese nähern sich Laura und Fabia im Gespräch und entwickeln gemeinsam eine Formulierung. Keine der beiden Hypothesen wird auf Anhieb formuliert, sondern einzelne Sprachbausteine werden zusammengesetzt. Die Interaktionseinheit beginnt direkt mit dem Einstieg von Laura in (75) mit dem ersten Formulierungsversuch für die Forschungshypothese im Kontext: „…Forschungshypothese ist ja dann eigentlich, dass die Fischöldiät zu einer höheren/…“. Sie bricht hier die Aussage ab, um die Frage anzuschließen, ob hier von Ergebnissen gesprochen werden müsse. Zunächst fällt ihr also die Formulierung nicht leicht. Fabia übernimmt den Turn, und geht nur kurz mit „… nee ich glaube es reicht, wenn wir schreiben…“ (76) auf Lauras Frage ein. Sie spricht erst aus, dass es eine Frage sein solle, macht dann einen Formulierungsversuch mit „… die Fischöldiät führt zu einer….“, bricht ebenfalls ab und schließt mit dem Verweis auf die Forschungsfrage, dass dies ähnlich zu formulieren sei. Im ersten Anlauf zeigen also beide Schwierigkeiten, die Forschungshypothese zu formulieren. Mit Blick auf die als „Sorgt die Fischöl-Diät für eine höhere Blutdrucksenkung ans die andere Diät?“ (53) formulierte Forschungsfrage stellt Laura daraufhin fest, dass diese nun als Feststellung, als Hypothese zu formulieren sei (77). Fabia bestätigt (78) und Laura formuliert daraufhin, dass „die Fischöldiät zu einer höheren Reduktion als die normale Öl-Diät [führt]“ (79) und hält sich damit sehr eng an die bereits formulierte Forschungsfrage. Da Forschungsfrage und Forschungshypothese in Bezug zueinander stehen, stellt dies eine gute Strategie dar, um eine Orientierung für die Formulierung der Forschungshypothese zu erhalten. Laura präzisiert dahingehend, dass sie „die andere Diät“ durch „die normale Diät“ ersetzt und auch so notiert (80).

In den nächsten beiden Turns (81-82) nehmen Fabia und Laura Bezug auf die Reflexion, die sie als Teil der Studienleistung für das Seminar verfassen mussten, was offensichtlich erst kurz zuvor geschah. Anscheinend hat ihr das geholfen, hier eine Formulierung zu finden.

In (83) beginnt Fabia dann mit der Formulierung der Nullhypothese, ohne diesen Bezug explizit herzustellen. „… Es gibt keinen Unterschied (…) zwischen den beiden Diäten.

Und die beiden Diäten führen zu keinem unterschiedlichen Blutdruck.“ Damit schlägt sie zwei Formulierungsalternativen vor, von denen die erste Formulierung sehr unspezifisch auf „keinen Unterschied zwischen den Diäten“ verweist und die zweite Formulierung das Resultat in den Blick nimmt, nämlich den nicht unterschiedlichen Blutdruck. Laura formuliert das zweite mit anderen Worten im nächsten Turn: „… Die haben keine Auswirkung auf die Blutdruckreduktion“ (84) und verbindet somit die Diäten mit der Wirkung. In den folgenden drei Turns versuchen die beiden sich einer genaueren Formulierung zu nähern, was in (88) von Laura mit „… ich hasse sowas“ kommentiert wird. Hiermit kann nur die Suche nach der richtigen Formulierung für die Nullhypothese gemeint sein. Fabia versucht daraufhin erneut eine Formulierung mit „Ähm, es lässt sich kein Unterschied feststellen zwischen den Auswirkungen (…) der beiden Diäten auf die Blutdruckreduktion.“ (89). Dies wäre eine gute Formulierung der Nullhypothese. Laura reagiert hierauf jedoch nur mit „Denk dir was aus“ (90), womit sie entweder zeigt, dass sie sich nicht mehr mit der Formulierung beschäftigen möchte und dies Laura überlässt oder möglicherweise die gerade vorgeschlagene Formulierung akzeptiert. Fabia betont in (91) noch einmal, dass ihr klar ist, worauf sie hinauswill („Man weiß ja, was man sagen will, aber…“), dass dies jedoch trotzdem schwierig zu formulieren ist, ist der Halbsatz, den sie auslässt. Sie beginnt eine weitere Formulierung mit „… Es gibt keinen Unterschied auf“, der jetzt wieder von Laura im nächsten Turn fortgeführt wird mit „Zwischen den beiden Diäten…“ (92) und noch präzisiert wird durch „… in der Auswirkung (..) zwischen den beiden Diäten“. „Es gibt keinen Unterschied auf die Blutdruckreduktion in der Auswirkung zwischen beiden Diäten“ (93) wird von den beiden schließlich als Nullhypothese notiert. Dies ist eine richtige Formulierung der Nullhypothese, die jedoch die im Baustein Inferenzstatistik gelernten Textbausteine außer Acht lässt. In der Formulierung ist weder Bezug genommen auf die zufällige Aufteilung der Personen auf die Gruppen noch auf die beobachteten Unterschiede. Positiv hervorzuheben ist, dass sie eine Formulierung der Nullhypothese im Kontext wählen.

Interaktionseinheit 8: Simulieren/Zufallsmaschine erstellen (Z. 94-172)

In der achten Interaktionseinheit wird die in der Nullhypothese ausgedrückte Situation mit Hilfe der Zufallsmaschine von TinkerPlots modelliert. Die Einheit wird in sechs Episoden gegliedert. In der ersten Episode besprechen die beiden Teilnehmerinnen ihr Vorgehen, in der zweiten Episode werden erste Einstellungen an der Zufallsmaschine vorgenommen. Die dritte Episode handelt von der Erstellung der zwei neuen Gruppen. In der vierten Episode wird die Einstellung ohne Zurücklegen diskutiert. In der fünften Episode befassen sich die Beiden noch einmal mit weiterer Planung und in der sechsten Episode wird die Durchgangsanzahl angepasst.

Episode 8-1: Besprechen des Vorgehens (Z. 94-101)

94 F: Jetzt müssen wir randomisieren. Wollen wir das mit TinkerPlots machen? (unv.) Da (*zeigt auf*

	die leeren Zettel) kann man gar nicht so oft machen, wie es nötig ist, oder?
95	L: Wenn du noch weißt, wie das funktioniert.
96	F: Ich kann es ja mal versuchen, sonst können wir ja fragen.
97	I: Ich bin ja hier, ich helfe euch gerne.
98	F: Brauchen wir die ganzen Werte hier (*zeigt mit dem Finger auf die Tabelle in TinkerPlots*)?
99	*Es wird versucht, das Touchpad des Computers zu benutzen, dieses ist jedoch deaktiviert.*
100	I: Äh das funktioniert nicht, du musst die Maus benutzen.
101	F: Ah ok. (..) Die ganzen Werte und dann müssen wir die in der Zufallsmaschine in zwei verschiedene Stapel einteilen, (...) oder? Ich mach mal die auf. (...) So. Da hatten wir doch Stapel.

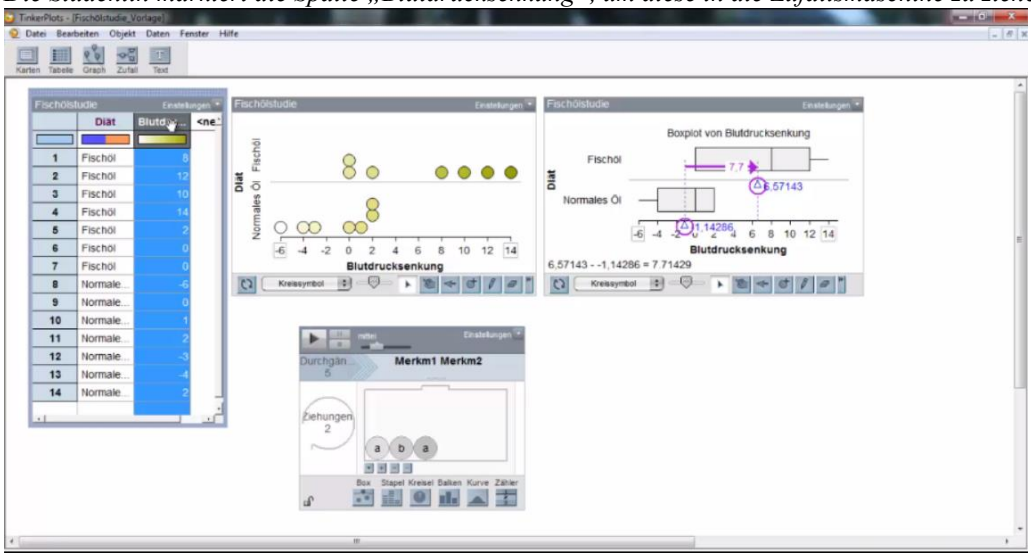
Fabia eröffnet die Einheit in (94) mit „Jetzt müssen wir randomisieren…“. Sie zeigt ihre Präferenz durch die erste Frage „… Wollen wir das mit TinkerPlots machen?“ und liefert gleich eine Begründung, weshalb sie nicht die händische Simulation über die Zettel machen möchte: „… Da (*zeigt auf die leeren Zettel*) kann man gar nicht so oft machen, wie es nötig ist, oder?“. Dies zeigt, dass sie genug Vertrauen in die eigene Kompetenz zu haben scheint, um den Randomisierungstest mit TinkerPlots durchführen zu können. Ihr scheint ebenfalls bewusst zu sein, dass die händische Simulation und die Computersimulation auf gleicher Ebene stehen, dass jedoch eine große Wiederholungszahl durch die händische Simulation nicht zu leisten ist. Sie quantifiziert an dieser Stelle nicht, wie häufig „es nötig“ ist, im Baustein Inferenzstatistik wurden hierzu Wiederholungszahlen von 1000 und 10 000 verwendet. Durch das nachgestellte Fragewort „oder?“ zeigt sie an, dass ihr Turn beendet ist und sie auf Zustimmung von Laura wartet. Laura übernimmt den Turn und antwortet „Wenn du noch weißt, wie das funktioniert.“ (95). Das lässt mehrere Deutungsalternativen zu. Erstens kann es bedeuten, dass sie einverstanden ist mit dem Vorschlag, TinkerPlots zu verwenden. Zweitens kann es bedeuten, dass sie selbst TinkerPlots nicht (mehr?) gut genug bedienen kann, um einen Randomisierungstest damit durchzuführen und dass sie hierbei erwartet, dass Fabia die Simulation durchführt. Drittens kann abgesehen von der technischen Bedienung der zweiten Deutungsalternative auch die generelle Durchführung eines Randomisierungstests gemeint sein, der nicht nur aus der technischen Umsetzung der Simulation besteht. Fabia scheint in ihrem Antwortturn auf die zweite Deutungsalternative zu antworten mit „Ich kann es ja mal versuchen, sonst können wir ja fragen.“ (96). Sie spricht sich selbst somit zu, dass sie mit TinkerPlots umgehen kann und erwartet ansonsten Hilfe durch die Interviewerin, die ihr von dieser im darauffolgenden Turn (97) zugesagt wird. Daraufhin wendet sich Fabia in (98) der Planung der Zufallsmaschine zu und führt dies in (101) näher aus: „… die ganzen Werte und dann müssen wir die in der Zufallsmaschine in zwei verschiedene Stapel einteilen…“. Dabei spricht sie auf Softwareebene und steigt direkt in die Erstellung der Zufallsmaschine ein.

Episode 8-2: Erstellung der Zufallsmaschine (Z. 102-118)

102	*Die Zufallsmaschine wird geöffnet.*

103 L: Die hier, ne?

104 *Die Studentin markiert die Spalte „Blutdrucksenkung“, um diese in die Zufallsmaschine zu ziehen.*



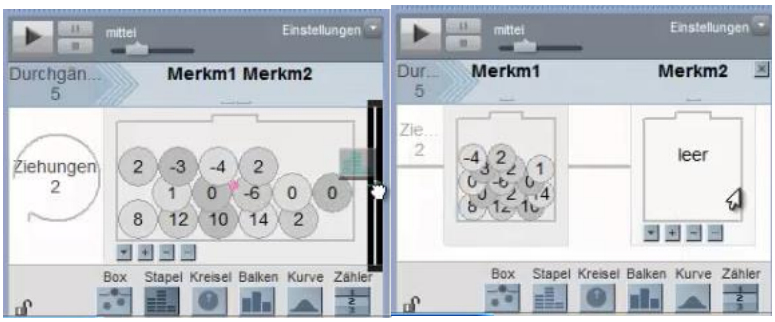
105 F: Wir haben die (*zeigt auf die Kugeln in der Zufallsmaschine*) glaube ich erst alle gelöscht. (...) So.

106 *Dann entfernt sie die Kugeln a, b, c aus der Zufallsmaschine. Es wird versucht, das Merkmal Blutdrucksenkung per Drag&Drop in die Zufallsmaschine zu ziehen, aber das funktioniert nicht.*

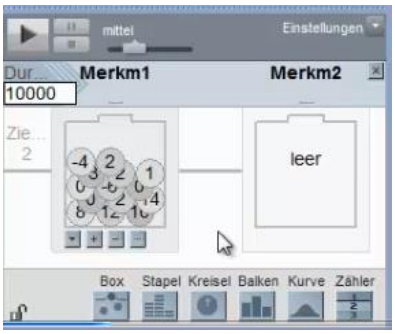
107 I: Jetzt kopieren. Reinziehen funktioniert nicht. Also einmal Steuerung C drücken. (..) Und dann darein klicken und Steuerung V.

108 *Die Studentin kopiert die Werte in die Zufallsmaschine.*

109 F: Ah ok. (..) Und dann hatten wir doch/.
110 L: /Den (*zeigt auf das Bauteil Stapel*) musst du da rein ziehen. (..) Nach rechts.
111 F: Ah ok.
112 *Anschließend markiert sie den Button „Stapel“ und zieht ihn rechts neben die Zufallsmaschine, so dass ein zweites Bauteil erscheint.*



113 L: (unv.) den Stapel da. Wie war das denn nochmal?
114 F: Müssen da die (*zeigt auf die Spalte Diät der Tabelle*) jetzt auch nochmal rein?
115 L: In verschiedene Stapel. Der hat das jetzt doch da zugefügt. Wie oft wollen wir das durchführen?
116 F: Tausend. Zehntausend.
117 *Die Studentin verändert die Anzahl der Wiederholungen in zehntausend.*



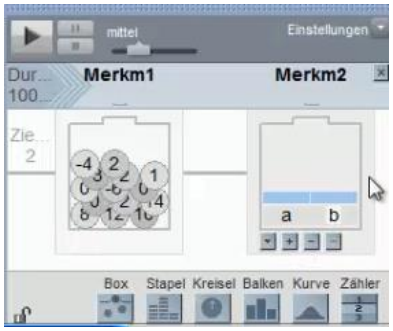
118 L: Das Gesetz der großen Zahlen muss man da beachten.

Die ersten Einstellungen an der Zufallsmaschine werden sehr flüssig durchgeführt. Eine Zufallsmaschine wird von Laura erstellt (102) und die drei vorhandenen Kugeln aus der Box gelöscht (105). Anschließend möchte sie die Blutdruckwerte des ursprünglichen

Experiments in die nun leere Box ziehen (106) und bekommt hierzu den Hinweis von der Interviewerin, dass dies mit Kopieren und Einfügen gemacht werden muss (107). In (110) wirft Laura ein, dass das Bauteil Stapel nun verwendet werden muss und zeigt damit, dass sie ebenfalls weiß, wie nun weitergemacht werden muss. Der Stapel wird daraufhin als zweites Bauteil in Reihe hinter der Box erstellt (112) und hier stockt nun das Vorgehen, weil sich beide wohl nicht mehr sicher sind, wie dieses Bauteil nun gefüllt werden muss. Laura formuliert das als „den Stapel da. Wie war das denn nochmal?“ (113), woraufhin Fabia fragend antwortet: „Müssen da die (*zeigt auf die Spalte Diät der Tabelle*) jetzt auch nochmal rein?“ (114), was als Vorschlag interpretiert werden kann, mit gleichem Vorgehen wie zuvor die Werte des Merkmals Diät des ursprünglichen Experiments in das Bauteil zu kopieren, ein Vorgehen, das komplett richtig wäre, jedoch nicht ausgeführt wird. Laura antwortet sogar „in verschiedene Stapel…“ (115), allerdings nimmt sie direkt die nächste Einstellung in den Blick und fragt nach der Anzahl, mit der sie „das“ durchführen wollen (115). Fabia antwortet mit „Tausend. Zehntausend“, was daraufhin deutet, dass sie diese beiden Zahlen als Wiederholungsanzahl im Kopf behalten hat. Vielleicht ist die zweite Zahl Zehntausend auch als eigener Verbesserungsvorschlag der erstgenannten Zahl gemeint. In (117) wird die Anzahl der Durchgänge auf 10000 gestellt mit dem Turn von Laura, dass man das Gesetz der großen Zahlen beachten müsse (118). Die Erstellung des zweiten Bauteils, in welchem die neuen Gruppen abgebildet werden sollen, stellt sich hier zunächst als Schwierigkeit heraus und wird in der folgenden Episode näher besprochen. Hier scheint ein semantischer Konflikt auf technischer Ebene vorzuliegen bezüglich der Anzahl der Ziehungen und der Anzahl der Durchgänge, die hier fehlinterpretiert werden. Durch die Art, die Simulation aufzubauen (sequentielle Simulation, vgl. Kapitel 3.5.2.2) muss die Wiederholung der Simulation über Messgrößen realisiert werden und kann nicht über die Anzahl der Durchgänge gesteuert werden, die in diesem Fall die Stichprobengröße des Experiments repräsentiert.

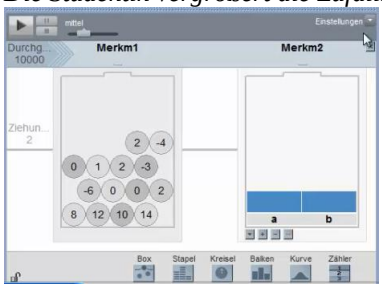
Episode 8-3: Modellierung zwei neuer Gruppen in der Zufallsmaschine (Z. 119-140)

119	F: Ähm. Wie macht man das, dass man in zwei Gruppen einteilt (*blickt zur Interviewerin*)?
120	I: Erkläre mir einmal was du vorhast.
121	F: Der soll das jetzt ja zufällig in zwei verschiedene Gruppen einteilen, die Werte (*zeigt auf das erste Bauteil der Zufallsmaschine*).
122	I: Das heißt du brauchst in der zweiten Urne sozusagen zwei verschiedene Gruppen.
123	F: Genau.
124	I: Das kannst du zum Beispiel mit einem Stapel machen. Mach mal zwei Mal auf plus. Das Bauteil Stapel hast du gerade schon reingezogen, oder?
125	F: Ja. Ah ja.
126	*Die Studentin klickt zweimal auf „+“ und es erscheinen zwei neue Felder a und b im zweiten Bauteil.*



127 I: Und jetzt kannst du noch die Anzahl ändern. Du brauchst ja siebenmal das eine und siebenmal das andere.
128 F: Ja genau.
129 I: Das geht über den kleinen Pfeil. Da gibt es "Anzahl anzeigen".
130 *Die Studentin klickt den kleinen Pfeil an und das Bauteil verschwindet.*

131 F: Was hab ich jetzt/ hab ich es jetzt weggemacht?
132 I: Äh mach mal rückgängig. Bearbeiten, rückgängig. (...) Halt zieh erst mal die Zufallsmaschine eine bisschen größer. Dann geht das glaube ich etwas besser.
133 *Die Studentin vergrößert die Zufallsmaschine.*



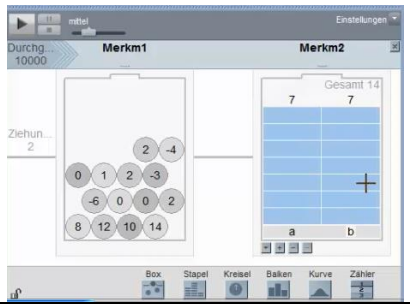
134 F: So. (...)
135 *Das Pfeilmenü wird geöffnet.*

136 F: Anzahlen anzeigen?
137 I: Mhm (bejahend).

138 *Die Studentin klickt "Anzahl anzeigen" an.*

139 F: Und dann können wir hier sieben angeben.

140 *Die Studentin verändert die Anzahlen in der Zufallsmaschine.*

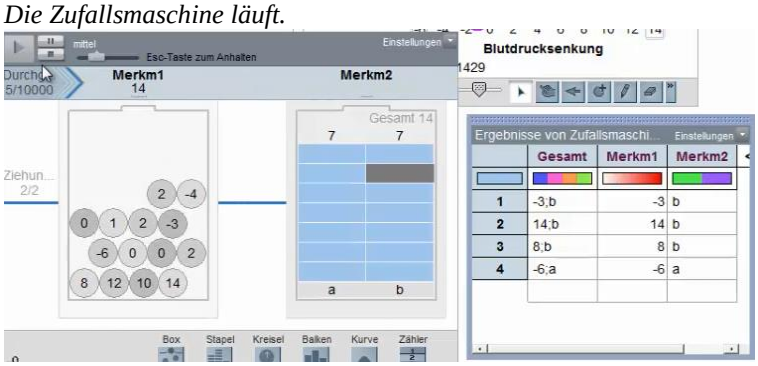


Nachdem die beiden gemerkt haben, dass sie nicht wissen, wie sie zwei neue Gruppen im zweiten Bauteil modellieren sollen, wendet sich Fabia mit diesem Problem an die Interviewerin (119). Die einzelnen Schritte, die nötig sind, werden im Folgenden von der Interviewerin technisch vorgegeben und von Fabia ausgeführt. So werden zwei Stapel innerhalb des Bauteils erstellt (126), die automatisch mit a und b beschriftet sind. Nach Vorgabe werden jedem Stapel sieben Elemente hinzugefügt (127-140). In dieser Episode ist viel technische Hilfe durch die Interviewerin nötig. Abgesehen vom Beginn der Episode, in der auf Statistikebene kurz erklärt wird, dass in zwei Gruppen eingeteilt werden soll, findet das Gespräch auf Softwareebene statt, wobei Bezüge zum Kontext hergestellt werden.

Episode 8-4: Die Einstellung ohne Zurücklegen (Z. 141-155)

141 L: Und jetzt wird das automatisch gezogen, ne.

142 *Die Zufallsmaschine läuft.*



143 I: Jetzt war beim Simulieren immer eine Sache, man muss sich immer überlegen, zieht man mit

oder ohne Zurücklegen.

144 F: Ja.

145 I: Müsst ihr hier mit oder ohne zurückziehen?

146 F: Ah. Mist. Der muss das wegmachen dann. Mach mal Stopp. (*Die Zufallsmaschine wird gestoppt*) (..) Ähm es muss ja dann eins weg sein, wenn er in Gruppen eingeteilt ist, oder? Also ohne Zurücklegen.

147 L: Ja ohne Zurücklegen, weil jede Person ist ja nur einmal dran.

148 F: Haben wir das hier eingestellt?

149 *Die Studentin stellt „ohne Zurücklegen“ ein.*

150 F: Ohne Zurücklegen ok.

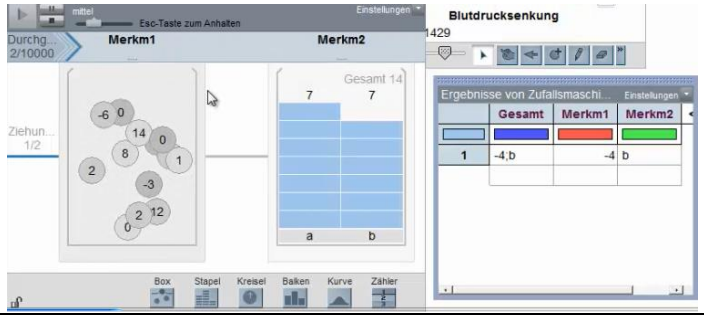
151 L: Ok dann kann ich die (*zeigt auf die bisherige Ergebnistabelle*) ja wieder komplett löschen.

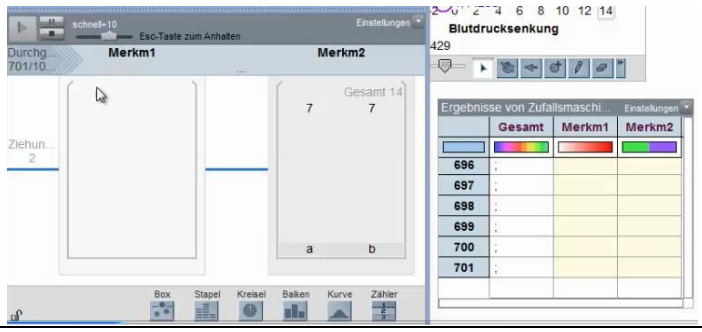
152 F: Hast du vorne auch ohne Zurücklegen eingestellt?

153 L: Bestimmt nicht.

154 *Auch für das erste Bauteil wird ohne Zurücklegen eingestellt.*

155 *Die Studentin lässt die Zufallsmaschine laufen.*

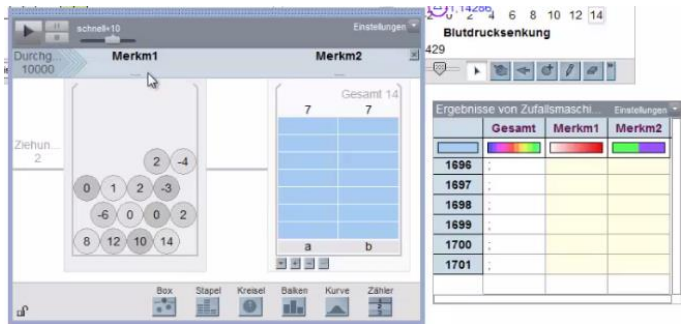




Nachdem die Bauteile korrekt erstellt sind, gibt die Interviewerin den Hinweis, dass über die Einstellung mit oder ohne Zurücklegen nachgedacht werden solle (143). Währenddessen läuft die Zufallsmaschine im Hintergrund auf mittlerer Geschwindigkeit. Dies scheint Fabia im Blick zu haben, denn auf die Frage der Interviewerin, ob mit oder ohne Zurücklegen eingestellt werden müsse (145) antwortet sie: „Ah. Mist. Der muss das wegmachen dann. Mach mal Stopp. (*Die Zufallsmaschine wird gestoppt*). Ähm es muss ja dann eins weg sein, wenn er in Gruppen eingeteilt ist, oder? Also ohne Zurücklegen.“ (146). Bei den bisherigen Einstellungen der Zufallsmaschine (n = 10000 Durchgänge, beide Bauteile standardmäßig mit Zurücklegen) wird immer zufällig ein Element aus jedem Bauteil ausgewählt. Mit „Der muss das wegmachen dann…“ bezieht sich Fabia wohl auf ihre Erwartung, dass nämlich bei mittlerer Geschwindigkeit der Zufallsmaschine erkennbar sein muss, dass in der vorliegenden Situation nach gezogenem Element eines Bauteils dieses eliminiert wird, was bei der aktuellen Einstellung der Zufallsmaschine nicht der Fall ist. Deshalb veranlasst Fabia das zu der Aussage „Mach mal Stopp“ und daraufhin erfolgt die Erklärung „Ähm es muss ja dann eins weg sein, wenn er in Gruppen eingeteilt ist, oder?“, und sie gibt am Ende ihres Turns die Antwort auf die Frage der Interviewerin: „Also ohne Zurücklegen“. Auch Laura hat dies verstanden und fügt als Erklärung aus dem Kontext hinzu: „Ja ohne Zurücklegen, weil jede Person ist ja nur einmal dran.“ (147). Sie hat also vermutlich die Vorstellung im Kopf, dass hinter den Blutdruckwerten die Personen stehen, die nun auf die zwei neuen Gruppen aufgeteilt werden. Im Folgenden beschäftigen sich die beiden damit, in beiden Bauteilen die Einstellung ohne Zurücklegen vorzunehmen (148-154) und die Zufallsmaschine erneut laufen zu lassen (155). Der direkte Bezug zum Kontext hilft hier, die korrekten Einstellungen vorzunehmen.

Episode 8-5: Weiteres Planen (Z. 156-166)

156 F: Jetzt sortiert er ja die Werte, die wir haben den verschiedenen Gruppen zu. (...) Achso ja. Er muss das /nee/ er macht das ja jetzt richtig oft.
157 L: Eigentlich müssten wir dann erst die Gruppen auswerten, oder? Und dann muss er das nochmal machen.
158 F: Also wir müssen ja jetzt erst von der ermittelten Gruppe/ ich mache mal Stopp/.
159 *Die Zufallsmaschine wird gestoppt.*



160 L: Genau von einer Gruppe und dann von der Nächsten und dann von/ dann immer Differenzen.
161 F: Und dann müssen wir ja erst die Mee. Mittelwerte wollten wir machen oder?
162 L: Ja.
163 F: Arithmetisches Mittel.
164 L: Differenzen der arithmetischen Mittel. Müssen wir dann erst ermitteln. Und dann in diese Verteilung einzeichnen lassen.
165 F: Vielleicht hätten wir das doch mit der Hand machen sollen.
166 Alle: (lachen).

Nachdem die Zufallsmaschine läuft, erschließt sich Fabia den Sinn in (156): „Jetzt sortiert er ja die Werte, die wir haben, den verschiedenen Gruppen zu…“, wobei ihr etwas auffällt, denn sie stockt und führt fort „Achso ja. Er muss das/nee/er macht das ja jetzt richtig oft.“ Auch Laura fällt auf, dass etwas nicht stimmt, denn sie leitet ihren nächsten Turn mit „Eigentlich…“ ein und schlägt vor „Eigentlich müssten wir dann erst die Gruppen auswerten, oder? Und dann muss er das nochmal machen.“ (157). Hier tritt jetzt als Schwierigkeit hervor, dass die beiden 10000 Durchgänge eingestellt haben, durch die Einstellung ohne Zurücklegen sind die Bauteile nach 14 Durchgängen jedoch leer. Dies können die beiden an dieser Stelle zwar nicht benennen, aber ihnen fällt auf, dass die Simulation nicht stimmt, denn auch Fabia sagt in ihrem nächsten Turn „Also wir müssen ja jetzt erst von der ermittelten Gruppe/ich mache mal Stopp“ (158) und stoppt die Zufallsmaschine (159). Was sie hier mit Gruppe meint, ist nicht eindeutig, anscheinend meint sie die Gesamtgruppe an Versuchspersonen, nicht eine der beiden neuen Gruppen. Ähnlich stellt es sich auch in Lauras nächstem Turn dar, die vielleicht auch einfach nur die Wortwahl von Fabia übernimmt: „Genau von einer Gruppe und dann von der Nächsten und dann von/ dann immer Differenzen.“ (160). Eventuell ist hier auch gemeint, dass sie die einzelnen Gruppen hinsichtlich der Mittelwerte auswerten möchten, was Fabia im nächsten Turn erwähnt (161) und Laura durch die „Differenzen“ andeutet, denn es können nur die Differenzen zwischen den Mittelwerten gemeint sein. Nach einer Klärung, dass sie arithmetische Mittel betrachten wollten, präzisiert Laura: „Differenzen der arithmetischen Mittel. Müssen wir dann erst ermitteln. Und dann in diese Verteilung einzeichnen lassen.“ (164). Implizit definiert sie hier bereits die Testgröße als Differenzen der arithmetischen Mittel und macht sogar einen Vorgriff auf die Referenzverteilung, in der diese eingezeichnet werden. Dies lässt darauf schließen, dass sie das weitere Vorgehen im Blick hat. Ob dies für Fabia ebenfalls zutrifft, ist fraglich, denn sie wirft auf diese Planung von Laura ein, dass sie vielleicht doch lieber

von Hand hätten simulieren sollen (165). Das kann gedeutet werden als Versuch, sich das gesamte Vorgehen über den vermeintlich langsameren Prozess des händischen Randomisierens verdeutlichen zu wollen oder auch als Aussage darüber, dass sie den Gedankengängen von Laura nicht komplett folgen kann oder will. Das Lachen aller im abschließenden Turn (166) dieser Episode zeigt, dass alle diesen Vorschlag nicht ernst nehmen.

Episode 8-6: Durchgangsanzahl neu bestimmen (Z. 167-172)

167 L: Ich mach das erst mal nochmal weg. (*Die Ergebnistabelle wird gelöscht*) (...) Aber wenn ich jetzt hier nur einmal mache (*Die Durchgangsanzahl wird auf 1 geändert – anschließend wird versucht, auf Ziehungen zu klicken*). (...) oder muss man./ nee Ziehungen kann man nicht ändern und dann (unv.).

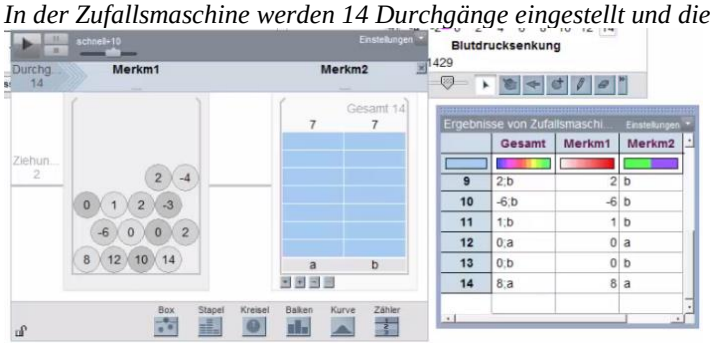
168 *Die Zufallsmaschine wird erneut gestartet. Ein Durchgang wird durchgeführt.*

169 F: Ah. Jetzt macht er. Jetzt nimmt der nur einen. (..) Das ist ja auch Quatsch.

170 L: Ich glaube du musst doch die Ziehung ändern, oder?

171 F: Nee Ziehung kann man nicht ändern, weil wir ja zwei von diesen Dingern haben. (…) Oder muss man das jetzt/ und wenn wir vierzehn machen?

172 *In der Zufallsmaschine werden 14 Durchgänge eingestellt und die Zufallsmaschine gestartet.*



Nachdem erkannt wurde, dass die gewählten Einstellungen noch nicht komplett richtig zu sein scheinen, begeben sich die Beiden in dieser Episode auf Fehlersuche auf Softwareebene. Laura ändert zunächst die Durchgangsanzahl auf 1 (167), ohne weitere Begründung, und versucht ebenfalls die Ziehungsanzahl zu ändern, die jedoch auf zwei festgelegt ist und nicht geändert werden kann, sobald mehrere Bauteile in einer Zufallsmaschine verwendet werden. Die Zufallsmaschine wird mit einem Durchgang laufen gelassen (168), was von Fabia mit „… Das ist ja auch Quatsch.“ (169) kommentiert wird. Laura wirft daraufhin noch einmal den Aspekt der Ziehung ein, der zu ändern sei (170), Fabia erklärt jedoch, dass die Ziehungsanzahl aufgrund der zwei Bauteile nicht

geändert werden könne (171). Die Funktionsweise der Zufallsmaschine ist ihr an dieser Stelle also klar. Sie macht daraufhin den Vorschlag, vierzehn zu nehmen (171), was nicht weiter kommentiert, sondern direkt umgesetzt wird, und die Zufallsmaschine wird gestartet (172). Eine Argumentation auf Kontext- oder Statistikebene mit Bezug zur ursprünglichen Gruppengröße hätte hier hilfreich sein können, findet jedoch nicht statt.

Interaktionseinheit 9: Teststatistik definieren und Referenzverteilung erstellen (Z. 173-272)

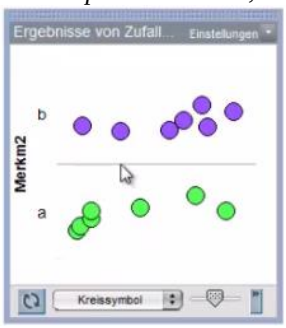
In der neunten Interaktionseinheit wird in insgesamt 10 Episoden die Teststatistik definiert und die Referenzverteilung erstellt. In der ersten Episode werden die simulierten Ergebnisse in einem Graphen dargestellt. In der zweiten Episode wird die Teststatistik in TinkerPlots bestimmt und anschließend in der dritten Episode das weitere Vorgehen geplant und die Messgrößen festgelegt. In der vierten Episode wird die Animation ausgeschaltet, um den Messgrößensammelvorgang zu beschleunigen. Die fünfte Episode behandelt die Anzahl der zu sammelnden Messgrößen. Während die Messgrößen gesammelt werden, wird in der sechsten Episode das Schema ausgefüllt. Es folgt ein kurzer Exkurs über die Lerneinheit zur Inferenzstatistik in der siebten Episode und in der achten, kurzen Episode wird das Schema weiter ausgefüllt. In der neunten Episode wird erneut die Anzahl der Messgrößen diskutiert und in der zehnten Episode wird schließlich die Referenzverteilung erstellt.

Episode 9-1: Darstellen der simulierten Ergebnisse (Z. 173-193)

173 F: Jetzt hat man ja zumindest schon mal einmal randomisiert. (*In der Ergebnistabelle wird hoch und runtergescrollt*). Das ist schon so lange her mit TinkerPlots. (...) Und jetzt können wir das ja theoretisch in die zwei Gruppen einteilen. (...)

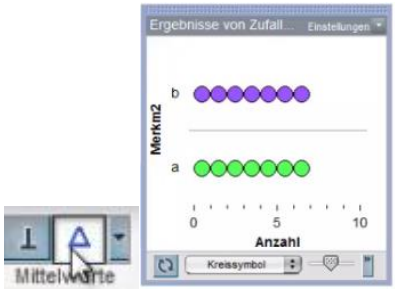
174 L: Da eine Graphik jetzt erst mal machen und da konnte man ja äh (...) irgendwo auf Ziehung nochmal machen.

175 *Ein Graph wird erstellt, die Fälle werden nach Merkm2 eingefärbt und werden getrennt.*

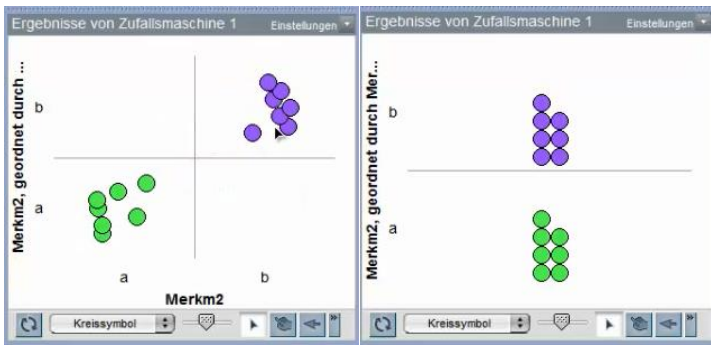


176 F: a und b. (...) Jetzt brauchen wir einen Mittelwert, ne? (…)

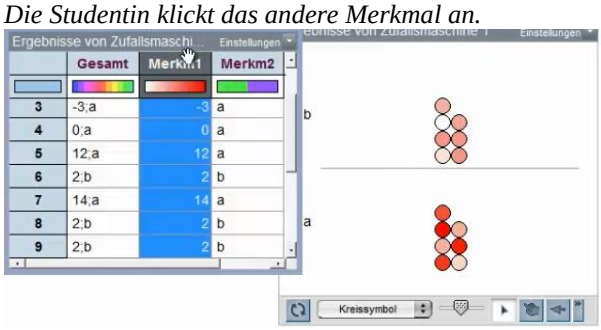
177 *Mit der Maus wird über die Funktionen für den Graph gefahren. Es wird horizontal gestapelt. Die Studentin wählt den Button „Mittelwerte“ aus. Dieser wird aber nicht angezeigt, weil die Darstellung nach den Gruppen geordnet ist.*



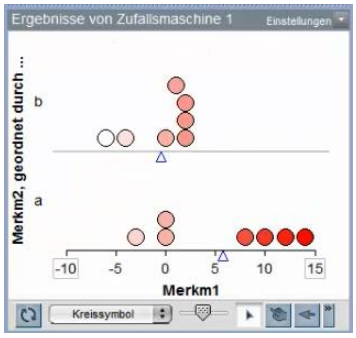
178 *Die Fälle werden einige Male angeklickt und hin und hergezogen. Es herrscht Schweigen.*
179 I: Was möchtest du machen?
180 F: Die Seiten einmal mit einer Linie (unv.)
181 *Die Studentin ordnet und stapelt die Ergebnisse der Zufallsmaschine hektisch auf verschiedene Arten.*



182 L: Was für eine Linie?
183 F: Warte. Das man so wie da. (...) Stapeln. (...) Bin ich zu blöd?
184 L: Sag noch mal/.
185 F: Es soll genauso dargestellt werden wie da (*zeigt auf die Graphik zu den beobachteten Daten*). Jetzt hab ich diese (*zeigt auf den Graph mit den Ergebnissen der Simulation*) in die zwei Gruppen. Sind sie ja jetzt schon.
186 I: Genau. Jetzt musst du noch auf das andere Merkmal erst klicken.
187 F: AH JA. Das fehlt.
188 I: Sonst weiß er nicht, was du sortieren möchtest.
189 *Die Studentin klickt das andere Merkmal an.*



190 F: Ah jetzt.
191 I: In die andere Richtung nach rechts schieben. Das willst du wahrscheinlich.
192 F: Ja.
193 *Anschließend werden die Werte gestapelt. Der Mittelwert wird automatisch eingeblendet, weil der Button bereits aktiviert war.*



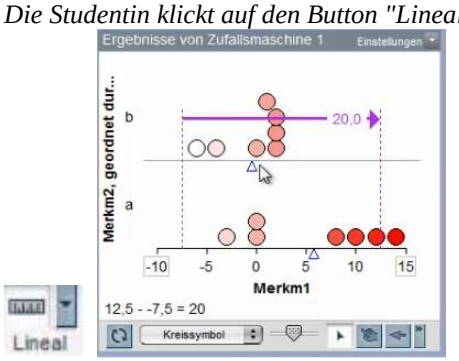
Zu Beginn dieser Episode und der Einheit stellt Fabia zunächst einmal den Status quo fest: „Jetzt hat man ja zumindest schon mal einmal randomisiert…“ (173). Ihr scheint also bewusst zu sein, dass die vorhandenen Blutdruckreduktionen nun einmal zufällig auf zwei neue Gruppen aufgeteilt wurden. Mit „… Und jetzt können wir das ja theoretisch in die zwei Gruppen einteilen“ meint sie wohl die graphische Darstellung der simulierten Ergebnisse auf Softwareebene, lässt allerdings auch den Bezug zur Statistikebene erkennen. Die folgende Aussage (174) von Laura zeigt, dass sie den gedanklichen Übergang zur Darstellung der simulierten Ergebnisse in einer Graphik leistet. Der zweite Teil ihres Turns bleibt interpretativ im Dunkeln, denn auf Ziehungen klicken war bereits vorher nicht möglich und ist auch jetzt nicht möglich. Auch nach der Turn-by-Turn Analyse ist nicht rekonstruierbar, was sie an dieser Stelle vorschlägt. Als nächstes wird ein Graph erstellt und bezüglich des Merkmals *Merkm2* eingefärbt (175). Daraufhin sagt Fabia „a und b. (…) Jetzt brauchen wir einen Mittelwert, ne?“ (176). Sie rekapituliert also zunächst, dass a und b die Ausprägungen des Merkmals sind, nach denen horizontal getrennt wurde. Es wäre hier die Gelegenheit gewesen, zu versprachlichen, dass a und b die beiden neuen Gruppen bezeichnen, aber dies wird nicht wahrgenommen. Dies kann entweder unterlassen worden sein, weil dieser Sachverhalt klar ist oder eben weil er unklar ist. Das weitere Vorgehen der Beiden legt jedoch eher nahe, dass ihnen die Bedeutung der Ausprägungen des Merkmals klar ist. Nach Fabias Forderung nach einem Mittelwert (176) wird dieser in der Werkzeugauswahl angeklickt, jedoch nicht in der Graphik angezeigt, weil nur ein kategoriales Merkmal dargestellt ist, von welchem kein Mittelwert angezeigt werden kann. Daraufhin geschieht erst einmal nichts, was als Ratlosigkeit gedeutet wird. So hat es auch die Interviewerin interpretiert, denn sie fragt als nächstes: „Was möchtest du machen?“ (179). Die nächste Aussage „Die Seiten einmal mit einer Linie“ (180) von Fabia ist wohl der Versuch anzudeuten, dass sie erwartet, dass weitere „Linien“ erscheinen, wenn sie Fälle nach rechts zieht, wie dies bei einem numerischen Merkmal der Fall wäre, weil dort durch das Ziehen eines Falls nach rechts eine Klasseneinteilung stattfinden würde. Zunächst werden die Fälle auf verschiedene Weisen angeordnet, jedoch immer mit der Einfärbung nach Merkmal 2, also den beiden neuen Gruppen (181-184). In (185) formuliert Fabia, dass sie die Ergebnisse so darstellen möchte, wie in der Graphik zum ursprünglichen Experiment, woraufhin sie

den Hinweis von der Interviewerin bekommt, dass sie dazu erst das andere Merkmal anklicken müsse (186). Dies wird lautstark von Fabia mit „AH JA. Das fehlt“ (187) kommentiert, was zeigt, dass ihr dieser elementare Schritt gefehlt hat und sie nun denkt, dass sie weiß, wie sie weiterarbeiten kann. Mit ein wenig weiterer Hilfe durch die Interviewerin stellt Fabia dann die Daten ähnlich strukturiert dar wie in der Darstellung der Daten des ursprünglichen Experiments. Während dieser Episode spielt sich fast die gesamte Kommunikation auf Softwareebene ab.

Episode 9-2: Bestimmen der Teststatistik (Z. 194-197)

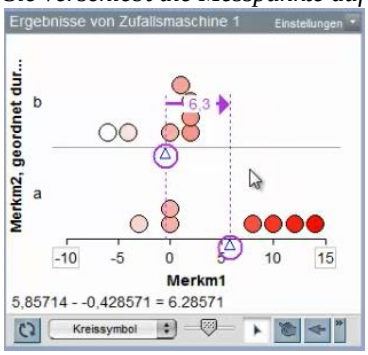
194 F: Und dann haben wir die Mittelwerte hier. (...) Und dann konnten wir mit diesem Lineal war das glaube ich. Dass man die Differenz der Mittelwerte berechnet hat.

195 *Die Studentin klickt auf den Button "Lineal".*



196 F: Da haben wir das ja glaube ich da hoch geschoben. (...) Und das dahin.

197 *Sie verschiebt die Messpunkte auf die Mittelwerte.*



Die Mittelwerte werden automatisch angezeigt, weil das Werkzeug noch aktiviert war. Dies erkennt Fabia in (194) und plant nun mit Hilfe des Lineals die Differenz der Mittelwerte zu berechnen. Dazu verwendet sie es kompetent (195-197) und bekommt die gewünschte Differenz korrekt angezeigt. Das Gespräch findet auf Statistik- und Softwareebene statt.

Episode 9-3: Planen und Messgrößen definieren (Z. 198-210)

198 L: Das müssten wir ja ganz oft machen theoretisch. Und die Differenz müssten wir ja noch in wieder eine neue Tabelle einfügen. Oder?

199 F: Geh mal auf Einstellung oben rechts bei dem.

200 I: Warte kurz. Lass mich kurz einmal eingreifen. Was hast du jetzt vor?

201 F: Also wir haben jetzt ja quasi die Differenz der Mittelwerte von dieser zufälligen Gruppe ge-

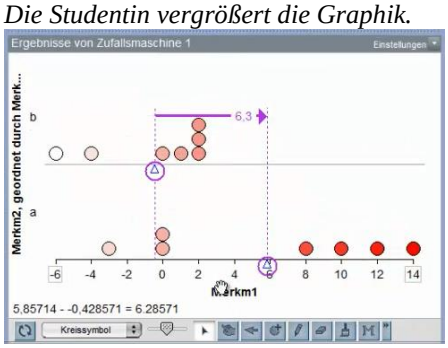
macht.

202 L: Von einer.

203 F: Von einer. Und eigentlich müssten wir das ja ganz oft/ ach wir müssten das jetzt immer wieder durchführen. Und damit wir ganz viele Differenzen dieser Mittelwerte haben. Und das kann man ja dann ganz schön in einer Referenzverteilung darstellen.

204 I: Gut das/ in TinkerPlots hieß das Messgrößen sammeln. Diese Differenz. Wenn du unten/ Ist da ein "M" zu sehen in der Graphik? Nee. Zieh die mal ein bisschen größer bitte. Ein Stück nach links.

205 *Die Studentin vergrößert die Graphik.*



206 I: Jetzt ist da ein "M" richtig? Ganz rechts unten.

207 F: Da. Ja.

208 *Die Studentin klickt auf das "M“.*

209 I: Ja. Und jetzt kannst du die Differenz die da oben angezeigt ist einfach anklicken und dann wird die in eine extra Tabelle gesammelt. Das wolltest du machen?

210 F: Ja. Genau.

Nachdem die Teststatistik definiert ist, plant Laura das weitere Vorgehen auf Softwareebene: „Das müssten wir ja ganz oft machen theoretisch. Und die Differenz müssten wir ja noch in wieder eine neue Tabelle einfügen. Oder?“ (198). Mit „Das müssten wir ja ganz oft machen theoretisch…“ ist wohl gemeint, dass die Blutdruckwerte zufällig neu auf zwei Gruppen aufgeteilt, die Mittelwerte gebildet und die Differenz der Mittelwerte berechnet wird. Dies wird zwar nicht versprachlicht, der Prozess scheint jedoch klar zu sein. Mit dem Einfügen der Differenzen in eine neue Tabelle hat sie wohl die Tabelle vor Augen, in der von TinkerPlots automatisch gesammelte Messgrößen protokolliert werden, auch wenn sie von dem korrekten Sprachangebot keinen Gebrauch macht. Nach einer Rückfrage der Interviewerin (200) erklärt Fabia auf Statistikebene noch einmal, was geschehen ist: „Also wir haben jetzt ja quasi die Differenz der Mittelwerte von dieser zufälligen Gruppe gemacht.“ (201), wobei sie wohl eher „von dieser zufälligen Einteilung in die Gruppen“ als „von dieser zufälligen Gruppe“ meint. Auch Lauras darauffolgender Einwurf „Von einer.“ (202) lässt sich nur in diese Richtung deuten, denn es wurde ja nicht in eine zufällige Gruppe eingeteilt, sondern in zwei. Mit Ausblick auf die Referenzverteilung wiederholt Fabia noch einmal das weitere Vorgehen (203) und bekommt daraufhin von der Interviewerin den Hinweis, dass dazu Messgrößen zu verwenden seien (204). Mit Anweisungen durch die Interviewerin wird anschlie-

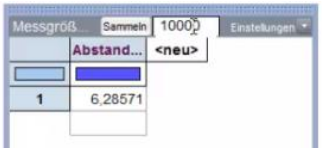
ßend die mit Hilfe des Lineals gemessene Differenz der Mittelwerte als Messgröße definiert (205-210).

Episode 9-4: Ausschalten der Animation (Z. 211-225)

211 L: Und da auf Sammeln, aber dann musst du erst die Animation ausstellen irgendwo.
212 *Das Fenster Info Zufallsmaschine wird geöffnet.*

213 I: Ganz oben unter Messgrößen. Genau das sind jetzt technische Sachen, um das zu beschleunigen. Einstellungen zu Messgrößen das oberste. Und jetzt hier der unterste Punkt, wo „Animation an“ steht. Das einmal aus.
214 F: Ah ok.
215 *Das Häkchen bei Animation an wird entfernt.*

216 L: Und dann kannst du da jetzt 10000 eingeben.
217 *Bei der Anzahl der zu sammelnden Messgrößen wird 10000 eingegeben.*



218 *Unbemerkt von den Studentinnen wird der Wert von TinkerPlots automatisch auf 5000 gesetzt.*
219 I: Ja warte noch nicht draufklicken. Das ist richtig. Mach nur die Sachen alle noch ganz klein. Also einmal anklicken und so ganz klein und noch kleiner noch kleiner noch kleiner. Nee das musst du von unten nach oben machen. (...) Genau.
220 *Die Zufallsmaschine wird minimiert.*

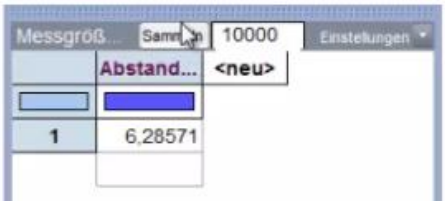
221 F: So und das auch? (*zeigt auf die Graphik*)
222 I: Ja. Das auch. Richtig. Ja weil der muss sonst die ganze Zeit irgendwas anzeigen.
223 *Alle weiteren Objekte werden minimiert.*
224 I: Super. Das wisst ihr noch richtig gut. Und die Tabelle auch noch, die dazu gehört.
225 F: Die hier.

Laura fällt nun offensichtlich ein, dass auf Softwareebene zum Messgrößensammeln gelernt wurde, die Animation auszustellen. Dies artikuliert sie in (211). Mit einem kur-

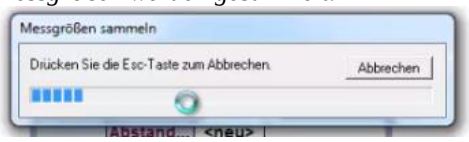
zen Hinweis der Interviewerin wird die Animation ausgeschaltet (212-215). Daraufhin sagt Laura, dass nun 10000 eingegeben werden könne (216), was ausgeführt wird (217). TinkerPlots begrenzt die zu sammelnde Anzahl an Messgrößen automatisch auf 5000, was von den Studierenden unbemerkt bleibt. Im Weiteren weist die Interviewerin noch daraufhin, die zur Simulation gehörigen Objekte zu minimieren, was von den Studierenden direkt ausgeführt wird (219-225). Diese Schritte beherrschen die Studierenden entgegen der in der a priori Analyse formulierten Erwartung sehr gut.

Episode 9-5: Anzahl Messgrößen (Z. 226-236)

226 I: Dann kannst du sammeln.
227 F: Wie viele haben wir jetzt?
228 L: Ist ja eigentlich egal, oder? Komm wir machen mal 10000.
229 *Die Studentin stellt erneut 10000 ein.*



230 *Wieder begrenzt TinkerPlots die Anzahl auf 5000.*
231 I: Wie viel hast du jetzt eingestellt?
232 F: 10000.
233 I: Mach mal. Dauert jetzt einen Moment.
234 Alle: (Lachen)
235 *Messgrößen werden gesammelt.*

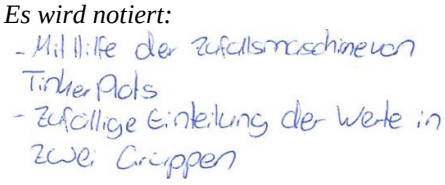


236 L: Ich habe immer lieber eine große Zahl genommen. Dann ist es etwas genauer.

Ohne große Diskussion werden 10000 Messgrößen eingestellt, was von TinkerPlots erneut auf 5000 begrenzt wird. Der Sammelvorgang wird gestartet.

Episode 9-6: Ausfüllen des Schemas (Z. 237-241)

237 F: Und dann müssen wir ja eigentlich nur einen Graph öffnen. Und dann haben wir doch die Referenzverteilung, ne?
238 L: Ja.
239 F: Und dann müssen wir nur noch den Wert, den wir haben/ eigentlich könnten wir das schon mal hier rein schreiben (*zeigt auf den Zettel mit dem Schema*), was wir gemacht haben. (unv.) mit Hilfe der Zufallsmaschine von TinkerPlots. (15 sec Pause) Zufällige Gruppeneinteilung, soll ich das so hinschreiben?
240 L: Zufällige Einteilung der Werte in zwei Gruppen.
241 *Es wird notiert:*



Während die Messgrößen gesammelt werden, plant Fabia kurz das weitere Vorgehen auf Softwareebene im Hinblick auf die Referenzverteilung (237) und bezieht sich nach einer kurzen Bestätigung von Laura (238) im nächsten Turn bereits auf den P-Wert, ohne ihn so zu bezeichnen: „Und dann müssen wir nur noch den Wert, den wir haben…“ (239). Dieser Halbsatz ließe sich vervollständigen als …*den wir [beobachtet] haben, in die Referenzverteilung eintragen*. Sie bricht jedoch mitten im Satz ab und schlägt stattdessen vor, das Schema auszufüllen. Sie notieren unter Randomisierung: „Mit Hilfe der Zufallsmaschine von TinkerPlots – Zufällige Einteilung der Werte in zwei Gruppen“. Hierbei gehen sie weder auf den Kontext des Experiments ein noch liefern sie eine genaue Beschreibung ihres Vorgehens in TinkerPlots. Die von ihnen gewählte Beschreibung wäre austauschbar bei sehr vielen Randomisierungstests anwendbar.

Episode 9-7: Exkurs über Verständnis während des Bausteins Inferenzstatistik (Z. 242-250)

In einem kurzen Exkurs reflektieren Laura und Fabia darüber, dass sie das Vorgehen beim Randomisierungstest im Seminar „erst so richtig verstanden haben“ (242). Und dass für dieses Verständnis das Vorgehen zunächst mit der händischen Simulation über Zettel und dann die Übertragung in TinkerPlots sehr hilfreich war (248).

Episode 9-8: Schema weiter ausfüllen (Z. 251-253)

251 *Im Hintergrund ist das Messgrößensammeln beendet. Die Studentinnen beschäftigen sich weiter mit dem Ausfüllen des Schemas.*
252 L: Differenz der Mittelwerte.
253 *Es wird notiert:*

Ohne Diskussion notieren die Beiden im Schema unter Teststatistik „Differenz der Mittelwerte“ (253).

Episode 9-9: Anzahl Messgrößen (Z. 254-262)

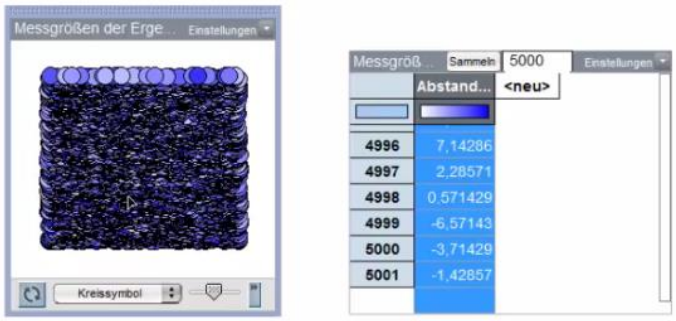
254 F: Ok. Er hat jetzt nur 5000 gemacht, ich glaube 10000 war ein bisschen zu viel.
255 L: Du hast doch 5000 eingegeben oder?
256 F: Nee ich habe 10000 eingegeben, aber er hat 5000 raus gemacht.
257 L: Das hatte er ja eben auch.
258 F: Eben auch. Eben hab ich auch erst 10000 eingegeben und dann kam das da.
259 L: Deswegen hab ich es nochmal geändert.
260 I: Du kannst ja nochmal 5000 sammeln, dann hat er insgesamt 10000, wenn du das gerne möchtest.
261 F: Ach ist egal.
262 L: 5000 ist ja eigentlich auch schon. (...) Ähm jetzt müssen wir die Werte nehmen, da brauchen wir einen Graphen.

Hier fällt Fabia auf Softwareebene auf, dass insgesamt nur 5000 Messgrößen gesammelt wurden (254). Sie stellen gemeinsam fest, dass bereits vorher die von ihnen eingegebe-

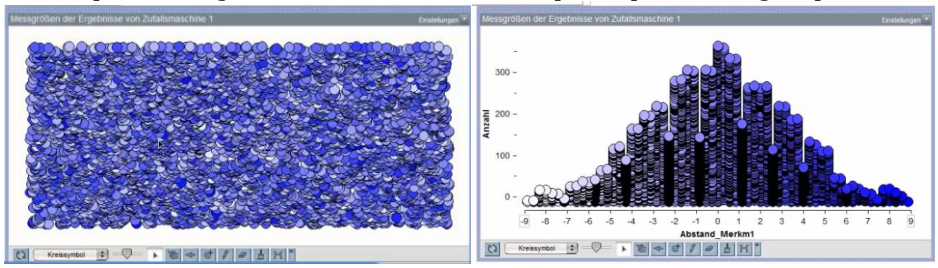
ne Anzahl von 10000 auf 5000 begrenzt wurde (255-259). Auf den Einwurf der Interviewerin, noch einmal 5000 zu sammeln, um auf die gewünschten 10000 zu kommen (260) antwortet Fabia in (261) mit „Ach ist egal.“, als würde sie keinen großen Unterschied darin sehen, ob sie nun 5000 oder 10000 Messgrößen sammeln würden. Auch Laura scheint dem zuzustimmen, denn sie beginnt den folgenden Turn (262) mit „5000 ist ja eigentlich auch schon (…) …“, was ergänzt werden könnte durch „genug“ oder „ausreichend“ oder „genau“. Statistisch wird nicht argumentiert, wieso 5000 als Anzahl der gesammelten Messgrößen für die Referenzverteilung „eigentlich auch schon“ ausreicht. Daraufhin leitet sie zur nächsten Episode über, dass nun ein Graph benötigt würde.

Episode 9-10: Referenzverteilung erstellen (Z. 263-272)

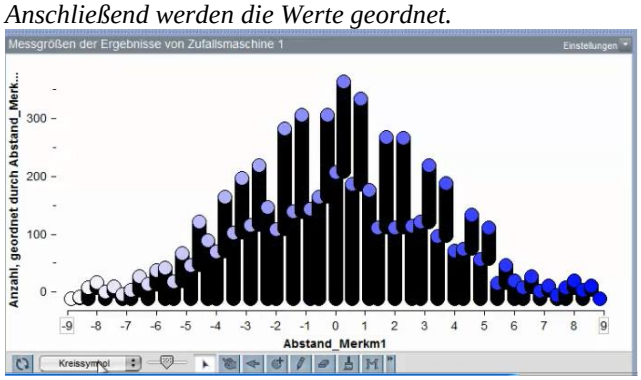
263 *Es wird ein Graph erstellt bei markierter Messgrößentabelle.*



264 F: Mach dir den gleich ein bisschen größer, dann ist es einfacher.
265 L: Ja ich mach das mal weg. Zur Seite.
266 *Der Graph wird vergrößert. Die Werte werden komplett separiert und gestapelt.*



267 F: Sollen wir es auch noch ordnen? Das ist schöner, oder?
268 L: Ja.
269 *Anschließend werden die Werte geordnet.*



270	F: Ach kann man das nicht hier/ (*es wird auf Kreissymbol im Graphobjekt gezeigt*)?
271	I: /Eigentlich kannst du das nicht ordnen, du hast es hier unten schon komplett einmal aufgetrennt auf die Achse. Es ist ja eigentlich schon geordnet.
272	F: Ja.

Mit sehr wenig Kommunikation auf Softwareebene und ohne Hilfe durch die Interviewerin wird die Referenzverteilung flüssig erstellt. Die angezeigte Strategie von Fabia „Mach dir den [Graphen] gleich ein bisschen größer, dann ist es einfacher.“ (264) ist positiv hervorzuheben, da dies einmal mehr ihre hohe Kompetenz im Umgang mit der Software zeigt.

Interaktionseinheit 10: Ermitteln des P-Werts (Z. 273-298)

In der zehnten Interaktionseinheit wird der P-Wert ermittelt. Diese Einheit wird in drei Episoden gegliedert, von denen die ersten beiden sehr kurz sind. In der ersten Episode wird das Vorgehen zum Ermitteln des P-Werts geplant, in der zweiten Episode wird das Schema im Hinblick auf die Referenzverteilung ausgefüllt und dann in der dritten, langen Episode wird der P-Wert in TinkerPlots ermittelt.

Episode 10-1: Planen des Vorgehens (Z. 273-275)

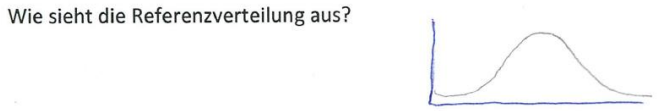
273	L: Ähm und jetzt müssen wir den Wert, den wir haben, hier eintragen (*es wird in die Verteilung gezeigt*). (...) Und dann müssen wir mit diesem Divider gucken wie viel Prozent. Und dann haben wir doch schon den P-Wert, oder?
274	F: Mhm (bejahend).
275	L: Was war nochmal der Wert? 7,7 oder?

Diese Interaktionseinheit beginnt mit einer kurzen Episode des Planens von Laura (273) auf statistischer und auf Softwareebene. Dabei führt sie alles kurz aus, vermittelt jedoch den Eindruck, dass sie einschätzen kann, welches die nächsten nötigen Schritte sind. Zunächst möchte sie „… den Wert, den wir haben, hier eintragen…“, womit nur die im ursprünglichen Experiment beobachtete Differenz von 7,7 mmHg gemeint sein kann, die sie in die gerade erstellte Referenzverteilung, beispielsweise mit Hilfe eine Linie, eintragen möchte. Als nächstes möchte sie mit dem „… Divider gucken, wie viel Prozent…“, was zum einen darauf schließen lässt, dass sie mit der englischen Version von TinkerPlots gearbeitet hat (sie verwendet das Wort Divider statt Einteiler), zum anderen lässt es darauf schließen, dass sie weiß, wie sie den Einteiler einsetzen muss, um herauszubekommen, wie viel Prozent *gleich oder größer* dem beobachteten Wert sind. Die so mit dem Einteiler ermittelte Prozentzahl identifiziert sie dann korrekterweise als P-Wert. Fabia bestätigt dies kurz und somit kann das Vorgehen als akzeptiert betrachtet werden. Lauras nächster Turn (275) zeigt, dass sie vorher in (273) tatsächlich den richtigen Wert gemeint hat.

Episode 10-2: Ausfüllen des Schemas (Z. 276-281)

276	F: Wie kann man das denn jetzt beschreiben (*zeigt auf das Schema*)?
277	L:/ Sagen wir einfach siehe Abbildung, oder?

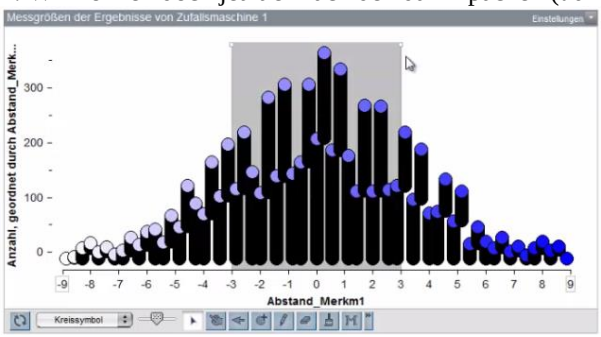
278 I: Genau. Du kannst einfach eine Skizze machen.
279 L: Ja mal doch einfach so eine/.
280 F: Gut dass du schreibst. Du kannst das besser als ich. (lachen)
281 *Es wird skizziert:*



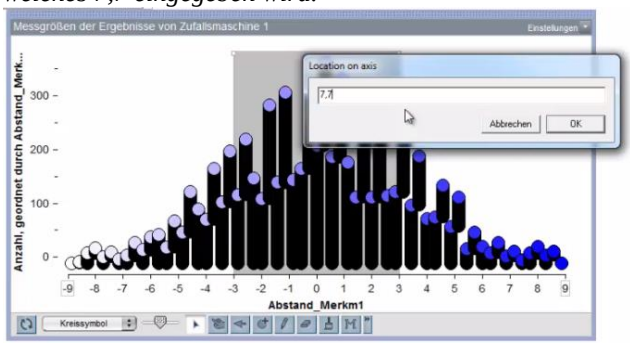
Für die Skizze zur Referenzverteilung im Randomisierungstestschema orientieren sich die Studierenden an der erstellten Graphik und skizzieren ohne Achsenbeschriftung und ohne Skala die Referenzverteilung wie in (281).

Episode 10-3: Ermitteln des P-Wert (Z. 282-298)

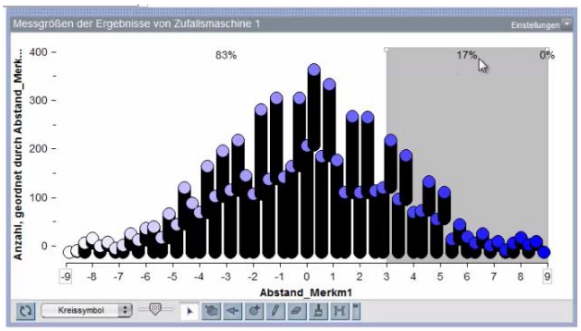
282 L: Ähm (...) wie haben wir denn die 7,7/ oder haben wir einfach den Divider darauf gepackt? Weißt du das noch?
283 F: Wir können doch jetzt einfach den dahin packen (*der Einteiler wird eingeblendet*) und 7,7.



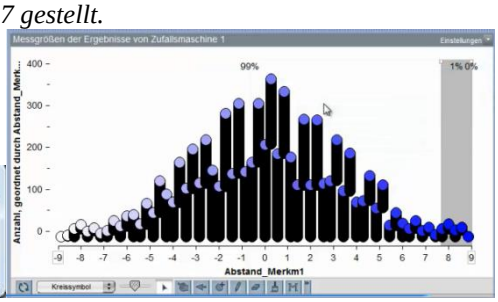
284 *Auf die linke Begrenzung des Einteilers wird doppelt geklickt, so dass sich das Fenster öffnet, in welches 7,7 eingegeben wird.*



285 *Nahezu gleichzeitig werden Prozente eingeblendet. Weil der linke Rand des Einteilers angeklickt wurde und mit einem Wert größer als der bisherige rechte Rand des Einteilers versehen wurde, verschiebt sich der Einteiler automatisch.*



286 L: Und dann musst du dir das glaube ich anzeigen lassen.
287 F: Dann die Prozentzahl. Und dann müsste das doch der P-Wert sein, oder? (*auf die 17% wird gezeigt*) Glaube ich?
288 L: Ich bin mir eigentlich sicher, ja.
289 Alle: (Lachen)
290 F: (4 sec Pause) Also wie groß ist der P-Wert? 17 Prozent. (...) Und das ist dann doch nicht signifikant, oder? Soll ich lieber? (3 sec Pause)
291 I: Moment. Jetzt sag mir erst mal, was du da angezeigt hast.
292 L: Ah das ist/ guck mal das hat der gar nicht gemacht.
293 *Der untere Wert des Einteilers wird auf 7,7 gestellt.*



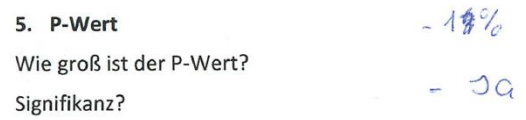
294 F: Ich hab mich gerade schon gewundert. 7,7.
295 L: Hat der eben anders gemacht. Jetzt passt es.
296 I: Du hast das andere Ende hochgezogen.
297 F: Ja das kann sein. (..) Ok das macht ein Prozent. Dann ist es signifikant.
298 L: Ja.

Zu Beginn dieser Episode klären Fabia und Laura kurz noch einmal, dass sie den Einteiler nutzen wollen (282-283) in Verbindung mit dem beobachteten Wert 7,7. Wegen eines Bedienfehlers wird zunächst ein Prozentwert von 17% angezeigt. Das liegt daran, dass das linke Ende des grauen Bereichs des Einteilers per Doppelklick ausgewählt wurde und mit einem Wert größer als das bisherige rechte Ende des Einteilers versehen wurde. Aus diesem Grund verschiebt sich der Einteiler auf einen Bereich von ungefähr 3 bis 9 und gibt hierfür einen Wert von 17% aus (284-286). In (287) fragt Fabia dann: … Und dann müsste das doch der P-Wert sein, oder? Glaube ich?“. Die doppelte Nachfrage innerhalb ihres Turns lässt sich interpretieren als Unsicherheit, ob sie nun den richten Wert als P-Wert identifiziert hat. Laura antwortet daraufhin: „Ich bin mir eigentlich sicher, ja.“ (288). In (290) benennt Fabia zum ersten Mal, dass sie hier einen P-Wert von 17% vorliegen haben und stockt. Danach versucht sie eine Interpretation mit „… Und das ist dann doch nicht signifikant, oder?“. Für einen P-Wert von 17% wäre dies eine richtige Interpretation, allerdings wurde hier nicht der richtige Anteil für den

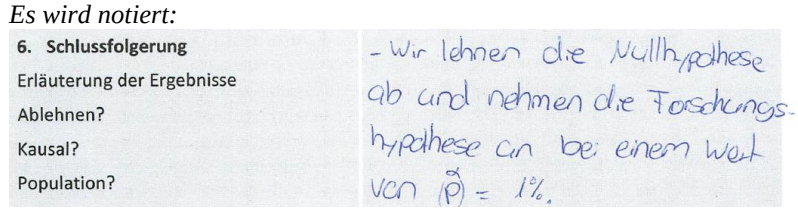
P-Wert bestimmt. Das hat auch die Interviewerin erkannt und greift kurz ein: „Moment. Jetzt sag mir erst mal, was du da angezeigt hast.“ (291), um Fabia und Laura darauf zu bringen, sich den mit dem Einteiler abgedeckten Bereich noch einmal genau anzusehen und gegebenenfalls ihre Einschätzung zu revidieren. Diese Strategie geht auf, denn im nächsten Turn (292) erkennt Laura, dass sie nicht gemessen haben, was sie dachten zu messen und stellt den Einteiler auf den richtigen Wert (293). Nachdem der vom Einteiler überdeckte Bereich nun 1% beträgt, führt Fabia an, dass sie sich schon gewundert habe (294). Dies lässt darauf schließen, dass sie eine bestimmte Erwartungshaltung zur Größe des P-Werts hatte, vermutlich, dass ein signifikanter P-Wert ermittelt wird, was nach dem im Baustein Inferenzstatistik ein Wert ≤ 10% gewesen wäre. Auch Laura bestätigt „… Jetzt passt es.“ (295) und zeigt somit ebenfalls, dass ihre offensichtlich vorhandenen Erwartungen nun erfüllt sind. Fabia fasst dann zusammen: „… Ok das macht ein Prozent. Dann ist es signifikant.“

Interaktionseinheit 11: Schlussfolgerungen ziehen (Z. 299-307)

299 F: Dann/ Weil er nur zu einer Wahrscheinlichkeit von einem Prozent stimmt die Nullhypothese.
300 L: Erst mal muss glaube ich signifikant eingetragen werden.
301 *Es wird notiert:*



302 F: Ja.
303 L: Also äh (...) weil die Nullhypothese nur zu einer Wahrscheinlichkeit von einem Prozent stimmt, lehnen wir diese ab und nehmen die Forschungshypothese an. (5 sec) Oder? Ich finde die Studie trotzdem irgendwie blöd, weil das so wenige Leute sind.
304 F: Ja, oder? (...) Und nur Männer. (lacht)
305 *Es wird notiert:*



306 L: Ja. Gut.
307 F: Wir sind fertig.

Die Interpretation des P-Werts beginnt Fabia direkt mit einer falschen Schlussfolgerung auf Statistikebene: „Dann/Weil er nur zu einer Wahrscheinlichkeit von einem Prozent stimmt die Nullhypothese.“ (299). Dies ist die typische Fehlvorstellung, dass der P-Wert die Wahrscheinlichkeit für die Nullhypothese angibt. Laura geht hierauf zunächst gar nicht ein, sondern beschäftigt sich erst mit dem Ausfüllen des Schemas und trägt dort unter *P-Wert* „1%“ und „Ja“ bei *Signifikant?* ein (300-301). Danach formuliert Laura eine Schlussfolgerung, ähnlich wie zuvor Fabia mit einer falschen Begründung „… weil die Nullhypothese nur zu einer Wahrscheinlichkeit von einem Prozent stimmt, lehnen wir diese ab und nehmen die Forschungshypothese an…“ (303). Es lässt sich

schließen, dass Beide gelernt haben, aus einem P-Wert von 1% zu schließen, dass die Nullhypothese abgelehnt und die Forschungshypothese angenommen werden kann. Eine korrekte Interpretation des P-Werts im Sinne des im Baustein Inferenzstatistik Gelernten wird jedoch von Beiden nicht geliefert. Im Randomisierungstestschema notieren sie nur die gezogene Schlussfolgerung als „Wir lehnen die Nullhypothese ab und nehmen die Forschungshypothese an bei einem Wert von P=1%“. Auf die weiteren Stichworte des Schemas (Kausal und Population) gehen beide nicht ein.

An die in (303) von Laura gezogene Schlussfolgerung über das Annehmen der Forschungshypothese schließt sich ein Kommentar auf Kontextebene zur Studie an: „… Ich finde die Studie trotzdem irgendwie blöd, weil das so wenig Leute sind.“. Darauf geht Fabia ein und antwortet: „Ja, oder? (…) Und nur Männer.“ (304). Sie sprechen hiermit also eine berechtigte Kritik an diesem Experiment indirekt an, dass es sich nicht um eine Zufallsstichprobe handelt, da nur Männer teilgenommen haben und der geringe Stichprobenumfang wird ebenfalls angesprochen, jedoch nicht thematisiert, wieso das „blöd“ ist. Beides führt jedoch nicht dazu, dass sie weitere Schlussfolgerungen in diese Richtung formulieren oder gar im Schema notieren.

Sie erklären anschließend das Durchführen des Randomisierungstests für beendet (307).

Interaktionseinheit 12: Zusammenhang zwischen den formulierten Hypothesen und den möglichen Erklärungen (Z. 309-314)

313	I: Ok. Und wenn ihr jetzt nochmal an vorher denkt. An diese Erklärungen, die wir hatten. Was steckt da jeweils für eine Erklärung dahinter? Auf was kann man die zwei Hypothesen jeweils beziehen? Also welche Erklärung steckt hinter der Nullhypothese? Und welche Erklärung steckt hinter der Forschungshypothese?
314	F: Also bei der Nullhypothese war es auf jeden Fall, dass der Unterschied nur auf die zufällige Einteilung zurückzuführen ist. Und bei der Forschungshypothese ist der Unterschiede wirklich auf die Auswirkung dieser Fischdiät zurückzuführen.

In der zwölften Interaktionseinheit wendet sich die Interviewerin gemäß Leitfaden Frage A1 nach den der Null- und der Forschungshypothese zugehörigen Erklärungen zu. Dies wird nur in den Zeilen (313-314) thematisiert, vorher wird Bezug genommen auf das nun folgende Interview entlang des Leitfadens und die von den Studierenden formulierten Hypothesen.

Fabia beantwortet die Frage und erklärt auf Kontext- und Statistikebene, dass die Nullhypothese auf die zufällige Einteilung zurückzuführen sei und die Forschungshypothese auf die Auswirkung der Fischdiät (314). Dabei stellt sie dies sprachlich knapp dar, ordnet die Erklärungen aber richtig zu.

Interaktionseinheit 13: Beschreiben der Zufallsmaschine (Z. 315-338)

Da während der Erstellung der Zufallsmaschine in Interaktionseinheit 8 wenig diskutiert und erklärt wurde, werden in der dreizehnten Interaktionseinheit nun noch Nachfragen

zur Zufallsmaschine durch die Interviewerin gemäß dem Leitfaden A2 gestellt. Diese Einheit wird in fünf kurze Episoden entlang der Interviewfragen strukturiert. In der ersten Episode geht es um die allgemeine Beschreibung der Zufallsmaschine, in der zweiten Episode wird der Frage nach der Durchgangsanzahl nachgegangen. In der dritten Episode wird besprochen, wieso *ohne Zurücklegen* eingestellt werden musste und in der vierten Episode wird über das Nichtumbenennen der Merkmale und der Etiketten für die zwei neuen Gruppen gesprochen. In der fünften Episode wird noch einmal rekapituliert, wieso die Werte zufällig auf zwei neue Gruppen aufgeteilt wurden. Während des größten Teils der Interaktionseinheit wird mit Bezug zum Kontext argumentiert.

315 I: Mhm (bejahend). Genau. Und jetzt interessiert mich nochmal was ihr in der Zufallsmaschine genau gemacht habt. Kannst du die nochmal groß machen bitte? Die ist ganz links. Dieses kleine Symbol.

316 F: Da. Ich schieb das mal da rüber.

317 Die Studentin arrangiert die Objekte neu und macht die Zufallsmaschine sichtbar.

318 I: Ok. Könnt ihr mir da nochmal bitte erklären, was da vor sich geht?

319 F: Also da sind ja alle Werte drin, die in dieser Studie waren. Also die von der normalen Diät und die von der Fischöldiät. Alle Ergebnisse. Und die wurden jetzt einfach zufällig in zwei andere Gruppen eingeteilt. Also man ist davon ausgegangen, dass einfach nur diese Werte da stehen und man hat jetzt irgendwie die Werte a und b zufällig zugeteilt. Sodass es nicht zurückzuführen ist, ob es jetzt ein normales Öl war oder das Fischöl.

320 I: Mhm (bejahend). Ok. Ähm (.) Ihr habt am Anfang so ein bisschen überlegt, was die Durchgangsanzahl sein muss. Warum muss das vierzehn sein? Warum habt ihr euch für vierzehn entschieden?

321 F: Weil es um vierzehn Personen geht. Und diese vierzehn mussten wir dann gleichmäßig auf zwei Gruppen aufteilen. (…)

322 I: Mhm (bejahend). Ok. Das heißt gleichmäßig auf zwei Gruppen aufteilen. Deswegen sind in dem zweiten Merkmal quasi für a und b sieben Etiketten, oder?

323 M& L: Ja.

324 F: Weil vorher auch sieben-sieben war. (…)

325 I: Mhm (bejahend). Das heißt um diese Gruppengröße die in der Studie/ dem Experiment war um das zu reproduzieren habt ihr das so abgebildet?

326 M& L: Ja.

327 I: Und warum musste jetzt ohne Zurücklegen eingestellt werden?

328 F: Weil jede/ Wir gucken uns immer erst mal nur eine Ziehung an und da ist jede Person nur einmal abgebildet. Also diese zwei Gruppen. In jeder Gruppe sind sieben Personen und jede Person ist auch nur einmal in jeder Gruppe.

329 L: Es sind ja auch nur sieben Plätze in jeder Gruppe da. Sonst wären in jeder Gruppe vielleicht mehr.

330 I: Mhm (bejahend). Das ist sozusagen die Erklärung warum hinten. Also in diesem zweiten Bauteil insgesamt nur vierzehn drin sind und ohne Zurücklegen gezogen wurde. Und vorne, wo ihr die Werte habt?

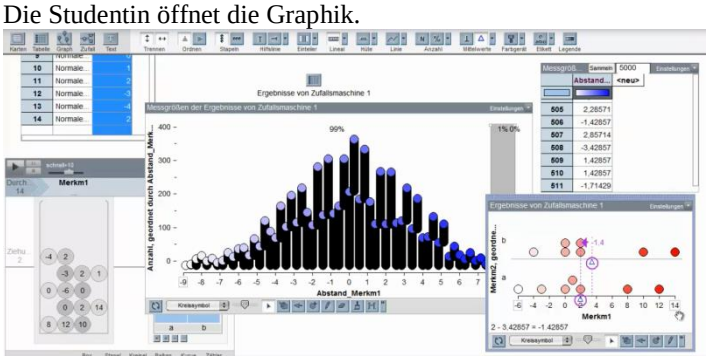
331 L: Weil es ja nur vierzehn Werte sind. Es darf ja jeder nur einmal verwendet werden. Man darf ja nicht zufällig die zwölf plötzlich drei Mal.

332 I: Ok. (...) Gut. Noch irgendwas, das ihr dazu beschreiben wollt? (..) Ah vielleicht noch. Ihr habt jetzt nichts irgendwo umbenannt, also ihr habt die Merkmale so belassen, wie sie automatisch sind und auch das a und b. Warum habt ihr das gemacht?

333 F: Ist doch eigentlich egal, wie die Gruppe jetzt heißt, oder? (..) Das ist eine Schönheitssache. (…)

334 I: Das ist völlig in Ordnung. Hat aber zwischendrin offensichtlich nicht zu wie irgendwelchen Irritationen geführt. Also als ihr dann die Graphik erstellt habt/ Machst du die nochmal auf?/ Sozusagen wie ihr es gesagt habt, diese eine Randomisierung, diese erste Randomisierung.(...) Da gab es aber soweit ich es nachvollziehen konnte, keine Probleme, weil ihr nicht genannt habt das war DIE Gruppe.

335 Die Studentin öffnet die Graphik.



336 F: Nee. Weil ob es jetzt a oder b ist oder irgendein Wert, das ist ja logisch, wir sprechen ja immer von den gleichen Werten. Dann muss man denen nicht noch Namen geben. Es sind ja nur zwei Merkmale. Dann ist das ja nicht so schwer. (lachen)

337 I: Okay. Ähm und jetzt nochmal die Frage: Was bedeutet es denn hier, dass man die Werte zufällig auf zwei neue Gruppen aufteilt?

338 L: Also man geht ja davon aus, dass die Nullhypothese stimmt. Das ist ja eigentlich diese Grundlage, die man hat. Dass man sagt. Es hat keine Auswirkung welche Gruppe man hat. Die Werte sind immer ähnlich. Egal in welcher Gruppe das zugeteilt ist. Und deshalb macht man das zufällig, weil man dann halt sieht, wie die Mittelwerte verteilt sind, wenn es egal ist, was für eine Auswirkung es hat. Welches Öl benutzt wird quasi.

Episode 13-1: Allgemeine Beschreibung der Zufallsmaschine (315-319)

Nachdem die Zufallsmaschine unter den minimierten Objekten gefunden und wieder vergrößert wurde (315-317), stellt die Interviewerin zunächst eine allgemeine Frage, „… was da vor sich geht“ (318). Fabia übernimmt den Turn in (318) und beschreibt die zwei Bauteile mit Bezug zum Kontext. Sie erkennt, dass die Blutdruckwerte losgelöst vom zuvor eingenommenen Öl betrachtet werden und diese zufällig auf zwei neue Gruppen (von ihr Werte genannt) zugeordnet werden.

Episode 13-2: Erklärung der Durchgangsanzahl (320-326)

Die nächste Frage betrifft die Durchgangsanzahl, die zunächst in den Episoden 8-4 und 8-5 eine Schwierigkeit dargestellt hat. In (321-326) erklären Laura und Fabia, dass vierzehn als Durchgangsanzahl gewählt wurde, weil es um vierzehn Personen gehe (321), die auf zwei Siebenergruppen aufgeteilt werden müssen (324). Auch diese Erklärungen sind gut und lassen Bezug zum Kontext der Studie erkennen.

Episode 13-3: Erklärung der Einstellung Ohne Zurücklegen (327-331)

Daran schließt sich die Frage der Interviewerin an, wieso die Einstellung ohne Zurücklegen ausgewählt wurde (327). Fabia argumentiert zunächst, dass jede Person nur einmal abgebildet werde und jede Person nur einmal in den Gruppen vorkommen dürfe (328). Laura argumentiert, dass es nur sieben Plätze in jeder Gruppe gebe (329) und sonst mehr Personen in einer Gruppe sein könnten. Dies wird von der Interviewerin als „Erklärung für das zweite Bauteil“ bezeichnet (330). Für das erste Bauteil mit den Blutdruckwerten antwortet Laura, dass jeder Wert nur einmal verwendet werden dürfe, damit nicht ein Wert, beispielsweise die Zwölf, dreimal gezogen würde (331). Die in dieser Episode genannten Erklärungen zeigen, dass die beiden Teilnehmerinnen sowohl statistisch als auch auf Softwareebene verstanden haben, was in der Zufallsmaschine geschieht und dies gut mit Bezug zum Kontext erläutern können.

Episode 13-4: Erklärung des Nichtumbenennens der Merkmale (332-336)

Es schließt sich die Frage der Interviewerin an, wieso weder die Merkmale, noch die Etiketten der neuen Gruppen umbenannt wurden (332). Fabia formuliert direkt: „Ist doch eigentlich egal, wie die Gruppe jetzt heißt, oder? Das ist eine Schönheitssache.“ (333). In (336) bekräftigt sie noch einmal „… ob es jetzt a oder b ist oder irgendein Wert, das ist ja logisch…“ und „… Es sind ja nur zwei Merkmale. Dann ist das ja nicht so schwer.“ Da während des Prozesses keine Schwierigkeiten aufgetreten sind, die auf eine nicht erfolgte Umbenennung der Merkmale oder der Gruppenetiketten zurückzuführen sind, stimmt diese Einschätzung wohl.

Episode 13-5: Erklärung der zufälligen Aufteilung (337-338)

Als letzte Frage dieser Interaktionseinheit stellt die Interviewerin: „… Was bedeutet es denn hier, dass man die Werte zufällig auf zwei neue Gruppen aufteilt?“ (337). Laura antwortet und führt alle nötigen Argumente an. Diese sind in der folgenden Abbildung 208 dargestellt.

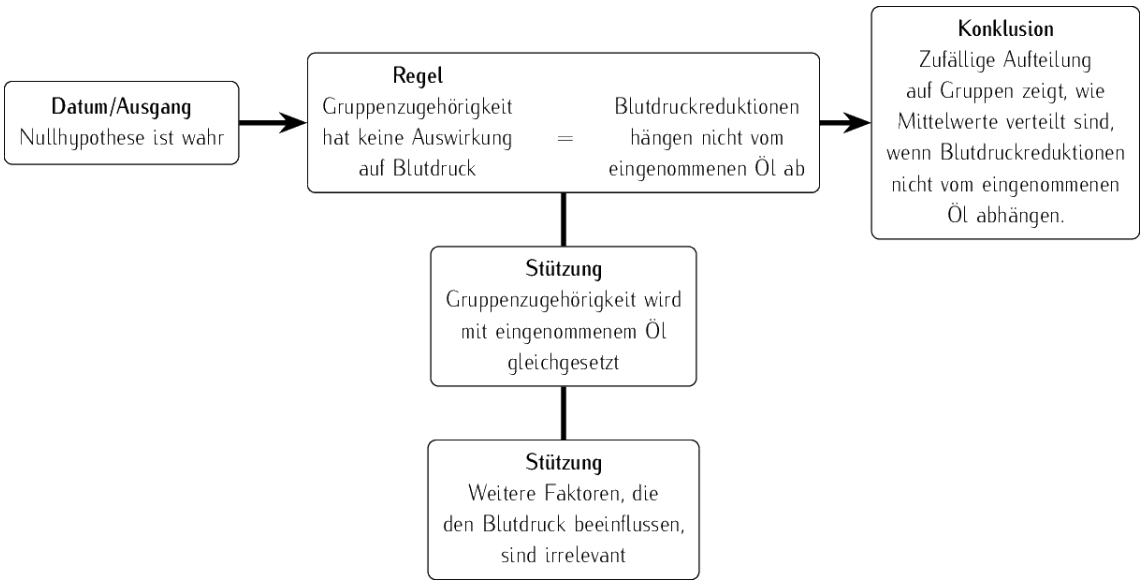


Abbildung 208: Rekonstruierte Argumentation von Laura zur zufälligen Aufteilung der Versuchspersonen auf die Gruppen

Interaktionseinheit 14: Verständnis über einzelne Punkte der Referenzverteilung (Z. 339-363)

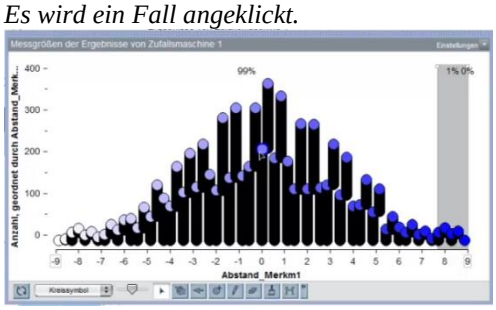
In dieser Interaktionseinheit wird der Leitfrage A3 des Interviewleitfadens nachgegangen. Diese Einheit wird in drei Episoden geteilt. In der ersten Episode werden die Bedeutungen zwei einzelner Fälle der Referenzverteilung erklärt (Frage A3a des Leitfadens) und in der zweiten Episode wird ein Fall aus dem Bereich des P-Werts ausgewählt und in Bezug auf die Nullhypothese interpretiert, im Wesentlichen formuliert von der Interviewerin (Frage A3b und A3c des Leitfadens). In der dritten Episode wird der Frage A3d des Leitfadens nachgegangen, ob ein Wert von 8 überraschend ist.

Episode 14-1: Erklären einzelner Punkte der Referenzverteilung (Z. 339-347)

339 I: Genau. Das war sozusagen der wichtige Nachzug. Dieser Bezug zu dem Kontext. Dass es egal ist, ob man Fischöl genommen hat oder nicht. Man tut sozusagen so als ob eine Person einen festen Wert hätte für die Blutdrucksenkung nach vier Wochen. Egal, ob sie Fischöl genommen hat oder nicht. Genau. Jetzt gucken wir nochmal auf die Referenzverteilung. Klick doch mal bitte einen Punkt über der Null an. (...) So ungefähr über der Null.

340 L: Achso hier bei der Null.

341 *Es wird ein Fall angeklickt.*



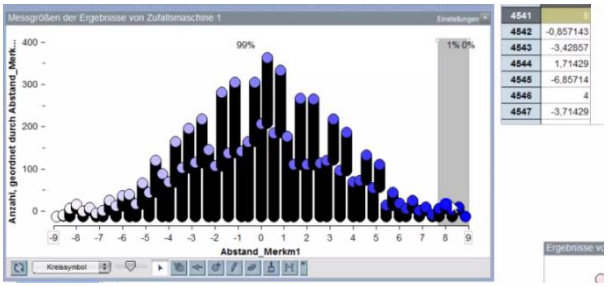
342 I: Ja irgendeinen. So. Wenn wir uns DEN angucken, könnt ihr bitte nochmal erklären, was dieser

eine Punkt bedeutet?

343 L: Ja das ist jetzt ja EIN ermittelter Abstand zwischen zwei Mittelwerten. Von einer dieser Gruppenzuteilung waren die Mittelwerte gleich quasi. Also gab es keinen Unterschied dazu.

344 I: Genau. (...) Und, wenn ihr jetzt einen Punkt anklickt, der da hinten im Bereich des Einteilers liegt. Irgendeinen der bei der neun ungefähr ist. (...) Könnt ihr den bitte auch nochmal erklären?

345 *Es wird ein Fall über der 8 markiert.*



346 L: Willst du vielleicht?

347 F: Ich muss überlegen. (...) Dass da halt ein Unterschied zwischen den beiden Gruppen, der (...) bei 8 liegt? Ein Unterschied der beiden Mittelwerte bei 8.

In (339) fasst die Interviewerin zunächst das Gesagte zur vorigen Interaktionseinheit zusammen und leitet dann über zur Referenzverteilung, zu der ein Punkt mit dem Wert 0 angeklickt wird (341) und erklärt werden soll, was dieser Punkt bedeute (342). Laura ergreift hierauf den Turn und antwortet auf statistischer Ebene korrekt, dass es sich um einen ermittelten Abstand zwischen zwei Mittelwerten handele und schließt daraus, dass in diesem Fall die Mittelwerte der beiden Gruppen gleich waren (342). Dies ist eine gute Erklärung auf Ebene der Statistik, die Kontextebene wird nicht einbezogen. In (344) fragt die Interviewerin danach, was ein weiterer Punkt bedeute (344), diesmal mit dem Wert 8 (345). Laura übernimmt den Turn in (346), gibt die Frage jedoch an Fabia weiter. Dies lässt sich auf zwei Arten interpretieren. Entweder kann sie selbst keine Antwort liefern und möchte deshalb, dass Fabia die Frage beantwortet, oder sie möchte höflicherweise Fabia die Möglichkeit geben, ihr Wissen darzustellen, weil Laura selbst gerade schon eine Frage sehr gut beantwortet hat. Aus dem bisherigen Gesprächsverlauf und mit Blick auf die weiteren Turns nach der Turn-by-Turn Analyse ist die zweite Interpretation wahrscheinlicher. Fabia nimmt die Zuweisung des Turns an und antwortet zögerlicher als Laura vorher, „dass da halt ein Unterschied zwischen den beiden Gruppen, der bei 8 liegt? Ein Unterschied der beiden Mittelwerte bei 8.“ (347). Auch dies ist eine richtige Interpretation des markierten Falls, Fabia antwortet ebenso wie Laura vorher auf statistischer Ebene. Ein Bezug zum Kontext wird von beiden nicht hergestellt. Man hätte hier beispielsweise formulieren können, dass im Fall der Differenz von 0 zwischen den Mittelwerten der beiden Gruppen die Blutdruckreduktionen in beiden Gruppen im Mittel ähnlich verteilt waren oder anders ausgedrückt, dass hier der Blutdruck in den beiden neuen Gruppen im Mittel weder erhöht noch gesenkt wurde.

Episode 14-2: Bezug auf die Nullhypothese für einen Fall mit dem Wert 8 (Z. 348-363)

348 I: Mhm (bejahend). Und was heißt das in Bezug auf die Nullhypothese?

349 F: Ja er liegt ja in dem Bereich und das war ja hier 7,7. (…)

350 L: Ja an dieser Stelle ist es halt so, dass die Nullhypothese/ dass man da zustimmen könnte. Also dass es ein ähnlicher Unterschied ist wie hier (*zeigt auf die beobachteten Daten des Experiments*) und dass man dann doch vom Unterschied sprechen könnte, wenn man denn nur diese Gruppeneinteilung hätte.

351 I: Was meinst du, dass man von einem Unterschied sprechen könnte?

352 L: Zwischen den Auswirkungen von der Fischöldiät und der normalen Diät, weil da ja der Unterschied der Mittelwerte sehr hoch ist. Also/.

353 I:/ Ja ich glaube jetzt vertauscht du gerade was. Weil was du vorhin gesagt hast war richtig. Das jetzt aber stimmt nicht. (...) Du hast die ganze Zeit/ oder ihr habt zusammen die ganze Zeit simuliert unter der Annahme, dass die Nullhypothese richtig ist, dass die stimmt. Habt ihr gesagt. Das also alles, was man beobachtet hier in/ von diesen vierzehn Leuten nur daran liegt, dass die zufällig auf die Gruppen eingeteilt wurden. Und das habt ihr hier simuliert. Und jetzt ist da hinten aber ein Wert rausgekommen. Meinetwegen der bei 8. Der sozusagen noch größer ist als der Unterschied, der hier beobachtet wurde. Bei dem Experiment.

354 L: Das heißt in dem Fall würde die Nullhypothese dann nicht stimmen quasi, wenn man nur von dieser einen ausgehen würde. Das ist halt dieser Bereich ein Prozent von diesen ganzen Werten, die wir ermittelt haben. Der halt in dem Bereich liegt. Wo man den deutlichen Unterschied sieht quasi. Also es gibt einen Unterschied.

355 F: Es gibt auf jeden Fall einen Unterschied von 8 und/.

356 I:/ Ich glaube ihr denkt gerade schon einen Schritt weiter, als ich denke. (...) Alleine durch die zufällige Zuordnung ist ein Unterschied von 8 zwischen den beiden Mittelwerten aufgetreten. Das hat immer noch nichts mit dem Fischöl oder nicht mit dem Fischöl zu tun. Das heißt an der Stelle/ anders als ihr gerade gesagt habt / ist die Nullhypothese verträglich mit dem, was simuliert wurde. Unter dieser zufälligen Zuordnung haben wir auch einen Wert beobachtet, der größer ist als das, was wir im Experiment beobachtet haben. Das heißt rein durch die zufällige Zuordnung kann es auch sein, dass man einen Wert wie 7,7 oder auch 8 oder was da rausgekommen ist, als Unterschied zwischen den Mittelwerten beobachtet.

357 F: Ja. Aber halt dann nur zu diesem einen Prozent.

358 I: Richtig. Das ist wieder die wichtige Feststellung. In einem Prozent der Fälle ist es vorgekommen, dass man NUR durch die zufällige Zuordnung der Werte auf zwei neue Gruppen so einen Wert wie 7,7 oder noch größer beobachtet hat.

359 M&L: Mhm (bejahend).

Nachdem die statistische Bedeutung eines Falls mit dem Wert 8 geklärt wurde, wird in dieser Episode von der Interviewerin gefragt, was dies in Bezug auf die Nullhypothese heiße (348). Fabia antwortet zunächst mit „Ja er liegt ja in dem Bereich und das war ja hier 7,7 (…).“ (349) und lässt weiteres offen, sie endet mit einer Pause. Mit „dem Bereich“ ist vermutlich der Bereich gemeint, der mit dem Einteiler markiert ist bzw. der Bereich, der durch den P-Wert beschrieben wird, also die Werte ab 7,7 gemeint. Laura übernimmt den Turn und erläutert ausführlicher, dass man da der Nullhypothese zustimmen könne (350). Sie verweist richtigerweise darauf, dass „es ein ähnlicher Unterschied ist wie hier“ indem sie auf die beobachteten Daten und den dort gemessenen Unterschied zwischen den beiden Gruppen zeigt und schließt daraus, dass „man dann doch vom Unterschied sprechen könnte, wenn man denn nur diese Gruppeneinteilung hätte.“ Dieser Schluss ist unklar, was auch von der Interviewerin so empfunden wird, denn sie fragt in (351), was Laura damit meine. Diese präzisiert: „Zwischen den Auswirkungen von der Fischöldiät und der normalen Diät, weil da ja der Unterschied der Mittelwerte sehr hoch ist. Also/“ (352). Zusammen mit Lauras vorherigem Turn lässt sich daraus

interpretieren, dass sie mit dem ersten Unterschied in (350, „ein ähnlicher Unterschied wie hier“) die Differenz der Mittelwerte zwischen den beiden Gruppen des ursprünglichen Experiments meint und mit dem zweiten Unterschied (350, „dass man doch von Unterschied sprechen könnte, wenn man denn nur diese Gruppeneinteilung hätte.“) ausdrücken möchte, dass in diesem Fall im Gegensatz zur in der Nullhypothese angenommenen Situation eben doch ein *Unterschied* zwischen den beiden Gruppen, respektive den beiden Öldiäten, zu beobachten ist. Dies wäre jedoch keine richtige Interpretation, wenn sie aus diesem einen Fall, der einen Wert größer als der beobachtete hat, darauf schlösse, dass die Nullhypothese nicht gilt. Diese Interpretation lässt außer Acht, dass unter der Annahme simuliert wurde, dass kein Unterschied zwischen den beiden Gruppen herrscht. In (353) greift die Interviewerin das auf und fasst die Simulation unter der Nullhypothese noch einmal zusammen. Laura geht darauf ein und antwortet: „Das heißt in dem Fall würde die Nullhypothese dann nicht stimmen quasi, wenn man nur von dieser einen ausgehen würde…“ (354). Diese Interpretation ist unsinnig, denn ein einzelner Fall lässt keine Schlussfolgerung zu. Als nächstes bezieht sie sich jedoch auf den Bereich, der durch den P-Wert beschrieben wird und quantifiziert ihn mit einem Prozent aller Werte, die sie ermittelt hätten. Im nächsten Turn (358) versucht die Interviewerin noch einmal, die vorliegende Situation zu erklären, insbesondere mit Blick darauf, dass die Nullhypothese, also die zufällige Zuordnung als Erklärung für die beobachteten Unterschiede die Grundlage der simulierten Werte ist. Nach einem Einwurf von Fabia (357) zur Größenordnung, in der ein Wert wie 7,7 oder größer simuliert wurde, liefert nun die Interviewerin die statistische Interpretation des hier aufgetretenen P-Werts von einem Prozent. Fabia und Laura bejahen dies (359) und damit ist diese Episode abgeschlossen und das hervorgebrachte Wissen gilt als akzeptiert. Die wesentlichen Inhalte hat hier jedoch die Interviewerin formuliert und nicht die beiden Teilnehmerinnen.

Episode 14-3: Überrascht ein Wert von 8? (Z. 360-363)

360	I: Ok. Überrascht euch so ein Punkt, der da in dem grauen Bereich liegt, oder die Punkte, die da liegen?
361	F: Es kommt immer drauf an. Wir haben das ja ziemlich oft durchgeführt. 5000 Mal. Da kann es ja dazu kommen, dass so eine hohe Differenz zustande kommt. Es sind ja jetzt auch nicht so viele Werte. Ich glaube, da kann das schon schnell. Deshalb ist es wichtig, dass man es oft machte. Wenn man es 10 Mal macht und einmal das Pech hat, dass dann so ein Wert rauskommt, dann hat das ja eine viel größere Auswirkung.
362	I: Mhm (bejahend). Also jetzt nicht so sehr überraschend für euch?
363	F: Nee.

In der letzten kurzen Episode dieser Einheit wird von der Interviewerin die Frage gestellt, ob ein Punkt des grauen Bereichs überraschend wäre (360). Fabia antwortet hierauf in (361) mit Bezug auf die von ihnen durchgeführte Wiederholungsanzahl von n = 5000, bei der es schon dazu kommen könne, „… dass so eine hohe Differenz zustande kommt. Es sind ja jetzt auch nicht so viele Werte…“. Diese Aussage lässt vorsichtige Einblicke auf eine bei Laura möglicherweise vorhandene Erwartungshaltung

bezüglich der Referenzverteilung zu, denn offensichtlich erwartet sie, dass einige wenige Werte ähnlich wie der im ursprünglichen Experiment beobachtete auftreten. Hier spielt gegebenenfalls auch eine Rolle, dass in den Beispielen, die im Baustein Inferenzstatistik gezeigt wurden, der P-Wert immer einen Wert zwischen 0,01 und 0,05 hatte, und im Beispiel, das im Seminar verwendet wurde, Werte zwischen 0,01 und 0,3 aufgetreten sind, also immer Fälle unter der Nullhypothese aufgetreten sind, die gleich oder größer dem ursprünglich beobachteten Wert waren. Implizit nimmt Fabia noch Bezug auf das Gesetz der großen Zahlen mit dem zweiten Teil der Aussage, indem sie erläutert, dass wenn man die Simulation zehn Mal durchführe, und davon ein Wert so groß sei, „dann hat das ja eine viel größere Auswirkung.“ (361). Dies ist eine gute Erklärung, wenn sie im Hinterkopf hat, dass in diesem Fall der P-Wert 10% betragen würde und somit der P-Wert eine Größe annähme, ab der die Nullhypothese nicht mehr abgelehnt würde.

Interaktionseinheit 15: Schlussfolgerungen ziehen (Z. 364-401)

Die letzte Interaktionseinheit des Interviews bezieht sich auf die Fragen A4-A9 des Leitfadens. Aufgrund der Leitfadenfragen wird die Einheit in fünf Episoden gegliedert. In der ersten Episode wird der statistischen Bedeutung des P-Werts nachgegangen (A4), in der zweiten Episode der Erklärung des P-Werts im Kontext (A5) und der Wirksamkeit von Fischöl (A6). In der dritten Episode wird thematisiert, wie sicher die Studierenden bezüglich ihrer Einschätzung sind (A7). In der vierten Episode wird im Wesentlichen von der Interviewerin eine alternative Interpretation des P-Werts geliefert (A8) und in der letzten Episode wird geklärt, was aus einem fiktiven P-Wert von 30% geschlossen werden könnte (A9).

Episode 15-1: Statistische Erläuterung des P-Werts (364-377)

364 I: Ok. Gut damit haben wir dieses Randomisieren und Wiederholen auch nochmal geklärt. Dann kommen wir jetzt nochmal zum Interpretieren, also zu dem Schlussfolgern. Das was ihr jetzt rausbekommen habt. Ihr habt gesagt, der P-Wert beträgt hier ein Prozent. Das ist richtig. Was war nochmal der P-Wert? Könnt ihr das nochmal formulieren?

365 L: Der misst die Wahrscheinlichkeit (...) ob die Nullhypothese stimmt. (murmelt)

366 I: Nee. Genau das nicht.

367 L: Nicht?

368 I: Nein. Wir können gar keine Wahrscheinlichkeiten für Hypothesen angeben. Das gibt es leider nicht. Hab ich versucht in der letzten Vorlesung deutlich zu machen. also das ist es gerade nicht. Der P-Wert ist NICHT die Wahrscheinlichkeit dafür, dass die Nullhypothese stimmt.

369 F: Also kann ich sagen, die ungefähre Wahrscheinlichkeit dafür, dass man die Nullhypothese ablehnen kann oder nicht. Annehmen kann oder nicht?

370 I: Nee, auch das nicht. Wir gehen auch bei dem P-Wert davon aus, dass die Nullhypothese richtig ist. Das stecken wir die ganze Zeit rein. Auch bei dem P-Wert. Der ist ja eine Wahrscheinlichkeit aus dieser Referenzverteilung. Und diese Referenzverteilung beruht darauf, dass ihr vorher diese zufällige Zuteilung, sprich die Nullhypothese, reingesteckt habt. Das heißt, auch der P-Wert ist eine Wahrscheinlichkeit dafür, irgendwas UNTER der Nullhypothese zu haben. Also unter der Voraussetzung, dass die Nullhypothese richtig ist.

371 L: Achso ja, die Wahrscheinlichkeit dafür, dass man einen solchen Wert oder einen extremeren

bekommt unter der Voraussetzung, dass die Nullhypothese stimmt.
372 I: Genau. Das war es. Aber ist auch inhaltlich klar was das bedeutet?
373 L&F: Ja.
374 I: Weil, ich sag mal, das ist eine Formulierung, die habe ich euch natürlich beigebracht in der Vorlesung. Wie man hört ist die auch ganz gut griffig.
375 F: (lacht) Ja jetzt ist es immer klar/.
376 I: / Wie gesagt, wichtig ist mir, dass ihr an der Stelle versteht, ihr steckt quasi die Nullhypothese als wahr von vorneherein rein.
377 F: Mhm (bejahend).

In (364) fordert die Interviewerin dazu auf, auszuformulieren, was der P-Wert angibt. Laura übernimmt den Turn und antwortet auf Statistikebene: „Der misst die Wahrscheinlichkeit (…) ob die Nullhypothese stimmt.“ (365). Dies ist eine typische Fehlvorstellung, die sie bereits vorher formuliert hat (Interaktionseinheit 11). Dass dies eine falsche Definition ist, wird von der Interviewerin in (366) gesagt, was eine ungläubige Nachfrage von Laura nach sich zieht (367). Offenbar hat sie diese Fehlvorstellung fest verankert. Dies war auch der Eindruck der Interviewerin, weshalb sie im darauffolgenden Turn (368) diese Fehlvorstellung thematisiert. Daraufhin versucht Fabia, den P-Wert auszuformulieren mit „Also kann ich sagen, die ungefähre Wahrscheinlichkeit dafür, dass man die Nullhypothese ablehnen kann oder nicht. Annehmen kann oder nicht?“ (369), womit sie erneut versucht, eine Wahrscheinlichkeitsangabe für die Nullhypothese zu machen, auch wenn sie diese nun nicht mehr quantifiziert. Auch dies wird jedoch von der Intervierwein im folgenden Turn (370) als falsch abgelehnt und sie betont, dass die Nullhypothese als wahr vorausgesetzt wird. Nun greift wieder Laura zum Turn (371) und formuliert den P-Wert als: „Achso ja, die Wahrscheinlichkeit dafür, dass man einen solchen Wert oder einen extremeren bekommt unter der Voraussetzung, dass die Nullhypothese stimmt.“ Dies ist nun eine richtige Definition des P-Werts und entspricht im Wortlaut dem, wie es im Baustein Inferenzstatistik gelehrt wurde. Da es sich nach einer auswendiggelernten Definition anhört, fragt die Interviewerin in (372), ob klar sei, was dies inhaltlich bedeute, was von Beiden in (373) bestätigt wird. Zur Bekräftigung wird dies in (375-377) verdeutlicht. Zum Kontext wird in dieser Episode kein Bezug genommen, das Gespräch findet auf einer allgemein statistischen Ebene statt.

Episode 15-2: Der P-Wert im Kontext (378-382)

378 I: Und jetzt kommt ihr dazu, wenn tatsächlich ein Unterschied von 7,7 beobachtet wurde, dann habe ich das bei einer zufälligen Zuteilung nur in einem Prozent der Fälle. Ja, das ist die Wahrscheinlichkeit, die hinter dem P-Wert steckt. Gut (*blickt auf den Zettel*) und ich glaube das haben wir gerade schon gemacht. Könnt ihr das mit dem P-Wert jetzt nochmal auf den Kontext hier beziehen, mit dem Fischöl?
379 F: Ja wir haben ja jetzt einen P-Wert von einem Prozent, der ist ziemlich niedrig und immer wenn der P-Wert 10 Prozent oder niedriger ist, dann können wir die Nullhypothese ablehnen. Und wenn der P-Wert aber größer als 10 Prozent ist, stimmt die Nullhypothese.
380 I: es ist so auch nicht ganz richtig. Ja willst du was sagen?
381 L: Man kann es so formulieren, wenn man davon ausgeht, dass es keine Auswirkung hat welches

Öl man benutzt, dann kann zu einem Prozent der Fälle so ein extremer Wert rauskommen wie da ist in den Mittelwerten. In dem Unterschied der Mittelwerten?

382 I: Genau. Was du gerade gesagt hast, war richtig. Man würde hier tatsächlich die Nullhypothese ablehnen, aber wenn der P-Wert größer ist als 10 Prozent ist, dann kann man gar nichts aussagen. Dann kann die Nullhypothese richtig sein, aber es kann auch andere Erklärungen geben. Also sie kann auch immer noch falsch sein. Dann lässt sich keine Aussage treffen. Aber das ist immer dieses Bestreben: Nee ich will ja irgendwas annehmen, ich will ja hinterher irgendwas sagen können. Deshalb sind Forscher auch so erpicht darauf, dass sie signifikante Ergebnisse bekommen, denn dann können sie immer sagen: Ja gut wir lehnen die Nullhypothese ab und damit gilt dann unsere Forschungshypothese. Ok. Das ist sozusagen die EINE Interpretation. (...) Was heißt das jetzt in Bezug auf das Fischöl. Lässt sich jetzt also sagen, dass Fischöl eine blutdrucksenkende Wirkung hat?

383 M&L: Ja.

384 F: Ja, das würde ich schon sagen. Aber nur bei diesen.

385 L: Weil wir es mit diesen durchgeführt haben.

In der nun folgenden Episode wird von der Interviewerin nach dem Bezug des P-Werts zum Kontext gefragt (377). Fabia übernimmt den Turn (378) und bezieht sich zunächst auf die im Baustein Inferenzstatistik gelernte Kategorisierung des P-Werts, um eine Aussage über Signifikanz zu treffen. Mit dieser Begründung argumentiert sie, dass die Nullhypothese bei einem P-Wert kleiner gleich zehn Prozent abgelehnt werden könne und dass die Nullhypothese bei einem P-Wert größer als zehn Prozent stimme. Der erste Teil der Aussage ist richtig, im zweiten Teil zeigt sich eine weitere Fehlvorstellung, dass nämlich bei einem P-Wert größer als zehn Prozent die Nullhypothese als wahr betrachtet würde. Dies wird nur kurz kommentiert von der Interviewerin (380), daraufhin meldet sich jedoch Laura zu Wort (381). Sie formuliert nun korrekt, wie der P-Wert in Höhe von einem Prozent interpretiert werden kann, ohne jedoch daraus eine Schlussfolgerung zu ziehen. Dies geschieht in (382) durch die Interviewerin. Sie liefert auch eine Schlussfolgerung für einen fiktiven P-Wert größer als zehn Prozent. Am Ende dieses Turns kommt sie nun noch auf die Frage zu sprechen, ob nun gesagt werden könne, dass Fischöl eine blutdrucksenkende Wirkung habe. Beide Teilnehmerinnen antworten hierauf mit „Ja.“ (383) und Fabia schränkt ein „… Aber nur bei diesen.“ (384), was von Laura direkt aufgenommen wird: „Weil wir es mit diesen durchgeführt haben.“ (385). Sie schränken also richtigerweise die Reichweite der Schlussfolgerung, dass Fischöl wirksam ist, um den Blutdruck zu senken, auf die Stichprobe ein, mit der das Experiment durchgeführt wurde. So wurde es im Baustein Inferenzstatistik thematisiert und hier von den beiden Studierenden wiedergegeben.

Episode 15-3: Sicherheit bezüglich der Einschätzung (386-390)

386 I: Genau. (..) Das ist richtig. Wie sicher seid ihr euch dabei? Das haben wir nicht in der Vorlesung gelernt.

387 L: Eigentlich schon sicher, weil wir das ja 5000 Mal durchgeführt haben. (...) Und ein Prozent ist ja schon / also nur anhand diesen hier jetzt. Wenn ich mir nur diese Werte (*zeigt auf das Arbeitsblatt*) angeguckt hätte, hätte ich gedacht es sind viel zu wenige Leute. Um es auf die Allgemeinheit zu schließen vor allem.

388 I: Aber wir bleiben jetzt erst mal bei diesen.

389 L: Ja doch, dann schon.
390 F: Da schon recht sicher, eigentlich.

An die Interpretation, dass Fischöl wirksam ist, um den Blutdruck der Versuchspersonen zu senken, schließt sich die Frage der Interviewerin an, wie sicher die Beiden bezüglich dieser Interpretation seien (386). Mit Bezug auf die Wiederholungszahl von 5000 antwortet Laura (387), dass sie „schon sicher“ sei. Dann bezieht sie noch einmal die Stichprobengröße des Experiments ein und formuliert erneut, dass sie diese als zu gering ansieht, um auf die Allgemeinheit zu schließen. Einschränkend formuliert die Interviewerin im nächsten Turn, dass bei den Versuchspersonen geblieben werden solle (388), woraufhin Beide antworten, dass sie für diese Personen schon sicher seien (389-390). Es zeigt sich in der Wortwahl eine gewisse Abneigung, eine absolute Aussage, wie beispielsweise *Ich bin sicher* zu treffen.

Episode 15-4: Alternative Interpretation des P-Werts (391-395)

391 I: Ok. Und jetzt hätte ich gerne noch die alternative Interpretation. Also die eine Interpretation habt ihr gerade genannt. Wenn ich einen P-Wert von einem Prozent in der Studie habe, kann ich schließen, dass die Nullhypothese nicht die plausible Erklärung ist, sondern dass ich sagen kann, Fischöl ist wirksam. Was könnte noch sein? Was wäre eine andere Interpretation? Des P-Werts.
392 F: Mit dem gleichen P-Wert?
393 I: Mit dem gleichen P-Wert. (...) Wenn ihr jetzt überlegt, dass die Nullhypothese wahr ist.
394 L: Ja dann müssten wir davon ausgehen, dass in den Gruppen irgendwas anderes zusätzlich gemacht wurde. Dass es noch andere Faktoren außer das Öl gibt, die diese Unterschiede der Gruppen beeinflusst haben.
395 I: Ok. Das ist tatsächlich eine Erklärung, an die habe ich gar nicht gedacht. Aber das ist richtig. Also selbst, wenn man eine Studie noch so kontrolliert/ anlegt, irgendwas kann man nicht beeinflussen. Solche Faktoren könnten das sein. Ich kann natürlich kontrollieren, ob die Leute viel Sport oder wenig Sport machen, wie alt sie sind. Das kann ich alles kontrollieren, aber hat der heute Morgen einen Anruf gekriegt und ist deshalb gestresst und hat einen höheren oder niedrigen Blutdruck, kann ich nicht unbedingt beeinflussen. Was ich eigentlich meinte war (...) es könnte trotzdem an der zufälligen Zuordnung der Leute zu den Gruppen gelegen haben. Dann hätte ich etwas Seltenes beobachtet. Das ist auch möglich. Also in einem Prozent kommt ein solcher Wert oder größer raus, wenn ich die Leute zufällig aufteile. Das heißt ich könnte hier ein seltenes Ergebnis gehabt haben. Und tatsächlich läge nicht am Fischöl, auch wenn ihr euch recht sicher wart. Ne das war dieses Entscheiden unter Unsicherheit. Es bleibt diese Unsicherheit, dass ich nie 100 Prozent sicher weiß, dass die Forschungs- oder die Nullhypothese eben richtig sind. Bei der Nullhypothese weiß ich es eh nicht. Ok. Und jetzt darauf bezogen, wenn wir jetzt nicht einen P-Wert von einem Prozent gehabt hätten, sondern einen P-Wert von 30 Prozent, was würdet ihr dann schließen?

Eine alternative Interpretation des P-Werts, wie zum Beispiel dass etwas Seltenes im ursprünglichen Experiment passiert ist, dass es aber trotzdem sein kann, dass die Nullhypothese die plausible Erklärung ist, wird von Fabia und Laura nicht geliefert auf die Frage der Interviewerin (391-393). Ihnen scheint nicht klar zu sein, dass die Frage in diese Richtung gemeint ist, stattdessen argumentieren sie damit, dass „es noch andere Faktoren außer das Öl gibt, die diese Unterschiede der Gruppen beeinflusst haben.“ (394). Die mögliche Erklärung, dass es trotzdem an der zufälligen Zuordnung liegen kann, ziehen sie also nicht in Betracht, sondern sie verweisen auf konfundierende Vari-

ablen, die Einfluss auf die Blutdruckwerte genommen haben könnten. Von der Interviewerin wird dies im folgenden Turn (395) als richtig eingestuft und erläutert. Die ursprünglich von ihr gesuchte Erklärung nennt sie im gleichen Turn selbst und erläutert diese auch.

Episode 15-5: Schlussfolgerungen aus einem P-Wert von 30% (395-402)

395	I: … Und jetzt darauf bezogen, wenn wir jetzt nicht einen P-Wert von einem Prozent gehabt hätten, sondern einen P-Wert von 30 Prozent, was würdet ihr dann schließen?
396	F: Dann kann man gar nichts sagen, eigentlich.
397	I: Kannst du das ausführen? Ist richtig.
398	L: Ja also man kann da zum Beispiel nicht sagen, dass man die Nullhypothese ablehnt, weil es ja zu einer recht großen Wahrscheinlichkeit von 30 Prozent zu einem solchen Ergebnis kommen kann. Aber man kann auch nicht sagen, dass es gar keinen Unterschied gibt. Weil man das nicht feststellen konnte, weil man nichts sicher sagen konnte quasi.
399	I: Mhm (bejahend). Noch was zu ergänzen?
400	L: Nö.
401	F: Nee, eigentlich nicht.
402	I: Ok. Gut. Dann sind wir inhaltlich durch, vielen Dank.

Die abschließende Frage der Interviewerin, was aus einem fiktiven P-Wert von 30 Prozent zu schließen sei (395) beantwortet Fabia mit „Dann kann man gar nichts sagen, eigentlich.“ (396), was eine korrekte Schlussfolgerung darstellt. Auf die Frage nach einer Ausführung dazu (397) antwortet nun Laura und führt zwei Möglichkeiten aus. Einerseits nennt sie, dass die Nullhypothese nicht abgelehnt werden könne, „… weil es ja zu einer recht großen Wahrscheinlichkeit von 30 Prozent zu einem solchen Ergebnis kommt…“ und andererseits, dass man auch nicht sagen könne, dass es keinen Unterschied [zwischen den beiden Diäten] gebe, weil man das nicht feststellen könne (398). Auch dies ist richtig und damit endet das Interview.

11.3.2 TinkerPlots-Benutzung zur Simulation von Fabia und Laura

Aus den vorherigen Analysen wird in diesem Abschnitt entlang den in Kapitel 10.4.4 vorgestellten Kategorien analysiert, welche der notwendigen Schritte in TinkerPlots in den Interaktionseinheiten 8, 9 und 10 von Fabia und Laura selbständig, mit leichter Intervention oder mit starker Intervention durch die Interviewerin durchgeführt werden.

Tabelle 35: Notwendige Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Fabia und Laura

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen der Zufallsmaschine	Ja	-	-
Vorhandene Kugeln der Zufallsmaschine löschen	Ja	-	-
Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren	-	-	Ja
Ein zweites Bauteil der Zufallsmaschine hinzufügen	Ja	-	-
Zweites Bauteil füllen	-	-	Ja
Die Einstellung ohne Zurücklegen bei beiden Bauteilen	-	Ja	-

auswählen			
Durchgangsanzahl festlegen	Ja	-	-

Beim Erstellen der Zufallsmaschine arbeiten Fabia und Laura weitgehend selbständig. Beim Einfügen der ursprünglichen Blutdruckwerte benötigen sie konkrete Anweisungen, ebenso dabei, das zweite Bauteil zu füllen, in welchem die neuen Gruppen abgebildet werden. Auf die Einstellung ohne Zurücklegen wird von der Interviewerin lediglich hingewiesen, so dass dies nur als leichte Intervention gewertet wird. Schwierigkeiten, die auftreten, handeln sie zunächst unter sich aus und kommen dadurch auch zu Lösungen.

Tabelle 36: Optionale Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Fabia und Laura

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einstellen der Geschwindigkeit	Ja	-	-
Umbenennen der Merkmale Merkm1 und Merkm2	Ja	-	-

Das Einstellen der Geschwindigkeit nehmen Fabia und Laura selbständig vor. Ein Umbenennen der Merkmale erfolgt nicht.

Tabelle 37: Notwendige Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Fabia und Laura

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen einer Graphik zur Simulation	Ja	-	-
Das erste Merkmal auswählen und darstellen	Ja	-	-
Das zweite Merkmal auswählen und in der Graphik passend darstellen	-	Ja	-
Die Mittelwerte einblenden	Ja	-	-
Differenz der Mittelwerte mit Hilfe der Linealfunktion messen	Ja	-	-
Messgröße definieren	-	-	Ja
Anzahl der zu sammelnden Messgrößen eintragen	Ja	-	-
Messgrößen sammeln	Ja	-	-
Erstellen einer Graphik zu den gesammelten Messgrößen	Ja	-	-
Referenzverteilung darstellen	Ja	-	-

Auch hinsichtlich des Darstellens der simulierten Ergebnisse und dem Ermitteln der Differenz der Mittelwerte der zwei neuen Gruppen arbeiten Fabia und Laura selbständig. Sie benötigen einen Hinweis, das zweite Merkmal in der Tabelle zu markieren, danach können sie selbständig weiterarbeiten. Vor allem die Verwendung der Linealfunktion, um die Differenz der Mittelwerte zu ermitteln, geschieht sehr zügig und ohne Diskussion in nur vier Turns. Bezüglich des Messgrößendefinierens wurden starke Interventionen durch die Interviewerin gemacht, allerdings ist nicht klar, ob Fabia und Laura dies nicht auch mit wesentlich weniger starken Interventionen hinbekommen hätten. Sonstige Interventionen sind nicht nötig.

Tabelle 38: Optionale Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Fabia und Laura

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Fälle in der Graphik stapeln	Ja	-	-
Beschleunigen des Messgrößensammelns	-	Ja	-

Das Beschleunigen des Messgrößensammelvorgangs wird von Laura und Fabia selbst initiiert, sie benötigen lediglich den Hinweis, wo der Reiter für das Einstellungsmenü der Messgrößen zu finden ist.

Tabelle 39: Notwendige Schritte in TinkerPlots zur Ermittlung des P-Werts von Fabia und Laura

Schritt	**Keine Intervention**	**Leichte Intervention**	**Starke Intervention**
Einteiler verwenden	-	Ja	-
Prozente einblenden	Ja	-	-

Am Ende verschiebt sich aufgrund eines unbemerkten Bedienfehlers der Einteiler, so dass zunächst ein P-Wert von 17% ermittelt wurde. Hierauf geschieht eine leichte Intervention durch die Interviewerin, die dazu führt, dass Fabia und Laura ihren Fehler sofort bemerken und korrigieren.

11.3.3 Fallzusammenfassung Fabia und Laura

Der Bearbeitungsprozess von Fabia und Laura wurde in den beiden vorigen Abschnitten ausführlich dargelegt und interpretiert. In diesem Abschnitt wird der allgemeine Eindruck zu diesem Paar geschildert und eine Fallzusammenfassung bezüglich besonderer Auffälligkeiten (Voigt 1984) gegeben.

Fabia und Laura arbeiten in allen Interaktionseinheiten recht zügig und gehen kaum über das ihrer Meinung nach direkt Geforderte, beispielsweise durch die im Randomisierungstestplan vorgegebenen Stichworte, hinaus. Beim Arbeiten mit TinkerPlots fällt auf, dass sie wenig Hilfe benötigen und viel selbständig arbeiten, weshalb die Interaktionseinheiten 8, 9 und 10 nicht besonders lang dauern. Die Schlussfolgerungen aus dem erhaltenen P-Wert werden nur sehr knapp und rudimentär gezogen, weshalb insbesondere die Einheit 11 sehr kurz ausfällt. Dem wird in der letzten Einheit versucht Rechnung zu tragen, in welcher hierzu alle Fragen des Leitfadens diskutiert werden, jedoch hat hier die Interviewerin einen hohen Redeanteil, wie in der Analyse zu sehen war.

Eine Schwierigkeit besteht für dieses Paar darin, die geforderten Sachverhalte auf Statistik- und Kontextebene richtig zu formulieren. Sie verwenden viel Zeit darauf, gemeinsam Formulierungen zu suchen, die jedoch auch nach der gemeinsamen Aushandlung häufig noch holprig klingen. Eine passende Sprache scheint eine Hürde für die Beiden darzustellen, insbesondere die Verbindung der verschiedenen Ebenen.

Häufig ist es Fabia, die die Stichworte des Schemas aufnimmt oder vorliest und somit eine neue Gesprächsepisode eröffnet. Dadurch übergibt sie den Turn häufig an Laura, von der sie eine Antwort oder Formulierung erwartet, die diese in der Regel liefert. Dies kann als Muster gedeutet werden, nach welchem dieses Paar bereits in der Vergangenheit gearbeitet hat, da es sich durch fast alle Interaktionseinheiten durchzieht. Laura liefert die ersten Formulierungsvorschläge, die von Fabia bestätigt oder ergänzt werden.

Das Konzept der zufälligen Aufteilung der Versuchspersonen auf die Gruppen scheint für das Paar zunächst eine Schwierigkeit darzustellen, die sich im Laufe des Bearbeitungsprozesses jedoch abbaut. Zu Beginn des Interviews haben sie in der zweiten Interaktionseinheit große Schwierigkeiten, die zufällige Aufteilung zu erklären. Sie nehmen die Perspektive von Versuchspersonen ein, scheinen hier jedoch die Fehlvorstellung zu haben, dass diese zufällig ausgewählt wurden. Mit Hilfe der Interviewerin arbeiten sie als Begründung für die zufällige Aufteilung die daraus resultierende Gleichverteilung relevanter Merkmale in den beiden Gruppen heraus. Sie können dies jedoch zu Beginn nur ungenau formulieren und liefern auch keine eigenen Beispiele relevanter Merkmale. Beim Formulieren möglicher Erklärungen für die beobachteten Unterschiede in der dritten Interaktionseinheit nennen sie zwar die zufällige Aufteilung als eine solche Erklärung, hierbei scheint es sich jedoch um ein rezeptartiges Aufsagen von in der Veranstaltung gelernten Formulierungen zu handeln. Sie nennen dies selbst als Schwierigkeit. Im Gespräch zur Nullhypothese wird das Konzept nicht adressiert, allerdings wird später im Interview deutlich, dass sie die zufällige Aufteilung als Erklärung der Nullhypothese ansehen. Nachdem die selbständige Durchführung des Randomisierungstests abgeschlossen ist und im dritten Teil wieder Interviewfragen gestellt werden, entsteht der Eindruck, dass nach dem Bearbeitungsprozess das Konzept der zufälligen Aufteilung wesentlich besser verstanden ist und im Gegensatz zum Beginn des Interviews auch korrekt und ausführlich (Episode 13-5) erklärt werden kann. Scheinbar war diese Durchführung nötig, um das Konzept, das möglicherweise im Vorfeld nicht komplett durchdrungen war, besser verstehen zu können. Während der Simulation nehmen Fabia und Laura Bezug auf die händische Randomisierung, die im Seminar durchgeführt wurde und auch die Erinnerung daran hat möglicherweise entscheidend zum Verstehen des Konzepts der zufälligen Aufteilung beigetragen.

Bezüglich der Null- und Forschungshypothese benötigen Fabia und Laura mehrere Formulierungsanläufe, ohne dass sie jedoch prinzipielle Schwierigkeiten mit diesem Konzept zu haben scheinen. Sie verwechseln die beiden Hypothesen nicht (wie beispielsweise als Schwierigkeit von Vallecillos (1999) berichtet), allerdings adressieren sie in der Nullhypothese nicht die zufällige Aufteilung, was die von ihnen aufgestellte Nullhypothese nicht falsch macht sondern eher zeigt, dass wie gerade beschrieben das Konzept der zufälligen Aufteilung eine Schwierigkeit darstellt.

Schlussfolgerungen werden von ihnen nur knapp gezogen. Mit Verweis auf ein signifikantes Ergebnis wird die Nullhypothese abgelehnt und die Forschungshypothese angenommen. Dies ist ein korrektes Vorgehen bei dem erhaltenen P-Wert von einem Prozent, gleichzeitig formulieren Fabia und Laura Kritik am Design des Experiments im Hinblick auf die Auswahl der Versuchspersonen und die geringe Stichprobengröße. Im letzten Teil des Interviews wird deutlich, dass jedoch bezüglich des P-Werts bei diesem Paar eine typische Fehlvorstellung zum P-Wert vorliegt, denn sie interpretieren den P-Wert irrtümlich als Maß für die Wahrscheinlichkeit der Nullhypothese (365). Ein Bezug zum Kontext wird von Fabia und Laura nicht selbständig vollzogen, auch auf entsprechende Nachfragen durch die Interviewerin im letzten Teil des Interviews reagieren die Beiden auf statistischer Ebene, indem sie auf das Ablehnen der Nullhypothese verweisen aufgrund des kleinen P-Werts. Schlussfolgerungen über die Wirksamkeit des Fischöls werden von den Beiden nicht gezogen, möglicherweise liegt das daran, dass sie auch hier auf statistischer Ebene einem Rezept folgen (bei einem P-Wert kleiner als 10 Prozent kann die Nullhypothese abgelehnt werden), das sie nicht auf den Kontext übertragen können.

11.4 Fallstudie Mandy und Alisa

In diesem Abschnitt werden die zusammenfassenden Interpretationen der 15 Interaktionseinheiten von Mandy und Alisa präsentiert. Das Vorgehen ist dabei analog zu dem der Analyse von Rebecca und Selina und Laura und Fabia.

11.4.1 Zusammenfassende Interpretationen der 15 Interaktionseinheiten von Mandy und Alisa

Interaktionseinheit 1: Vorstellung und Vorlesen des Aufgabenblatts (Z. 1-3)

In der ersten Interaktionseinheit wird festgestellt, dass die beiden Teilnehmerinnen, Mandy und Alisa, bereits im Vorfeld miteinander vertraut waren und häufiger ihre Hausaufgaben miteinander besprochen, allerdings nicht gemeinsam gelöst haben. Die Interviewerin liest das Aufgabenblatt komplett vor und daran schließt sich die zweite Interaktionseinheit an.

Interaktionseinheit 2: Begründung für die zufällige Aufteilung auf die Gruppen (Z. 4-7)

3 I: […] Meine erste Frage ist: Wieso wurden die Versuchspersonen, die hier teilgenommen haben, zufällig auf die zwei Gruppen aufgeteilt?

4 M: Naja also damit man kein anderes Merkmal hat um zu sagen, wenn ich die jetzt nach der Haarfarbe aufteilen würde. Dass ich sagen könnte, ja es liegt irgendwie an der Haarfarbe. Dass man mit Öldiäten generell vielleicht besser den Blutdruck senken kann. Also so würde ich die zufällige Zuordnung erklären.

5 A: Ja ich hätte es so ähnlich gesagt, dass man kein anderes Merkmal irgendwie bewusst hat.

6 M: Mhm (bejahend). Genau oder wenn man des jetzt hätte, Frauen reagieren besser darauf. Dann

	wüsste man ja nicht woran es tatsächlich liegt. Zum Beispiel ich weiß jetzt nicht, ob Haarfarbe/.
7	A: / Ja aber so. Um das in Worte zu fassen, was in meinem Kopf war. (…)

Die zweite Interaktionseinheit beginnt mit der Frage der Interviewerin am Ende ihres Turns (3) nach der Begründung, wieso die Versuchspersonen zufällig auf die zwei Gruppen aufgeteilt wurden. Mandy ergreift hierauf den Turn (4) und antwortet mit Bezug zum Kontext. Sie beginnt mit einer allgemeinen Erklärung („… damit man kein anderes Merkmal hat…“) und führt dies dann mit einem Beispiel aus: „…, wenn ich die jetzt nach der Haarfarbe aufteilen würde. Dass ich sagen könnte, ja es liegt irgendwie an der Haarfarbe….“. Danach fährt sie fort: „Dass man mit Öldiäten generell vielleicht besser den Blutdruck senken kann…“. Alisa bekräftigt diese Aussage im darauffolgenden Turn (5) und stellt den Aspekt eines allgemeinen Merkmals noch einmal heraus und fügt an, dass „man kein anderes Merkmal irgendwie bewusst hat.“ Interessant ist hier das Anfügen des Adverbs „bewusst“, das darauf hinweist, dass sie möglicherweise die Vorstellung hat, dass zwischen „bewussten“ und „unbewussten“ Merkmalen unterschieden werden muss. Da keine Nachfrage durch die Interviewerin hierzu gestellt wurde, lässt sich nachträglich nur vage interpretieren, was sie hiermit gemeint haben könnte. Beispielsweise kann sie die Vorstellung haben, dass sichtbare Merkmale, wie zum Beispiel die von Mandy genannte Haarfarbe oder das im nächsten Turn von Mandy genannte Merkmal Geschlecht eine bewusste Einteilung der Gruppen nach diesen Merkmalen zulässt. Ein unbewusstes Merkmal wäre nach dieser Logik das, was sich nicht messen oder kategorisieren lässt, wie zum Beispiel die individuelle Reaktion einer Versuchsperson auf einen bestimmten Wirkstoff. Aufgeschlüsselt nach dem Toulminschen Argumentationsschema gehen Mandy und Alisa davon aus, dass eine zufällige Aufteilung der Versuchspersonen auf die Gruppen erfolgt ist und gelangen durch die Regel, dass andere Merkmale den Blutdruck beeinflussen können über die nächste Regel, dass durch die zufällige Aufteilung andere Merkmale als die durchgeführte Öldiät als Erklärung ausgeschlossen werden zu der Konklusion, dass die Öldiät das einzige Merkmal bleibt, das den Blutdruck beeinflusst habe. Diese Argumentation ist in Abbildung 209 dargestellt.

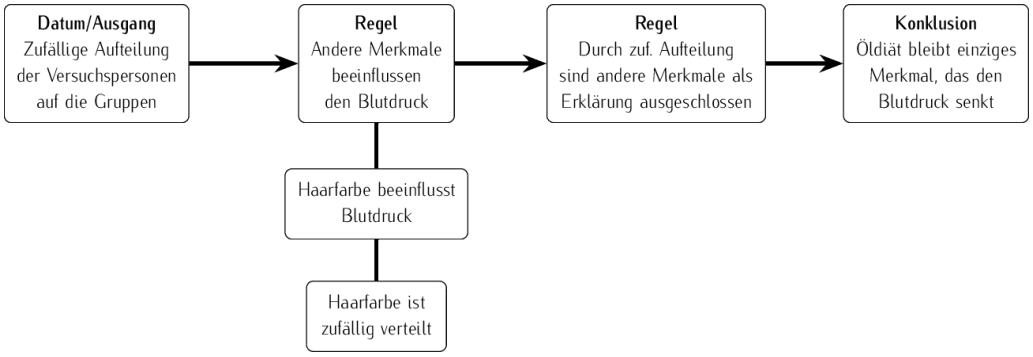


Abbildung 209: Rekonstruierte Argumentation von Mandy und Alisa zur zufälligen Aufteilung der Versuchspersonen auf die Gruppen

Interaktionseinheit 3: Mögliche Erklärungen für die beobachteten Unterschiede (Z. 8-19)

8	I: Ja. Gut. (..) Dann ist die Frage, wenn wir uns diesen Unterschied von 7,7 angucken zwischen den Mittelwerten beider Gruppen. Welche möglichen Erklärungen gibt es für diese beobachteten Unterschiede?
9	A: Also einmal zum Beispiel, dass Fischöl wirklich besser anschlägt als das normale Öl. Also dass deshalb die Blutdrucksenkung beim Fischöl höher war als beim normalen Öl also vom Durchschnitt her.
10	M: Ja also. Das sollte es zumindest aussagen, denke ich.
11	I: Fallen euch noch andere mögliche Erklärungen ein, woran es liegen könnte?
12	M: Ja also es wäre grundsätzlich natürlich auch möglich, dass jetzt einfach zufälligerweise die Personen in der Fischölgruppe besser darauf reagieren oder generell empfänglicher für Öl Diäten sind. Aber im Prinzip ist es wahrscheinlicher, dass Fischöl tatsächlich erfolgreicher ist als normales Öl.
13	A: Ja gerade weil der Unterschied so hoch ist, hätte ich da auch darauf geschlossen.
14	I: Wenn du sagst, dass die vielleicht empfänglicher sind. Das ist so ein bisschen so etwas, das man nicht kontrollieren kann als Forscher. Also meinst du damit, dass das quasi etwas ist, das die im Vorfeld nicht kontrolliert haben und das da halt lauter Personen in einer Gruppe sind/.
15	M: Ja genau. Also dass die zufällige Einteilung unglücklicherweise nicht so funktioniert hat, sondern dass zufälligerweise genau die Leute in der Gruppe waren in der es eh besser funktioniert. Und wenn man die Gruppen tauschen würde, dass es dann anders aussehen würde.
16	I: Mhm (bejahend). Du meinst, wenn man die Teilnehmer zwischen den Gruppen austauschen würde?
17	M: Oder wenn man jetzt die komplette normale Öl gegen die Fischöl auswechseln würde. Das hatte ich jetzt gemeint. Nicht einzeln.
18	I: Mhm (bejahend). Ok. Und das wäre als mögliche Erklärung tatsächlich noch diese zufällige Zuordnung. Dass wenn man verschiedene Stichproben hat, dass man eben Schwankungen darin hat.
19	M: Ja. Ja also das meinte ich damit so ein bisschen. Also dass es einfach an der Auswahl der Teilnehmer liegt.

Die nächste Interaktionseinheit wird eingeleitet durch die Frage der Interviewerin (8) nach möglichen Erklärungen für die beobachteten Unterschiede zwischen den Mittelwerten beider Gruppen. Diesmal ergreift zuerst Alisa den Turn und nennt als erste Erklärung mit Bezug zum Kontext, „…dass Fischöl wirklich besser anschlägt als das normale Öl…“ (9) und erläutert hierzu weiter, indem sie auf Statistikebene die ganze Verteilung anhand der Mittelwerte in den Blick nimmt: „…Also dass deshalb die Blutdrucksenkung beim Fischöl höher war als beim normalen Öl also vom Durchschnitt her.“ Dies wird von Mandy im von ihr als nächstes übernommenen Turn (10) als Erklärung akzeptiert, allerdings wird keine weitere Erklärung genannt, weshalb nun die Interviewerin erneut den Turn übernimmt und fragt, ob ihnen noch andere mögliche Erklärungen einfielen (11). Diesmal übernimmt Mandy den Turn und antwortet: „Ja also es wäre grundsätzlich natürlich auch möglich, dass jetzt einfach zufälligerweise die Personen in der Fischölgruppe besser darauf reagieren oder generell empfänglicher für Öl Diäten sind. Aber im Prinzip ist es wahrscheinlicher, dass Fischöl tatsächlich erfolgreicher ist, als normales Öl.“ (12). Implizit nennt sie in ihrer Antwort die natürlich vorkommende Schwankung mit der Personen auf Öl reagieren, und führt die beobachteten

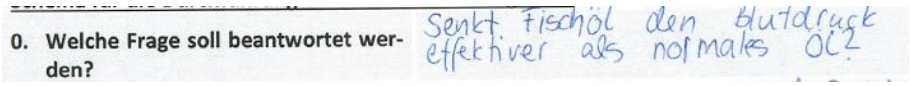
Unterschiede darauf zurück, dass nun „zufälligerweise“ diejenigen, die besser auf Öl-Diäten reagieren, in der Fischöl-Gruppe gelandet seien. Der zweite Satz ihrer Antwort macht jedoch deutlich, dass sie die zuvor genannte Erklärung für wahrscheinlicher hält. In Verbindung mit der vorigen Interaktionseinheit könnte dies ein Beispiel für Merkmale sein, nach denen nicht „bewusst“ gruppiert werden kann. Alisa bestätigt Mandys letzten Turn und nennt die Höhe des beobachteten Unterschieds als Begründung für die erste Erklärung (13). Im Folgenden wird durch die Interviewerin weiter nachgefragt, was mit „empfänglicher“ gemeint sei (14). Mandy übernimmt den Turn und antwortet (15). In ihrer Antwort geht Mandy davon aus, dass kein Unterschied zwischen der Wirkung der beiden Öldiäten herrscht, dass aber nun in der Fischölgruppe Personen gelandet sind, bei denen eine Senkung des Blutdrucks zufällig höher war als bei denen der anderen Gruppe. Sie formuliert dies jedoch in anderen Worten, und ob dies ihr selbst klar ist, dass hinter ihrer Aussage diese Erklärung steht, bleibt zweifelhaft. Auch die Aussage am Ende des Turns (15) „… Und wenn man die Gruppen tauschen würde, dass es dann anders aussehen würde“ ist ein Indiz, das die vorherige Interpretation stützt. So interpretiert dies auch die Interviewerin, denn sie fasst den Gesprächsverlauf zusammen wie folgt: „…Und das wäre als mögliche Erklärung tatsächlich noch diese zufällige Zuordnung. Dass wenn man verschiedene Stichproben hat, dass man eben Schwankungen darin hat.“ (17). Hinter dieser Erklärung stehen jedoch beliebig austauschbare Stichproben, nicht nur komplette Gruppentausche, wie es vorher von Mandy (16) genannt wurde. Dies wird jedoch nicht im Gespräch geklärt. Mandy akzeptiert und bestätigt jedoch die von der Interviewerin gelieferte Interpretation und bekräftigt, „… dass es einfach an der Auswahl der Teilnehmer liegt.“ (19). Somit stellt sie implizit einen Bezug zur Teilnehmerauswahl her, dass diese nämlich nicht zufällig ausgewählt wurden. Erneut ist es jedoch schwierig zu interpretieren, ob ihr dieser Sachverhalt klar ist.

Interaktionseinheit 4: Überleitung zum Randomisierungstest (Z. 20-22)

In der darauf folgenden kurzen Interaktionseinheit wird durch die Interviewerin übergeleitet zum nun durchzuführenden Randomisierungstest. Auf die Nachfrage von Mandy, ob sie den Randomisierungstest händisch machen oder mit TinkerPlots durchführen sollen (21) gibt die Interviewerin die Antwort, dass sie dies machen dürften, wie sie möchten (22). Die Interviewerin macht also keine Vorgabe, wie an den Randomisierungstest herangegangen werden soll.

Interaktionseinheit 5: Forschungsfrage notieren (Z. 23-28)

23 M: Ok. Welche Frage soll beantwortet werden?
24 A: Ist Fischöl oder bringt Fischöl mehr oder senkt Fischöl den Blutdruck/.
25 M: /Effektiver als/.
26 A: /Normales Öl?
27 *Die Studentinnen notieren:*

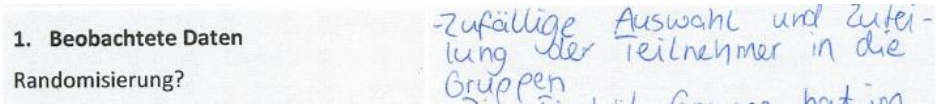


28 M: Ok.

Mandy und Alisa beginnen direkt mit dem Vorlesen des ersten Schritts des Randomisierungstestschemas in (23). Alisa schlägt im nächsten Turn verschiedene Formulierungen einer Frage im Kontext vor mit den Worten „Ist Fischöl oder bringt Fischöl mehr oder senkt Fischöl den Blutdruck.“ (24). Abwechselnd ergänzen sie Bruchstücke der Formulierung und notieren als Forschungsfrage ohne weiter darüber zu sprechen: „Senkt Fischöl den Blutdruck effektiver als normales Öl“ (27). Damit hat sie eine gute Formulierung der Forschungsfrage gefunden.

Interaktionseinheit 6: Beobachtete Daten (Z. 29-43)

29 A: Randomisierung.
30 M: Müssten wir machen.
31 I: Ihr seid hier noch bei den beobachteten Daten. Also hier geht es noch darum, was in dem Experiment berichtet wurde.
32 M: Achso also ob hier erst mal eine Randomisierung stattfand?
33 I: Genau.
34 A: Ja wenn das eine zufällige Zuordnung ist, dann ja.
35 M: Also zufällige Auswahl der männlichen Teilnehmer. Das schreiben wir erst mal. (...) (unv.) Zuteilung vielleicht.
36 A: Ja. In die Gruppen auch.
37 *Es wird notiert:*



38 A: Und welcher Unterschied?
39 M: Die Fischölgruppe schneidet tendenziell besser ab, als die normale Öl Gruppe.
40 A: Ja oder hat im Durchschnitt eine höhere Blutdrucksenkung.
41 M: Ist Durchschnitt, nee?
42 A: Ja also arithmetisches.
43 *Es wird notiert:*

Die sechste Interaktionseinheit eröffnet diesmal Alisa mit dem Stichwort „Randomisierung“ des Randomisierungstestschemas (29), woraufhin Mandy den Turn übernimmt mit den Worten „Müssten wir machen.“ (30). Dies kann interpretiert werden, dass sie dieses Stichwort als Aufforderung ansieht, nun eine Randomisierung der Werte vorzunehmen. Dies ist jedoch eine falsche Vorstellung von Mandy, denn das Stichwort steht unter dem Punkt *1) Beobachtete Daten* des Schemas und bezieht sich darauf, ob für das Design des Experiments die Gruppen randomisiert wurden. Deshalb greift an dieser Stelle auch die Interviewerin ein und weist daraufhin, dass dies bezüglich der beobachteten Daten aufgefasst werden solle und sagt mit Bezug zum Kontext „… Also hier geht es noch darum, was in dem Experiment berichtet wurde.“ (31) um weitere Missver-

ständnisse auszuschließen und sicher zu stellen, dass die beiden Studierenden sich mit den beobachteten Daten des Experiments auseinander setzen. Daraufhin fragt Mandy nach: „Achso also ob hier erst mal eine Randomisierung stattfand?“ (32). Dies kann als Frage auf zwei Ebenen interpretiert werden. Entweder meint sie mit der Frage, ob sie dies nur notieren solle, da für sie möglicherweise völlig klar war durch die Aufgabenstellung oder das bisherige Gespräch, dass die Gruppen des Experiments randomisiert wurden. Oder sie stellt die Frage auf einer inhaltlichen Ebene, weil ihr eben nicht klar ist, dass die Gruppen randomisiert wurden. In diesem Sinne ist die Antwort „Genau.“ (33) der Interviewerin nicht hilfreich, allerdings übernimmt danach Alisa den Turn und beantwortet die von Mandy gestellte Frage (32) auf der inhaltlichen Ebene mit „Ja wenn das eine zufällige Zuordnung ist, dann ja.“ (34). Sie erkennt demnach, dass die Versuchspersonen zufällig auf die Gruppen aufgeteilt wurden und ordnet dies dem Stichwort Randomisierung zu. Mandy schlägt daraufhin eine Formulierung vor für das Ausfüllen des Schemas (35), die von Alisa ergänzt wird (36) und schließlich notiert wird als „Zufällige Auswahl und Zuteilung der Teilnehmer in die Gruppen“ (37). Dies ist jedoch nur teilweise richtig, denn im Aufgabentext steht lediglich, dass es sich um Freiwillige handelt, nicht aber, dass diese zufällig ausgewählt wurden. Dieser Fehler ist jedoch weder der Interviewerin noch den beiden Teilnehmerinnen aufgefallen und wird auch später bei den Schlussfolgerungen keine Rolle spielen.

Nach dem Ausfüllen des Schemas fragt Alisa „Und welcher Unterschied?“ (38), womit sie entweder nur unvollständig die Frage des Randomisierungstestschemas vorliest (diese lautet: *„Welcher Unterschied lässt sich feststellen?“*) und somit eine Turnübergabe an Mandy darstellen würde. Oder sie stellt hiermit ihre Unsicherheit dar, auf welchen Unterschied die Frage abzielt. Mandy antwortet direkt inhaltlich „Die Fischölgruppe schneidet tendenziell besser ab als die normale Ölgruppe.“ (39), was eine qualitative Bewertung darstellt und erneut die gesamte Verteilung in den Blick nimmt. Alisa ergreift direkt den Turn im Anschluss von Mandys Aussage und ergänzt „Ja oder hat im Durchschnitt eine höhere Blutdrucksenkung.“ (40), womit sie Mandys Aussage auf den Durchschnitt präzisiert. Quantifiziert wird der Unterschied jedoch von keiner der Beiden. Am Ende der Interaktionseinheit notieren sie „Die Fischöl-Gruppe hat im Durchschnitt eine höhere Blutdruckreduktion als die normale Öl“ (43), was durch die Angabe der Differenz von 7,7 zwischen den beiden Mittelwerten hätte konkretisiert werden können, durch das Weglassen jedoch nicht falsch ist.

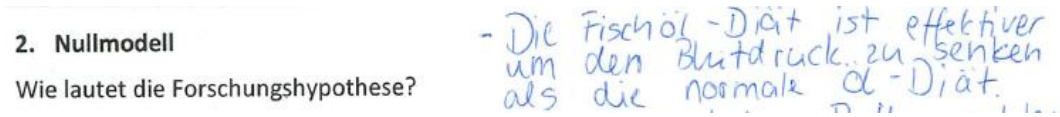
Interaktionseinheit 7: Nullmodell (Z. 44-67)

Null- und Forschungshypothese werden in der siebten Interaktionseinheit aufgestellt.

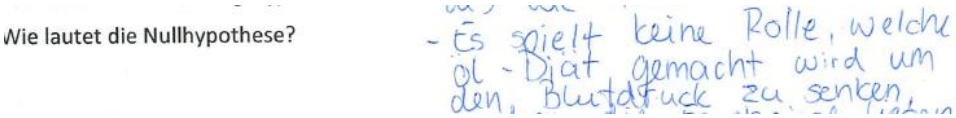
---

44 A: Also die Forschungshypothese ist dann, dass die Fischöldiät eigentlich wie hier oben effektiver ist, um Blutdruck zu senken als eine normale Öl Diät.

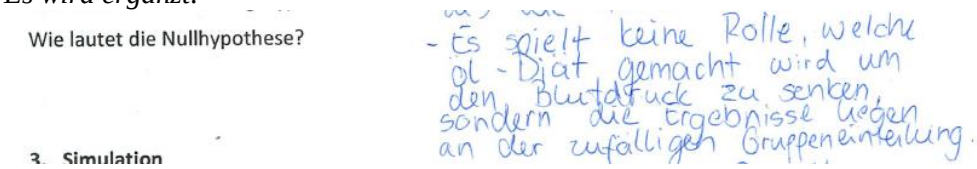
45 *Es wird notiert:*



46 A: Die Nullhypothese ist ja dann, dass es nicht wichtig ist, welches Öl es ist, oder dass es keinen Unterschied macht. Sondern dass es lediglich der zufälligen/.
47 M:/ Zuteilung geschuldet ist quasi.
48 A: Ja.
49 M: Also es spielt keine Rolle/.
50 A:/ Ja die Art des Öles.
51 M: Es spielt keine Rolle, welche Art der Öl-Diät gemacht wird/.
52 A: /Ja/.
53 M: /Um den Blutdruck zu senken.
54 A: Ja.
55 M: Sollen wir noch hinzufügen, dass: Es spielt keine Rolle bla bla bla, sondern liegt an der zufälligen Einteilung, oder?
56 *Es wird notiert:*



57 A: (unv.) mehr passt jetzt da nicht rein.
58 M: Ist gut so.
59 A: Ok. Die Simulation/.
60 I: /Nehmt das noch dazu.
61 A: Ok. Sondern/.
62 M:/ Das Ergebnis liegt an der zufälligen/.
63 A: / Oder die Ergebnisse?
64 M: Meinetwegen die Ergebnisse. Sondern die Ergebnisse/.
65 A:/ Liegen an der zufälligen Einteilung. An der zufälligen Gruppeneinteilung (...) der Teilnehmer, oder reicht das?
66 *Es wird ergänzt:*



67 M: Reicht.

Nur mit der kurzen Überleitung zur Statistikebene „Also die Forschungshypothese ist dann…“ (44) wendet sich Alisa in der nächsten Interaktionseinheit direkt der Formulierung der Forschungshypothese zu und nennt mit Bezug zum Kontext „… dass die Fischöldiät eigentlich wie hier oben effektiver ist, um Blutdruck zu senken als eine normale Öl-Diät.“ Das wird ohne weiteren Kommentar direkt im Schema notiert (45), woraufhin sich Alisa im nächsten Turn der Nullhypothese zuwendet. Die Formulierung der Nullhypothese dauert nun deutlich länger als die der Forschungshypothese. Alisa bietet Formulierungsalternativen an: „Die Nullhypothese ist ja dann, dass es nicht wichtig ist, welches Öl es ist, oder dass es keinen Unterschied macht. Sondern dass es lediglich der zufälligen.“ (46). Hierauf fällt ihr Mandy ins Wort und zeigt damit, dass sie ihrem Ge-

danken komplett folgt, denn sie vervollständigt Alisas Formulierung mit „Zuteilung geschuldet ist quasi.“ (47). Alisa bestätigt dies (48) und anschließend handeln die beiden weiter die Formulierung der Nullhypothese aus und nähern sich mit „es spielt keine Rolle“ (Mandy, 49) „Ja die Art des Öls“ (Alisa, 50), zusammengefasst von Mandy durch „Es spielt keine Rolle, welche Art der Öldiät gemacht wird.“ (51) und noch einmal ergänzt durch die Wendung „Um den Blutdruck zu senken“ von Mandy (53). Als Formulierungsalternativen für die Nullhypothese wurden in dieser Episode genannt:

- Es ist nicht wichtig, …
- Es macht keinen Unterschied, …
- Es spielt keine Rolle, …

Diese sind alle sprachlich als gleichwertig einzuschätzen. Letztlich notieren Mandy und Alisa „Es spielt keine Rolle, welche Öl-Diät gemacht wird um den Blutdruck zu senken.“ (56), was sie eigentlich noch ergänzen wollten wie von Mandy vorgeschlagen durch „… sondern liegt an der zufälligen Einteilung“ (55), nun aber aus Platzgründen weggelassen werden soll (57). Somit scheint Alisa diesen Zusatz als nicht so wichtig einzustufen, wozu Mandy zustimmt mit der Bewertung des Aufgeschriebenen als „Ist gut so.“ (58) und Alisa überleiten möchte zur Simulation (59), aber von der Interviewerin unterbrochen wird mit „Nehmt das noch dazu.“ (60), womit nur der Zusatz der zufälligen Zuordnung gemeint sein kann. Ohne dies in Frage zu stellen, formulieren Mandy und Alisa hinzu „…sondern die Ergebnisse liegen an der zufälligen Gruppeneinteilung.“ (61-66). Mandys „Reicht.“ (67) macht deutlich, dass die nun gefundene Formulierung von ihr als abgeschlossen betrachtet wird, womit sie endgültig die nächste Interaktionseinheit einläutet.

Interaktionseinheit 8: Simulieren/Zufallsmaschine erstellen (Z. 68-227)

In der achten Interaktionseinheit wird simuliert. Mandy und Alisa wählen hierzu zunächst den Zugang über TinkerPlots, wechseln dann jedoch schnell zur händischen Simulation. Erst nach einer händischen Randomisierung der Werte gehen sie zu TinkerPlots über, für dessen Bedienung sie auf die Hilfe der Interviewerin angewiesen sind. Die Interaktionseinheit wird in neun Episoden gegliedert. In der ersten Episode planen die Beiden ihr Vorgehen und starten einen ersten Versuch mit TinkerPlots zu arbeiten. In der zweiten Episode führen sie nach einem Input durch die Interviewerin die Simulation einmal händisch durch, wodurch dies eine sehr lange Episode ist. Der gedankliche Transfer der händischen Simulation zu TinkerPlots findet in der dritten Episode statt, konkret wird die Zufallsmaschine in der vierten Episode erstellt. Die fünfte Episode behandelt das Umbenennen der Merkmale. In der sechsten Episode wird die Einstellung ohne Zurücklegen vorgenommen und in der siebten Episode wird die Anzahl der nötigen Durchgänge bestimmt. In der achten Episode validieren Mandy und Alisa kurz die

von ihnen erstellte Zufallsmaschine und in der neunten Episode wird das weitere Vorgehen geplant und das Schema ausgefüllt.

Episode 8-1: Planen des Vorgehens (Z. 68-82)

68 A: Dann Simulation.
69 M: Ja. Wie wollen wir randomisieren?
70 A: Ja. Also erst mal ist das doch so, dass man die Werte, die man hat, mischt/.
71 M:/ Und dann wieder in neue Gruppen sortiert.
72 A: Genau und dann davon das/.
73 M:/ Den Durchschnitt in dem Fall.
74 A: Ja, oder wenn man es mit den Zettelchen macht den Median.
75 M: Ja ich würde das eigentlich lieber mit TinkerPlots machen, aber ich bin mir nicht mehr sicher, wie die einzelnen Schritte waren, aber das können wir ja. Können wir dann ja schauen. (…)
76 I: Ihr fangt einfach an mit dem, was ihr könnt. Ihr braucht erst mal eine Zufallsmaschine.
77 A: Stimmt.
78 *Die Studentinnen öffnen als erstes eine Zufallsmaschine.*

79 I: Und dann beschreibt ihr mir am besten was ihr machen wollt, dann helfe ich euch.
80 A: Merkmale müssen ja dann (...)/.
81 M: Ich bin total verwirrt gerade.
82 A:/ Warte wie war das nochmal?

Diese Interaktionseinheit wird im Vergleich zu den anderen deutlich eingeleitet. Alisa sagt „Dann Simulation“ (68) und übergibt damit den Turn an Mandy, da sie nichts weiter dazu sagt. Mandy greift dies auf mit „Ja. Wie wollen wir randomisieren?“ (69) und macht damit auch keinen Vorschlag bezüglich des weiteren Vorgehens, sondern gibt den Turn wieder zurück an Alisa. Sie greift dabei vermutlich den Beginn des Interviews auf, in welchem geklärt wurde, dass sie sowohl händisch als auch mit Unterstützung durch TinkerPlots simulieren dürfen (21-22). Es ist nun an Alisa, die Frage zu beantworten, sie weicht jedoch aus und führt erst einmal aus, welches ihre Ausgangslage ist: „Ja. Also erst mal ist das doch so, dass man die Werte, die man hat, mischt.“ (70). Hinter dieser Aussage steckt das Gedankengebilde, dass alle Werte, die während des Experiments aufgetreten sind, unabhängig von der Gruppenzugehörigkeit zusammengenommen und durchmischt werden, als Bild dafür, dass sie in eine zufällige Reihenfolge gebracht werden. Zwar wird die Unabhängigkeit von der Gruppe, und damit vom eingenommenen Öl, nicht ausgesprochen, aber dies wurde während des Bausteins Inferenzstatistik ausführlich thematisiert und scheint im Hintergrund präsent zu sein. Es kann jedoch auch einfach nur die Erinnerung an das prozessorientierte Vorgehen sein, wel-

ches im Seminar durchgeführt wurde, was durch die spätere Referenz auf die Zettel in (74) von Alisa als Interpretation gestützt würde. Mandy folgt Alisas Gedankengang und führt den nächsten Schritt nach dem Durchmischen aller Werte aus mit „Und dann wieder in neue Gruppen sortiert.“ (71), wobei sie mit „sortiert“ sicherlich nicht eine bewusste Zuteilung der Werte nach einem bestimmten Kriterium meint, sondern schlicht ein anderes Wort für *zuordnen* verwendet. Einander ergänzend planen sie weiter, indem Alisa in (72) beginnt mit „Genau und dann davon das“ und Mandy ergänzt „Den Durchschnitt in dem Fall.“ (73). Woraufhin wieder Alisa verdeutlicht und Bezug nimmt auf das händische Randomisieren „Ja oder wenn man es mit den Zettelchen macht den Median“ (74), wobei es jetzt keinen Grund gäbe, bei der händischen Simulation nicht das arithmetische Mittel zu verwenden, außer dass im Seminar beispielhaft der Median genutzt wurde, weil sich dieser schneller bestimmen lässt (vgl. Kapitel 8.2.2.2). Durch die „Zettelchen“ hat jetzt Alisa zum ersten Mal die händische Simulation ins Spiel gebracht. Mandy nimmt dieses Stichwort direkt auf und entgegnet, dass sie es lieber mit TinkerPlots machen würde (75), fügt jedoch gleich hinzu, dass sie sich bezüglich der einzelnen Schritte nicht mehr sicher sei. Trotzdem scheint es der von ihr favorisierte Weg zu sein, denn sie fügt noch an „… Können wir dann ja schauen.“, vermutlich in dem Bewusstsein, dass zuvor von der Interviewerin Hilfe bezüglich der technischen Umsetzung angeboten wurde. Sie liefert allerdings keine Begründung, wieso sie lieber TinkerPlots nutzen würde. Beide vermeiden also zunächst die Antwort, ob sie eine händische Simulation durchführen oder mit TinkerPlots arbeiten möchten, indem sie sich den Prozess wieder ins Gedächtnis rufen. Erst in (75) äußert Mandy ihre Meinung hierzu.

Nach einer kurzen Pause nach Mandys Turn meint die Interviewerin: „Ihr fangt einfach an mit dem, was ihr könnt…“ (76) und gibt die konkrete Starthilfe auf Softwareebene: „… Ihr braucht erst mal eine Zufallsmaschine.“ (76). Alisas darauffolgendes „Stimmt.“ (77) lässt sich so deuten, dass sie mit dem Vorgehen einverstanden ist, erst einmal zu beginnen und dann zu sehen, wie weit sie kommen, oder dass sie dieses Vorgehen sogar selbst vorgeschlagen hätte. Sie nimmt es als Startsignal, denn daraufhin zieht sie eine Zufallsmaschine in die Arbeitsfläche von TinkerPlots (78). Da hieraufhin jedoch nichts weiter geschieht, übernimmt erneut die Interviewerin den Turn und sagt: „Und dann beschreibt ihr mir am besten was ihr machen wollt, dann helfe ich euch.“ (79). Sie hat gemerkt, dass wohl große Unsicherheit bezüglich der Bedienung der Zufallsmaschine herrscht. Um die Schritte nicht vorzugeben, sollen nun die Studierenden beschreiben, was getan werden soll. Vielleicht mit Blick auf die Beschriftung innerhalb der Zufallsmaschine beginnt Alisa mit einem etwas ziellosen „Merkmale müssen ja dann“ (80), woraufhin erneut eine Pause eintritt. Mandy äußert sich nach der Pause mit „Ich bin total verwirrt gerade.“ (81). Alisa fragt daraufhin „Warte wie war das nochmal?“ (82) und zeigt damit, dass sie ebenfalls nicht weiß, was zu tun ist.

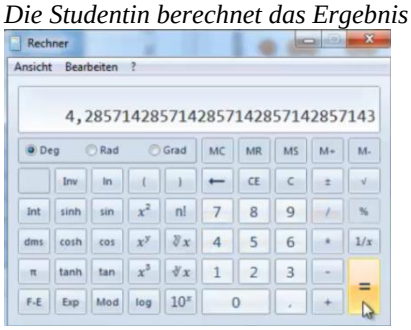
Episode 8-2: Händische Simulation (Z. 83-150)

83 I: Vielleicht nutzt ihr die Zettel doch, um das nochmal zu machen.
84 A: Ja.
85 I: Vielleicht hilft euch das nochmal.
86 M: (unv.)
87 I: Ihr könnt noch einen Stift haben.
88 A: Genau. Ich schreib. Willst du die Unteren aufschreiben, ich schreib die Oberen?
89 M: Ja.
90 *Die vorbereiteten leeren Zettel werden mit den Werten des Merkmals Blutdrucksenkung beschriftet.*
91 A: Ok. Jetzt müssen wir die mischen.
92 M: Ja. (...) Dann mach ich mal Notizen.
93 A: Und dann wieder in zwei Gruppen aufteilen, ne?
94 M: Ja. Wie wollen wir sie nennen?
95 A: A und B?
96 M: A und B.
97 A: Und dann/.
98 M:/ Verteil die erst mal auf zwei Haufen.
99 A: Ok.
100 *Die verdeckten Zettel werden gemischt und verdeckt abwechselnd auf zwei Haufen aufgeteilt.*

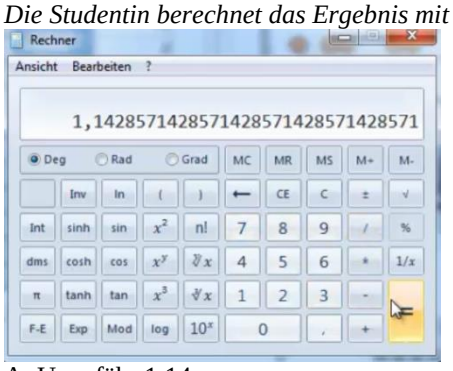
101 A: Alles klar. Ok. (...) Wo ist der Taschenrechner?
102 I: Der ist (...) da.
103 A: Ok.
104 *Die Studentin öffnet den Taschenrechner.*

105 *Studentin (M) nimmt das leere Blatt und beschriftet es mit A und B als Spaltenüberschriften.*
106 M: Ich trag einfach schon mal ein. Oder sollen wir nur die Durchschnittswerte eintragen?
107 A: Ja würde ich sagen, oder?
108 M: Ja.
109 A: Willst du mir die hier einmal diktieren aus Gruppe B?
110 M: Die Nullen lass ich mal weg. 12 + 6 + 8 - 6 + 2 + 14.
111 *Studentin (A) tippt die Zahlen im Taschenrechner des Computers ein.*
112 A: Gleich 30. Geteilt/ ist das geteilt?

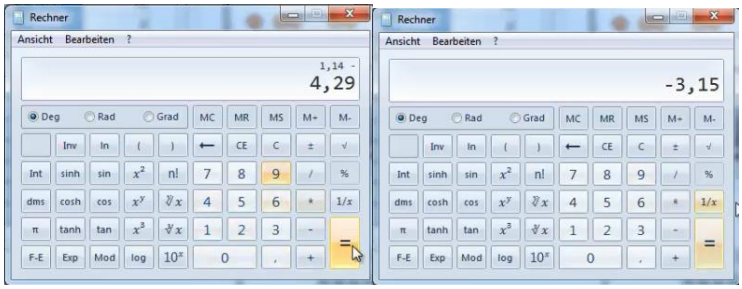
113 M: Ja. Durch 7.
114 A: Gleich. Ungefähr 4,29.
115 *Die Studentin berechnet das Ergebnis mit dem Taschenrechner.*



116 M: Ist gekauft. Ist das Gruppe A jetzt?
117 A: B. Das war Gruppe B.
118 M: Ja ist ja eigentlich egal.
119 A: So und das gleiche jetzt mit A. Alles klar?
120 M: Ja -4 oder fang lieber mit einer/ 10.
121 A: 10.
122 M: +2.
123 A: + 2.
124 M: +2.
125 A: +2.
126 M: +1.
127 A: +1.
128 M: -3.
129 A: -3.
130 M: -4.
131 A: -4.
132 M: Ja das war es.
133 A: Ok geteilt durch 7, ne?
134 M: Ja.
135 *Die Studentin berechnet das Ergebnis mit dem Taschenrechner.*



136 A: Ungefähr 1,14.
137 M: So, wollen wir nochmal randomisieren?
138 A: Musste man da nicht noch was machen? Den Unterschied dazwischen, oder?
139 M: Ja, aber das können wir jetzt machen oder später. (...) Dann lass uns aber beide gerundeten Werte nehmen. Also 1,14.
140 A: Nee umgekehrt. DAS Minus DAS, oder?
141 M: Hauptsache wir behalten die Reihenfolge bei. Mach so rum. Mach A minus B.
142 A: Ok. 1,14 minus 4,29. 3,15.
143 *Die Studentin gibt die Zahlen in den Taschenrechner ein.*



144 M: Wie soll ich den Wert jetzt dazwischen schreiben? Differenz ist -3,15. Also schon mal ein bisschen niedriger/.
145 Die Differenz wird auf dem Blatt Papier als einzelner Wert notiert.
146 A:/ Ja als 7,7.
147 M: Ja. Ok.
148 A: Ja.
149 M: Sollen wir so nochmal oder sollen wir es jetzt am Computer probieren?
150 A: Ja ich weiß nicht mehr genau.

Nachdem in der vorigen Episode sehr große Unsicherheit über die Benutzung der Zufallsmaschine herrschte, schlägt nun die Interviewerin zu Beginn dieser Episode vor, die Zettel zu nutzen (83), womit gemeint ist, die Simulation händisch durchzuführen. Alisa stimmt mit „Ja.“ (84) zu und damit ist vereinbart, dass Mandy und Alisa zunächst eine händische Simulation durchführen. Zunächst nutzen sie die leeren kleinen Notizzettel und notieren hierauf die Blutdruckreduktionen, die im Experiment aufgetreten sind, wobei Alisa Mandy die „unteren“ zuteilt (88), das sind mit Bezug auf den Aufgabenzettel die Werte der „Normale Öldiät“ Gruppe und sich selbst die oberen Werte zuteilt, das sind die der Fischölgruppe. Außer den reinen Zahlen, ein Wert pro Zettel, wird nichts weiter von den Beiden auf den Zetteln notiert (90). Sie stellen somit also Unabhängigkeit von den Gruppen bzw. von den Öldiäten her. Dies entspricht dem Vorgehen, wie es im Baustein Inferenzstatistik gezeigt und durchgeführt wurde. Alisa plant den weiteren Prozess: „Jetzt müssen wir die mischen“ (91) und „dann wieder in zwei Gruppen aufteilen“ (93). Mandy erwähnt zwischendrin (92), dass sie Notizen mache. Die zwei neuen Gruppen werden nach einem Vorschlag von Alisa (95) A und B genannt und daraufhin die Zettel verdeckt gemischt und zufällig abwechselnd auf zwei Stapel aufgeteilt (97-100). Dies entspricht dem Vorgang, der auch im Seminar durchgeführt wurde. Anschließend einigen Mandy und Alisa sich darauf, nur den Durchschnittswert jeder Gruppe zu notieren (106). In einem längeren Dialog (108-136) ermitteln sie 4,29 als Mittelwert ihrer neuen Gruppe B und 1,14 als Mittelwert ihrer neuen Gruppe A, wobei sie den Taschenrechner des Laptops zur Berechnung nutzen. Es wird an keiner Stelle diskutiert, dass der Mittelwert (und nicht der Median) verwendet wird und auch das Ziel (Bildung der Differenz der Mittelwerte, um dies mit dem beobachteten Wert des ursprünglichen Experiments zu vergleichen) wird nicht ausgesprochen. Nachdem die Mittelwerte berechnet wurden, fragt Mandy, ob nochmal randomisiert werden solle (137). Entweder ist ihr nicht klar, wie nun weiter gearbeitet werden soll, dass nämlich die Dif-

ferenz zwischen den beiden Mittelwerten der neuen Gruppen gebildet werden muss, oder sie möchte tatsächliche die händische Simulation mehrmals durchführen und mehrere neue Mittelwerte bestimmen, bevor sie die Teststatistik berechnet. Alisa antwortet mit einer Gegenfrage im nächsten Turn „ Musste man da nicht noch was machen? Den Unterschied dazwischen, oder?“ (138) und zeigt damit, dass sie zunächst die Teststatistik bestimmen möchte. Mandy scheint dies im Blick gehabt zu haben, also das weitere Vorgehen doch geplant zu haben, denn sie antwortet daraufhin: „Ja, aber das können wir jetzt machen oder später.“ (139). Nun soll also die Differenz bestimmt werden und Mandy beginnt mit dem gerundeten Wert 1,14, von dem sie dann logischerweise den größeren Wert 4,29 der anderen neuen Gruppe abziehen müsste. Alisa schlägt vor, dies umzukehren, so dass 4,29 – 1,14 gerechnet werden würde (140). Aus welchem Grund sie das vorschlägt, wird nicht klar. Entweder hat sie im Blick, dass sich durch das Vertauschen von Subtrahend und Minuend eine positive Zahl ergeben würde, oder sie möchte lieber den Mittelwert der Gruppe A vom Mittelwert der Gruppe B abziehen. Mandy scheint es auch nicht klar zu sein, denn sie antwortet „Hauptsache wir behalten die Reihenfolge bei. Mach so rum. Mach A minus B.“ (141), womit sie zunächst richtig feststellt, dass die Differenz immer in die gleiche Richtung gebildet werden muss bei erneuter Randomisierung und mit „mach A minus B“ wohl meint, dass der Mittelwert der Gruppe B vom Mittelwert der Gruppe A abgezogen werden soll. Das Ergebnis -3,15 wird eingeschätzt als „… schon mal ein bisschen niedriger.“ (144), wodurch sie Bezug nimmt auf die beobachtete Differenz des ursprünglichen Experiments, was durch Alisa im nächsten Turn (146) expliziert wird. Auch wenn dieser Wert vorher bei den beobachteten Daten nicht explizit notiert wurde, so hatten ihn die Beiden doch offensichtlich im Blick. Daraufhin fragt Mandy: „Sollen wir so nochmal oder sollen wir es jetzt am Computer probieren?“. Das kann als Bereitschaft interpretiert werden, dass sie sich den Prozess nun noch einmal verdeutlicht hat und bereit ist, die Software zu nutzen, um den Randomisierungstest durchzuführen. Die händische Simulation hat bei den beiden gut funktioniert, allerdings haben sie diese nicht bis zur Erstellung einer Referenzverteilung weiter geführt, so dass nicht festgestellt werden kann, ob ihnen die weiteren Schritte händisch klar sind.

Episode 8-3: Planen des Transfers der händischen Simulation in den Computer (Z. 150-159)

151 I: Also das dient ja vor allem dazu, dass ihr euch nochmal in den Kopf gerufen habt: Was macht ihr eigentlich? Und jetzt müsst ihr versuchen das, was ihr gerade von Hand gemacht habt/ was sehr gut geklappt hat/ in den Computer zu übertragen.
152 A: Ok.
153 M: Ich würde schätzen wir haben zwei Merkmale.
154 I: Nämlich?
155 M: Fischöl und normales Öl oder A und B. Da bin ich nicht ganz sicher muss ich sagen.
156 A: Ich dachte, die Werte kommen alle in die Box. Hat man nicht?/ Nee warte.
157 I: Überlegt mal gerade völlig ohne das. Was ihr sozusagen gemacht habt. Was habt ihr im ersten

	Schritt gemacht?
158	A: Alle Werte zusammen genommen.
159	M: Ja, wenn wir uns vorstellen wir haben die jetzt aus der Urne genommen, dann packen wir die jetzt erst mal in die Urne rein.

Die Interviewerin fasst den Zweck der händischen Simulation im nächsten Turn zusammen: „Also das dient ja vor allem dazu, dass ihr euch nochmal in den Kopf gerufen habt, was macht ihr eigentlich“. Und sie nennt als nächstes Ziel die Kombination mit der Softwareebene: „Und jetzt müsst ihr versuchen das, was ihr gerade von Hand gemacht habt, was sehr gut geklappt hat, in den Computer zu übertragen.“ (151), wodurch sie gleichermaßen motiviert und das Vorangegangene positiv bewertet. Alisas „Ok.“ (152) im darauf folgenden Turn lässt sich einerseits interpretieren, dass sie die Zusammenfassung und die Bewertung des Vergangenen durch die Interviewerin annimmt, andererseits aber auch, dass ihr das Ziel bzw. das weitere Vorgehen (in den Computer übertragen) nicht ganz klar war, und sie das von der Interviewerin vorgegebene nun als ihres akzeptiert. Dies wird aus ihrer kurzen Antwort interpretiert, da sie verbal das weitere Vorgehen nicht selbst aufgreift. Mandy hingegen versucht einen ersten Ansatz für das weitere Vorgehen zu finden und formuliert vorsichtig „Ich würde schätzen wir haben zwei Merkmale.“ (153), was sich sowohl auf die kontextuelle Ebene als auch auf die Softwareebene beziehen kann. Im Kontext sind die zwei Merkmale erstens die Gruppenzugehörigkeit, also mit der Bezeichnung von Mandy und Alisa die neuen Gruppen, und zweitens die Blutdruckwerte. Gemeint sein kann auch, dass sie an die Zufallsmaschine denkt, in der in der unbearbeiteten Ausgangssituation standardmäßig zwei Merkmale vorhanden sind. Auf die Nachfrage der Interviewerin „Nämlich?“ (154) erläutert sie nun diese zwei „Merkmale“ mit „Fischöl und normales Öl oder A und B. Da bin ich nicht ganz sicher muss ich sagen.“ (155), wodurch sie eine Fehlvorstellung zum Begriff des Merkmals präsentiert und inhaltliche Unsicherheit zum Ausdruck bringt. Offensichtlich meint sie mit „Merkmal“ „Ausprägung“, denn das, was sie aufzählt, wären Ausprägungen eines Merkmals „Gruppenzugehörigkeit“. Die Ausprägungen „Fischöl“ und „normales Öl“ gehören zu dem Merkmal Gruppenzugehörigkeit des ursprünglichen Experiments, die Ausprägungen „A und B“ ließen sich gut für die Randomisierung verwenden, der die Nullhypothese zugrunde liegt. Diese Unterscheidung scheint Mandy jedoch nicht klar zu sein. Natürlich wäre es auch möglich, im Simulationsmodell mit den Bezeichnungen *Fischöl* und *normales Öl* für die zwei neuen Gruppen zu arbeiten, allerdings wäre hier die Gefahr sehr groß, dass dies in Bezug auf die jeweilige Einnahme des Öls interpretiert würde, von der im Nullmodell ja gerade Unabhängigkeit angenommen wird, weshalb diese Bezeichnung eine unglückliche Wahl wäre. Alisa übernimmt den Turn und wirft ein, ohne auf Mandys Äußerung einzugehen, „Ich dachte, die Werte kommen alle in die Box…“ (156) und stockt dann erst einmal. „Die Werte“ bezieht sich vermutlich auf die Blutdruckreduktionen, die im Experiment aufgetreten sind, und die sie korrekt an dieser Stelle losgelöst von der Gruppenzugehö-

rigkeit, respektive vom eingenommenen Öl, betrachtet. Aufgrund des Stockens von Alisa und den unstrukturierten Ansätzen der Beiden greift die Interviewerin noch einmal ein und fordert dazu auf, noch einmal zu überlegen, was gerade von den Beiden händisch gemacht wurde (157). Alisa antwortet kurz und bezieht sich erneut (korrekt) auf alle Werte, die zusammengenommen worden seien (158) und Mandy führt daraufhin die Vorstellung aus, die mit dem Bild einer Urne verknüpft dem Vorgang zugrunde liegt als „Ja, wenn wir uns vorstellen wir haben die jetzt aus der Urne genommen…“ (159) und kommt nun anhand des gedanklichen Rekonstruierens der zuvor durchgeführten händischen Simulation in Verbindung mit dem Bild der Urne zu einem ersten umsetzbaren Ansatz bezüglich der Simulation in TinkerPlots „… dann packen wir die jetzt erst mal in die Urne rein.“ (159). Hiermit bezieht sie sich nun mit Hilfe eines Bildes der statistischen Ebene (Urne) auf die Softwareebene, in der die Urne „Box“ heißt, aber die gleiche Funktion erfüllt. Somit ist mit Unterstützung der Interviewerin ein Ansatz gefunden worden.

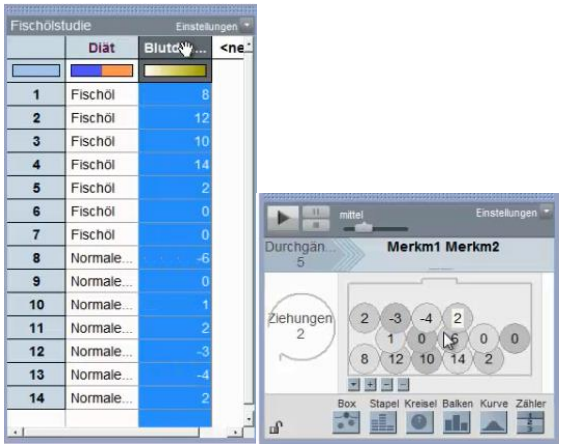
Episode 8-4: Erstellen der Zufallsmaschine (Z. 160-173)

160 I: Genau. Dazu macht ihr das erst mal alles raus mit dem Minus.

161 *Alle Kugeln werden aus der Zufallsmaschine gelöscht.*

162 I: Markiert die in der Tabelle. Da sind die ja alle drinnen. Genau und jetzt Steuerung C drücken zum Kopieren. Und darein klicken, wo leer steht und Steuerung V.

163 *Die Werte aus der Tabelle werden markiert und die Box eingefügt.*



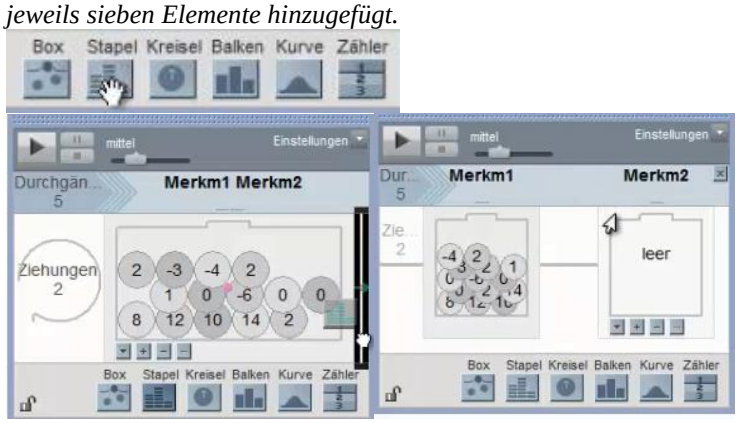
164 I: Jetzt habt ihr die alle drin. So, was war jetzt der zweite Schritt, den ihr gemacht habt?

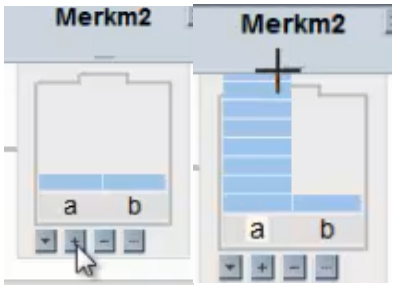
165 M: Dann haben wir die gezogen und in A und B eingeteilt. Also machen wir statt Merkmal 1 und 2/ nein/.

166 I: Es ist völlig richtig. In TinkerPlots macht ihr das jetzt, indem ihr ein zweites Bauteil dahinter setzt, z.B. den Stapel. Ihr habt jetzt sozusagen jedem ein neues Gruppenetikett hinzugefügt.

Entweder Gruppe A oder Gruppe B und das macht ihr jetzt. Hier machst du zwei Mal auf das Plus bitte. Das sind jetzt eure Stapel und da musst du jetzt mit der Maus entsprechend hochziehen, damit du 7 drin hast.

167 *Die Studentin fügt der Zufallsmaschine ein weiteres Bauteil hinzu. Anschließend werden zwei Stapel hinzugefügt, die von TinkerPlots automatisch mit a und b bezeichnet sind. Ihnen werden jeweils sieben Elemente hinzugefügt.*





168 A: 1, 2, 3, 4, 5, 6, 7.

169 I: Kannst es ansonsten auch genau mit dem kleinen Pfeil darunter eingeben. Da gibt es Anzahlen anzeigen ganz oben.

170 A: Ja.

171 I: Und da kannst du jetzt 7 eingeben.

172 A: Ah steht schon. Ja und dann hier auch 7. Alles klar.

173 *Die Studentin korrigiert die Anzahl der einzelnen Stapel, da diese vorher nicht 7 betrugen.*

Zu Beginn dieser Episode gibt die Interviewerin konkrete Anweisungen auf Softwareebene, wie die Zufallsmaschine zu bearbeiten ist, um wie in der vorigen Episode formuliert, „alle Werte zusammen genommen“ in die „Urne zu packen“ (158-159), da dies von Mandy und Alisa nicht selbständig in Angriff genommen wird. Zunächst gibt sie die Anweisung, alle Kugeln aus der Zufallsmaschine zu löschen (160), dann erläutert sie, wie die ursprünglichen Werte zu kopieren und in die leere Zufallsmaschine einzufügen sind (162), was beides von den beiden Teilnehmerinnen direkt im Anschluss an die Turns ohne weitere Nachfrage ausgeführt wird. Nachdem die Teilnehmerinnen hiernach wieder nicht selbständig weiter arbeiten, übernimmt die Interviewerin erneut den Turn und fragt nach dem nächsten Schritt, der händisch gemacht wurde (164). Mandy antwortet mit „Dann haben wir die gezogen und in A und B eingeteilt. Also machen wir statt Merkmal 1 und 2, nein.“ (165). Sie rekapituliert im ersten Teil des Turns richtig das Vorgehen, das Einteilen der Werte in zwei neuen Gruppen A und B und versucht im zweiten Teil des Turns dies auf die Software zu übertragen. In der Zufallsmaschine sind immer noch per Standardeinstellung zwei Merkmale ‚Merkm1‘ und ‚Merkm2‘ zu sehen, die Mandy nun vermutlich erstens umbenennen möchte („statt Merkmal1 und 2“) oder mit Werten füllen möchte. Da sie wiederum abbricht und auch Alisa den Turn nicht übernimmt, erläutert die Interviewerin das weitere Vorgehen in TinkerPlots (166). An dieser Stelle wird direkt von ihr auf statistischer Ebene erläutert, das mit einem zweiten Bauteil, dem Stapel, die Etiketten Gruppe A oder Gruppe B hinzugefügt werden, ohne den Studentinnen Gelegenheit zu geben, dies für sich selbst zu formulieren. Am Ende ihres Turns finden sich wiederum Anweisungen auf Softwareebene, wie das Stapelbauteil zu erstellen ist und es wird auch vorgegeben, dass jeder Stapel sieben Elemente enthalten muss. Das zweite Bauteil wird entsprechend von den Studentinnen im Folgenden (167-173) erstellt.

Episode 8-5: Umbenennen der Merkmale (Z. 174-192)

174 A: Sollen wir die Merkmale/ nee das ist ja/.
175 I: Was überlegst du? Denk laut bitte.
176 A: Ob man Merkmal 1 und 2 jetzt umbenennen soll. Aber wenn müsste ich/.
177 M: Eigentlich ist ja. Das (*zeigt auf das Bauteil Stapel der Zufallsmaschine*) sind auf jeden Fall die neuen Gruppen.
178 A: Ja aber in was? Also in was, wenn du das umbenennen würdest?
179 M: Das (zeigt auf das erste Bauteil, in welchem sich die Werte des Merkmals Blutdruckreduktion befinden) ist alles gemischt. Oder ja durcheinander.
180 I: Was ist es denn, was ihr darin habt?
181 M&A: Alle Werte.
182 I: Ja schreibt das doch auf.
183 *Merkm1 wird in „Werte“ umbenannt.*

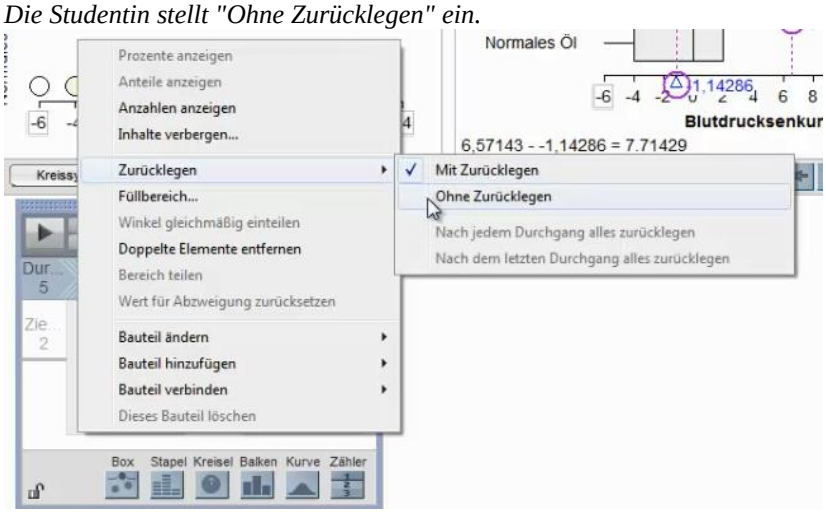
184 M: Und dann Nummer 2 ist die Neue.
185 A: Neue Gruppe.
186 M: Ja, weil A und B haben wir dann da unten ja.
187 A: Soll ich dann Gruppe, oder?
188 M: Neue Gruppe würde ich.
189 A: Ok neue Gruppe oder Gruppen?
190 M: Gruppen. Sind ja zwei.
191 A: Gut.
192 *Merkm2 wird in „Neue_Gruppen“ umbenannt.*

Nachdem die Zufallsmaschine in der vorigen Episode erstellt wurde, überlegt nun Alisa, ob die Merkmale 1 und 2 mit Bezug zum Kontext umbenannt werden sollen (174-176). Mandy greift dies auf und erläutert, indem sie auf das Bauteil Stapel zeigt, dass dies die neuen Gruppen seien (177). Sie zeigt hier, dass sie die Funktion des zweiten Bauteils verstanden hat, macht jedoch keinen Vorschlag für eine Umbenennung der Merkmale. Alisa fragt daraufhin konkret nach, ob und falls ja in was Mandy die Merkmale umbenennen würde (178), woraufhin Mandy nicht mit Benennungsvorschlägen antwortet, sondern sich zunächst die Funktion des ersten Bauteils bewusst macht, indem sie darauf zeigt und sagt „Das ist alles gemischt. Oder ja durcheinander.“ (179), womit sie durchaus meinen kann, dass die Werte beider ursprünglichen Gruppen hier „durcheinander“ in einer Urne sind und dadurch die ursprüngliche Gruppenzugehörigkeit aufgehoben ist. Da immer noch kein Vorschlag für eine Merkmalsbezeichnung gemacht wurde, obwohl dieser Bedarf angezeigt wurde, schaltet sich die Interviewerin erneut ein mit der Frage, was es denn sei, was darin liege (180). Beide antworten simultan „Beide Werte.“ (181) und erhalten hierauf den Vorschlag der Interviewerin „Ja schreibt das doch auf.“ (182),

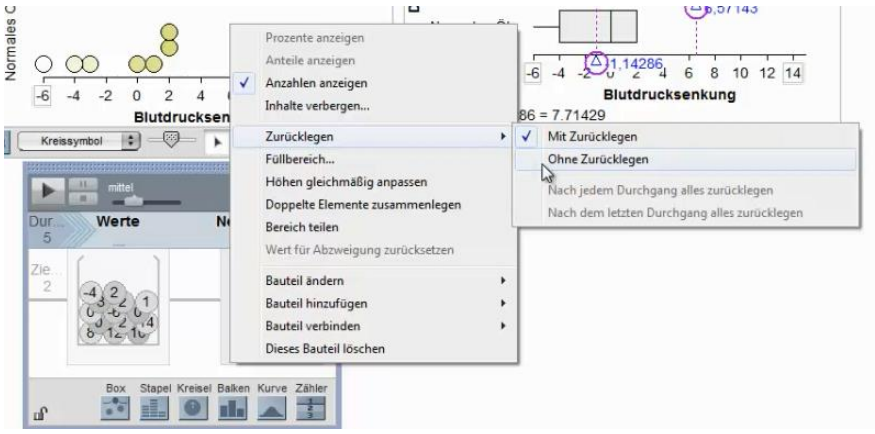
woraufhin das erste Merkmal in „Werte“ umbenannt wird (183). Hiernach diskutieren Mandy und Alisa wieder untereinander (184-192) und handeln aus, das zweite Merkmal in „neue Gruppen“ umzubenennen mit der Begründung, dass A und B bereits „da unten“ seien (186) und „Gruppen“ statt „Gruppe“, weil es ja zwei neue Gruppen gäbe (190). Für Mandy scheint der Vorschlag, das zweite Merkmal in „A und B“ umzubenennen, zumindest im Hinterkopf gewesen zu sein, weil sie dies explizit erwähnt. Ob sie nun eine Gleichbenennung des Merkmals mit seinen Ausprägungen vermeiden wollte oder ob ihr dies aus anderen Gründen unpassend erschien, wird nicht von ihr ausgeführt. Mit Bezug zum Kontext und mit Hilfe durch die Interviewerin wurden in dieser Episode Merkmalsbezeichnungen gefunden. Die hierüber geführte Diskussion hat sicherlich weiter zum Verständnis beigetragen.

Episode 8-6: Die Einstellung Ohne Zurücklegen (Z. 193-200)

193 I: Jetzt müsst ihr beim Simulieren immer noch überlegen, ob ihr mit oder ohne Zurücklegen zieht.
194 M& A: Ohne.
195 I: Das stellt ihr auch bei dem kleinen Pfeil ein. Da gibt’s irgendwo zurücklegen und dann könnt ihr das auswählen.
196 *Die Studentin stellt "Ohne Zurücklegen" ein.*



197 A: So. Fehlt jetzt noch irgendetwas?
198 I: Bei dem zweiten Bauteil musst du das auch noch machen.
199 A: Ok.
200 *Die Studentin führt die Aktion auch für das zweite Bauteil durch.*



Nachdem die Merkmalsumbenennung auf Softwareebene abgeschlossen ist und erneut die Studentinnen keinen Ansatz wissen, um selbständig weiter zu arbeiten, erwähnt die Interviewerin, dass beim Simulieren überlegt werden müsse, ob mit oder ohne Zurücklegen gezogen werden solle (193), um darauf aufmerksam zu machen, dass diese Einstellung noch vorzunehmen ist. Wieder simultan antworten Mandy und Alisa „Ohne.“ (194) und führen nach einer entsprechenden technischen Anweisung durch die Interviewerin (195) die Aktion für beide Bauteile durch (196-200). Es wird nicht diskutiert, aus welchem Grund diese Einstellung gewählt werden muss.

Episode 8-7: Bestimmen der Durchgangsanzahl (Z. 201-210)

201	M: Wie oft/ also wie viele Ziehungen? (*Sie versucht, auf die 2 bei Ziehungen zu klicken*) Was machen wir jetzt? Das kann man nicht ändern.
202	I: Nein. Der zieht jetzt in jedem Durchgang aus jeder Urne einmal. Aber die Frage ist wie viele Durchgänge braucht ihr?
203	M: Ja da ist doch jetzt eigentlich wieder je mehr Durchgänge wir haben desto/ also ein genaueres Bild kriegen wir. Weil jetzt wir ja quasi nur einen Durchgang gemacht.
204	A: Mhm (bejahend). Ja.
205	M: Also, wenn ich mich nicht täusche. (5 sec)
206	I: Ja aber jetzt müsst ihr euch kurz überlegen. Das ist genau was ihr eben gerade gemacht habt. Das heißt, ihr habt die alle verteilt. Das heißt, ihr könnt da nur eine 14 einstellen.
207	M: Achso. Ah ich dachte jetzt wie oft wir/.
208	A: / Aber dann machst du das/ ne nach 14 Mal ist das leer.
209	M: Ja. Ich hatte irgendwie einen Denkfehler. Alles klar. Ok dann/.
210	*Es werden 14 Durchgänge in der Zufallsmaschine eingestellt.*

Diese Episode wird eingeleitet von Mandy, die danach fragt, wie viele Ziehungen benötigt werden (201) und feststellt, dass sich die Anzahl der Ziehungen nicht ändern lässt, was daran liegt, dass TinkerPlots automatisch die Ziehungsanzahl auf die Anzahl der Bauteile setzt, sobald mehr als ein Bauteil verwendet wird. Die Interviewerin erläutert die Ziehungsanzahl zwei mit „Der zieht jetzt in jedem Durchgang aus jeder Urne einmal.“ (202) und schließt die Frage an, wie viele Durchgänge gebraucht werden. Mandy beantwortet dies qualitativ mit „… je mehr Durchgänge wir haben/desto ein genaueres Bild kriegen wir…“ (203), was darauf hindeutet, dass sie die Durchgangsanzahl hier

nicht mit der Reproduktion der Experimentdurchführung zusammenbringt, für die die vierzehn Blutdruckwerte neue Gruppenetiketten erhalten müssen, sondern mit der Gesamtwiederholungszahl verwechselt, mit der diese Randomisierung wiederholt werden soll. Dies wird von der Interviewerin aufgegriffen mit dem Hinweis auf den zuvor händisch durchgeführten Prozess und der konkreten Anweisung „… das heißt, ihr könnt da nur eine 14 einstellen.“ (206). Mandy thematisiert im Ansatz ihre Fehlvorstellung im nächsten Turn mit „Achso. Ah ich dachte jetzt wie oft wir.“ (207), was sich logisch ergänzen lässt mit „wie oft wir das Experiment durchführen sollen“, aber direkt von Alisa auf Softwareebene interpretiert wird, weil „… nach 14 Mal ist das leer“ (208), womit sie nur meinen kann, dass sich nach vierzehn Durchgängen in der Zufallsmaschine aufgrund der Einstellung ohne Zurücklegen keine Elemente mehr befinden.

In dieser Episode scheint bei Mandy eine Fehlvorstellung bezüglich der Funktionsweise der Zufallsmaschine vorzuliegen, die sich darin ausdrückt, dass Mandy die Ziehungsanzahl mit der Durchgangsanzahl und die Durchgangsanzahl mit der Wiederholungszahl des Experiments/des Tests verwechselt, so wie dies als mögliche Schwierigkeit in der a priori Analyse genannt wurde.

Episode 8-8: Validieren der erstellten Zufallsmaschine (Z. 211-213)

211 I: Dann lass doch einfach mal laufen und guck dir an, was das macht und, ob das macht, was ihr wollt.

212 *Die Studentinnen lassen die Zufallsmaschine langsam laufen.*

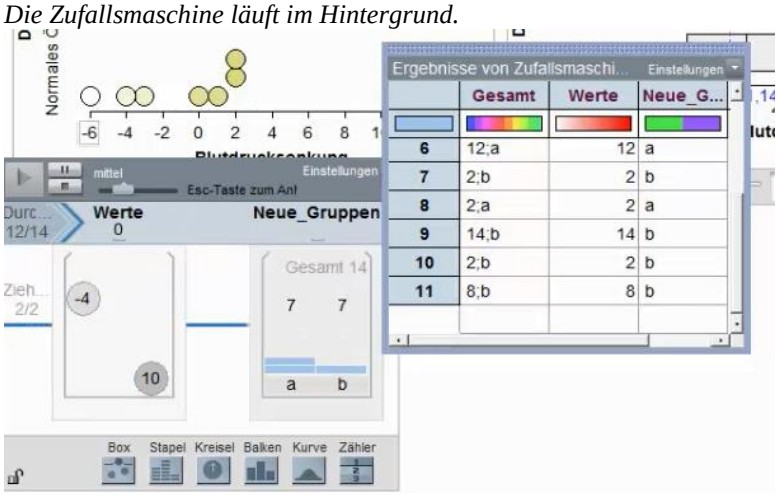
213 A: Ja macht er. Er teilt die Werte den Gruppen zu. Zufällig.

Aufgrund der Tatsache, dass bis hierher viel auf Softwareebene durch die Interviewerin vorgegeben wurde, schlägt diese nun vor „Dann lass doch einfach mal laufen und guck dir an, was das macht und ob das macht, was ihr wollt.“ (211). Daraufhin lassen Mandy und Alisa die Zufallsmaschine mit der Einstellung *mittlere Geschwindigkeit* laufen und beobachten, wie der Ziehungsprozess visualisiert wird. Daraufhin kommt Alisa zu der Feststellung „Ja macht er. Er teilt die Werte den Gruppen zu. Zufällig.“ (213), was als Validieren der erstellten Zufallsmaschine interpretiert wird. Das nachgeschobene Wort „Zufällig“ lässt den Schluss zu, dass sie die in der Nullhypothese ausgedrückte, der Si-

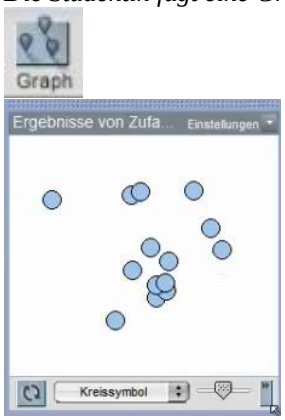
mulation zugrunde liegende, Situation im Hinterkopf hat, in der die zufällige Neuaufteilung der Werte auf zwei Gruppen beschrieben wird.

Episode 8-9: Weiteres Planen und Ausfüllen des Schemas (Z. 214-227)

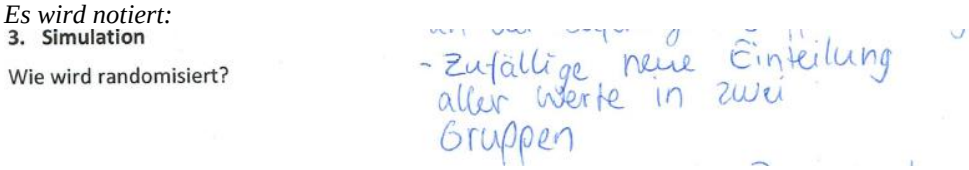
214 M: Ja. (...) Und dann müssen wir noch sagen, dass er die Differenz ausrechnen soll, bitte.
215 I: Genau. Also Auswertung macht ihr über eine Graphik. Das heißt, das müsst ihr jetzt wieder über eine Graphik darstellen.
216 A: Müssen wir das hier nicht schon mal eintragen mit der Simulation?
217 I: Könnt ihr machen.
218 *Die Zufallsmaschine läuft im Hintergrund.*



219 M: Was wollen wir denn darein schreiben? Wie wird randomisiert?
220 A: Ja ist das nicht quasi die Beschreibung, dass man die Werte zufällig in Gruppen A und B zuteilt. Also dass die zufällig neuen zugeteilt werden?
221 M: Dann schreib ich rein: Zufällige Einteilung in zwei Gruppen/ nee in neue Gruppen.
222 A: Zufällige Einteilung der Werte/ aller Werte.
223 M: Ich hol schon mal eine Graphik.
224 *Die Studentin fügt eine Graphik hinzu.*



225 M: Zufällige neue Einteilung aller Werte.
226 A: In 2 Gruppen.
227 *Es wird notiert:*



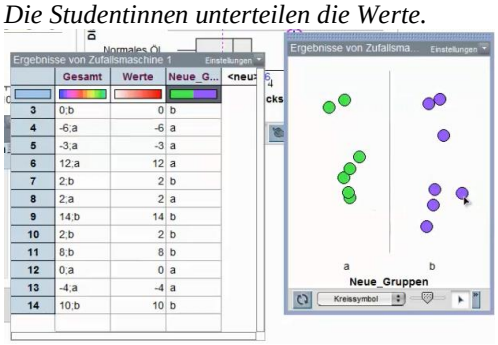
Nachdem in der vorigen Episode festgestellt wurde, dass die Zufallsmaschine zufriedenstellend arbeitet, plant nun Mandy das weitere Vorgehen auf Statistik- und Softwareebene, „dass er die Differenz ausrechnen soll.“ (214), womit sie vermutlich sprachlich verknappt dargestellt die Differenz der Mittelwerte der beiden neuen Gruppen meint und damit die Teststatistik im Blick hat, die in der Software definiert werden soll. Hierzu fügt die Interviewerin an, dass dies über eine Graphik geschehen müsse (215), da sie annimmt, dass Mandy und Alisa direkt damit weiterarbeiten möchten. Alisa wechselt allerdings erst zum Ausfüllen des Randomisierungstestschemas (216) und unterbricht somit die Gesprächsrichtung. Unter *3) Simulation* notieren Mandy und Alisa nach kurzer Aushandlung (219-222) „Zufällige neue Einteilung aller Werte in zwei Gruppen“ (227), was den Prozess gut zusammenfasst. Zwischenzeitlich (223-224) arbeitet Mandy bereits auf Softwareebene weiter, indem sie eine Graphik erstellt.

Interaktionseinheit 9: Teststatistik definieren und Referenzverteilung erstellen (Z. 228-334)

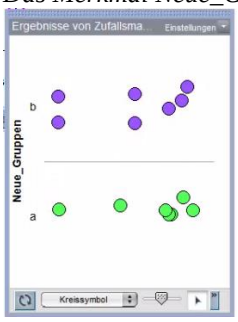
In der neunten Interaktionseinheit werden die Teststatistik definiert und die Referenzverteilung erstellt. Diese Einheit wird in sieben Episoden gegliedert. In der ersten Episode stellen Mandy und Alisa die simulierten Daten graphisch in TinkerPlots dar, in der zweiten Episode bestimmen sie die Teststatistik. Die Teststatistik als Messgröße zu definieren geschieht in der dritten Episode und in der vierten Episode wird die Anzahl der zu sammelnden Messgrößen festgelegt. Die fünfte Episode behandelt auf technischer Ebene die Einstellungen für das Beschleunigen des Messgrößensammelvorgangs und in der anschließenden sechsten Episode stellt die Geschwindigkeit des Messgrößensammelvorgangs eine Schwierigkeit dar, die jedoch gelöst werden kann. In der siebten Episode wird die Referenzverteilung erstellt.

Episode 9-1: Darstellen der Ergebnisse der Zufallsmaschine in einer Graphik (Z. 228-253)

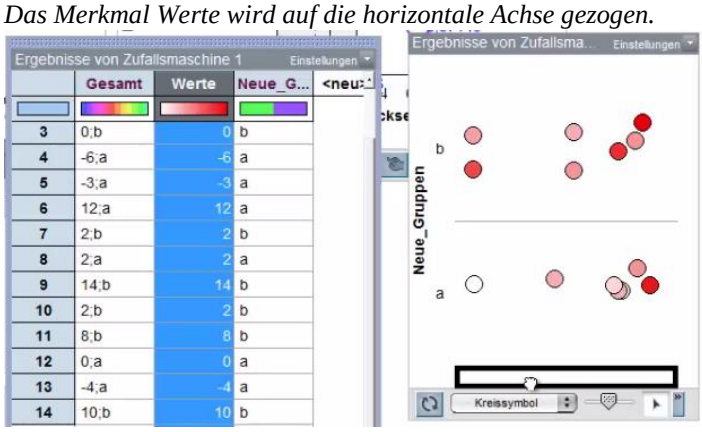
228 M: Ok. (...) Das sind jetzt unsere Werte?
229 A: Ja.
230 M: Jetzt müssen wir sie in A und B unterteilen und davon jeweils die Durchschnitte.
231 A: Also so was wie hier, ne? (*zeigt auf die Graphik der beobachteten Daten*)
232 M: Ja.
233 *Die Studentinnen unterteilen die Werte.*



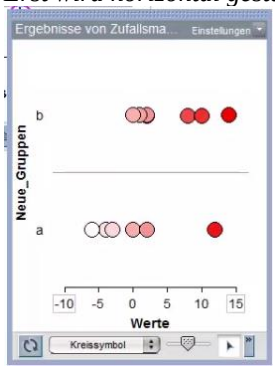
234 M: (unv.) Achso besser.
235 A: Also das kommt/.
236 M: / unten auf (unv.). Keine Ahnung.
237 *Die Graphik wird erneut durchmischt.*
238 A: Kann man das nicht rüber ziehen? Genau das kommt hierhin. Und die Werte/.
239 *Das Merkmal Neue_Gruppen wird auf die vertikale Achse gezogen.*



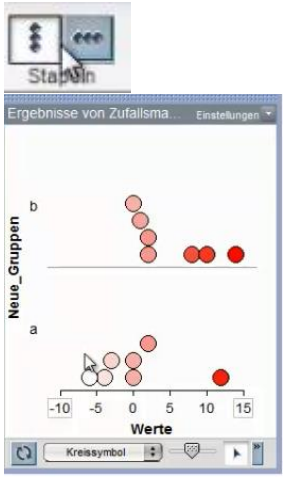
240 M:/ Was machst du?
241 *Das Merkmal Werte wird auf die horizontale Achse gezogen.*



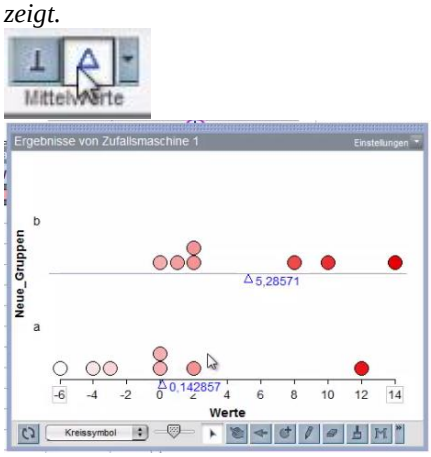
242 A: Hier unten. So wie da. Also die Werte kommen auf die X-Achse. Und dann muss man das auseinanderziehen.
243 M: Wunderschön.
244 A: Ja und dann kann man das ordnen.
245 M: Stapeln.
246 A: So oder so?
247 *Erst wird horizontal gestapelt.*



248 M: Nee mach mal das andere.
249 A: So.
250 *Dann wird vertikal gestapelt.*



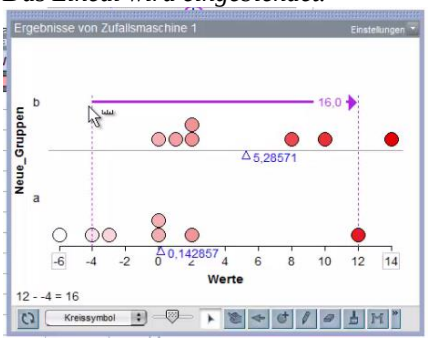
251 M: Und dann brauchen wir den Mittelwert. Also den Durchschnitt mit Zahlen.
252 A: Ja. Wert anzeigen.
253 *Die Graphik wird vergrößert, das arithmetische Mittel wird eingeblendet und der Wert angezeigt.*



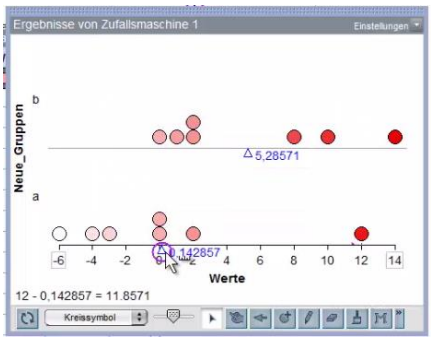
Wie um sich zu vergewissern, dass das bisherige richtig ist, eröffnet Mandy diese Episode mit der Frage „Das sind jetzt unsere Werte?“ (228), was direkt von Alisa bestätigt wird (229). Daraufhin macht Mandy einen Plan für das weitere Vorgehen „Jetzt müssen wir sie in A und B unterteilen und davon jeweils die Durchschnitte.“ (230). Alisa deutet auf die beobachteten Daten und fügt hinzu „Also so was wie hier, ne?“ (231), was von Mandy bestätigt wird und wodurch eine Zielstruktur für die von ihnen zu erstellende Graphik von ihnen selbst vorgegeben ist. Sie beginnen damit, Aktionen in der Graphik durchzuführen (273), wobei sie zunächst waagerecht nach den beiden neuen Gruppen trennen und anschließend senkrecht. Sie wählen das andere Merkmal aus (241), legen es auf die imaginäre X-Achse und separieren es komplett und stapeln senkrecht (242-250). Daraufhin wird der Mittelwert eingeblendet und mit dem Zahlenwert versehen (251-253). All dies wird sprachlich auf Softwareebene begleitet, ohne dass hier größere Schwierigkeiten erkennbar sind oder auf einer anderen als der Softwareebene kommuniziert wird.

Episode 9-2: Bestimmen der Teststatistik (Z. 254-273)

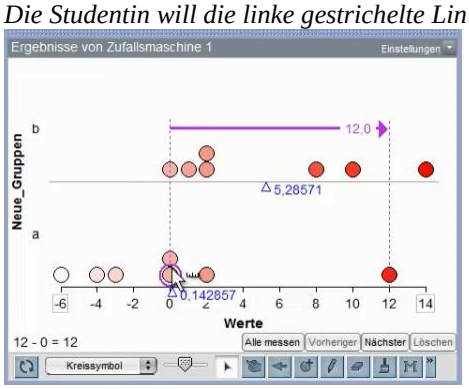
254 M: So und jetzt brauchen wir irgendwie die Differenz. (.)
255 I: Das geht mit dem Lineal.
256 *Das Lineal wird eingeblendet.*



257 A: Genau. Und dann muss man bis so ein Kreis kommt, oder?
258 I: Genau. Diese gestrichelte Linie ziehen. Auf das kleine Dreieck bis der Kreis kommt. Mach nochmal rückgängig.
259 *Durch das Verschieben der linken gestrichelten Linie ist das rechte Ende des Lineals verschwunden.*

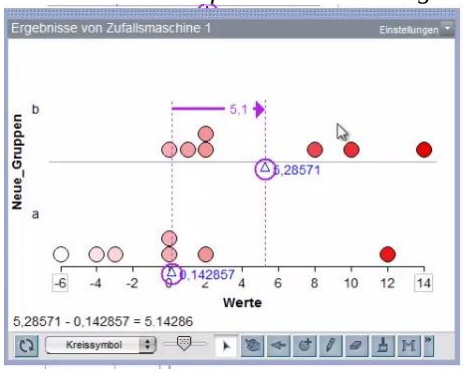


Das Lineal wird aus- und wieder eingeblendet.
260 M: So.
261 A: Oder mach einfach nochmal aus.
262 M: Zieh es nicht so weit runter.
263 A: Wie jetzt?
264 *Die Studentin will die linke gestrichelte Linie auf den Mittelwert ziehen.*



265 M: Aber nicht runterziehen.
266 I: Doch doch doch doch doch.
267 A: Sonst kommt der Kreis doch nicht, oder?
269 M: Doch.
270 A: Und jetzt das hier, ne?

271 M: Ja.
272 *Das Lineal wird auf beide Mittelwerte gezogen.*

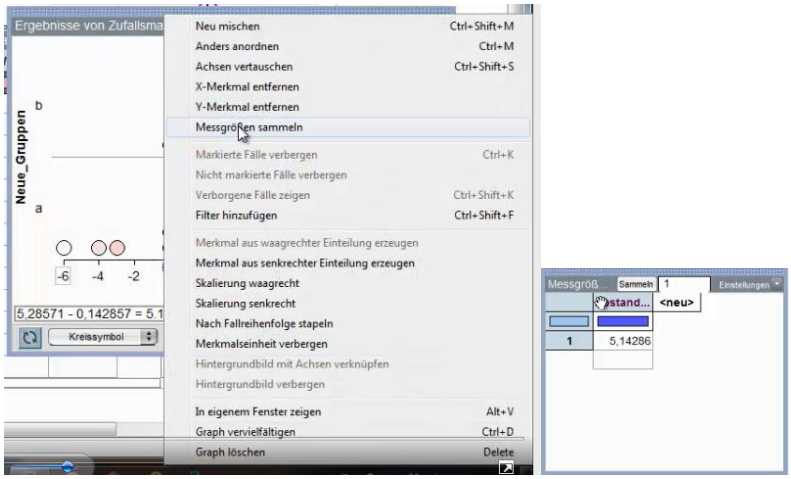


273 A: 5,1.

In der vorigen Episode wurden die Mittelwerte der beiden neuen Gruppen eingeblendet und daraufhin meint nun Mandy: „So und jetzt brauchen wir irgendwie die Differenz.“ (254), womit sie zum Ausdruck bringt, dass sie auf statistischer Ebene weiß, was nun folgen soll. Die Pause am Ende ihres Turns und die nicht erfolgte Turnübernahme durch Alisa veranlassen die Interviewerin, den Hinweis auf Softwareebene zu geben, dass dies mit dem Lineal ginge (255). Daraufhin wird das Lineal eingeblendet (256) und Alisa fragt nach: „Und dann muss man bis so ein Kreis kommt?“ (257), womit sie sich anscheinend noch an die Funktionsweise des Lineals in TinkerPlots erinnert, die genaue Bedienung jedoch nicht mehr parat hat. Die Interviewerin bestätigt sie und gibt Anweisung, wie das Lineal zu bedienen ist (258). Daraufhin arbeiten die beiden Teilnehmerinnen wieder selbständig und diskutieren über die Bedienung („Zieh es nicht so weit runter“ (262)) während sie die Aktionen ausführen. Am Ende nutzen sie das Lineal wie geplant, um den Abstand zwischen den Mittelwerten zu ermitteln und Alisa liest den Wert „5,1“ (273) ab. Sie bleiben hier stehen und formulieren dies nicht auf Kontext- oder Statistikebene. Beispielsweise hätten sie sagen können: „Jetzt haben wir eine Differenz von 5,1 zwischen den Mittelwerten der beiden neuen Gruppen“ – aber solchermaßen versprachlichen sie den Wert nicht.

Episode 9-3: Definieren der Messgröße (Z. 274-277)

274 M: So, das bräuchten wir jetzt häufiger, wenn es geht.
275 I: Genau das wollt ihr jetzt häufiger sammeln. Jetzt klickst du die 5,1 einmal an. Die Linie. Und einen Rechtsklick darauf, dann gibt es Messgrößen sammeln. Im oberen Drittel ist das irgendwo.
276 *Im Kontextmenü wird „Messgrößen sammeln“ ausgewählt. Die Tabelle „Messgrößen von…“ öffnet sich.*



277	A: Hab ich.

Mandy formuliert zu Beginn dieser Episode das Bedürfnis, dass sie „das jetzt häufiger [bräuchten], wenn es geht.“ (274).Wie zuvor lässt sich dies interpretieren, dass sie auf statistischer Ebene weiß, dass sie diesen Wert der Teststatistik nun sehr oft benötigt, um die Referenzverteilung zu erstellen, dass sie allerdings auf Ebene der Software nicht weiß, wie sie dahin gelangt. Die Interviewerin übernimmt den Turn und gibt genaue Anweisungen, wie nun eine Messgröße in TinkerPlots definiert wird (275). Diese werden ausgeführt (276) und von Alisa mit „Hab ich.“ (277) kommentiert. Auch dieser Kommentar wird so interpretiert, dass Alisa nun auf weitere Anweisungen wartet und nicht selbständig weiß, wie sie nun in TinkerPlots weiterarbeiten soll.

Episode 9-4: Anzahl der zu sammelnden Messgrößen festlegen (Z. 278-283)

278	I: Jetzt überlegt ihr wie häufig ihr das durchführen möchtet.
279	A: Ich weiß gar nicht wie oft haben wir das immer durchgeführt? Bei mir ist mein Laptop immer/ also hat aufgehört was zu machen/ deswegen.
280	I: Der hier macht was.
281	A: Ja aber ich weiß nicht mehr genau, weil ich da angefangen habe immer weiter runter zu gehen.
282	M: Das war doch dieses wo wir erst 1000 hatten und wo wir ausprobieren sollten, ob man nochmal 1000 draufmacht und noch eine Null.
283	A: Also 1000 geht. Dann machen wir 1000.

Die Interviewerin regt in dieser Episode an zu überlegen, wie häufig die Neuzuordnung der Werte zu den neuen Gruppen durchgeführt werden soll (278), implizit also wie viele Messgrößen gesammelt werden sollen. Alisa übernimmt den Turn und formuliert „Ich weiß gar nicht wie oft haben wird das immer durchgeführt? Bei mir ist mein Laptop immer/ also hat aufgehört was zu machen, deswegen.“ (279). Sie stellt keine Überlegungen an, welche Wiederholungszahl sinnvoll sein könnte, sondern versucht sich daran zu erinnern, was als Norm im Baustein Inferenzstatistik gesetzt wurde („wie oft haben wir das immer durchgeführt?“) und weist zudem darauf hin, dass dies bei ihr nicht funktioniert habe. Nach dem Hinweis der Interviewerin, dass der vorliegende Laptop

weiterarbeiten würde (280), drückt Alisa ihre Unsicherheit aus, „ich weiß nicht mehr genau“ (281), und gibt somit den Turn an Mandy weiter. Diese erinnert sich an den Wert 1000, der genommen werden sollte und dann „nochmal 1000 draufmacht und noch eine Null.“ (282), was möglicherweise bedeutet, dass sie zehntausend meint. Mit diesem konkreten Vorschlag erklärt sich Alisa einverstanden und sagt: „Also 1000 geht. Dann machen wir 1000.“ (283).

Episode 9-5: Beschleunigen des Messgrößensammelns durch Ausschalten der Animation (Z. 284-297)

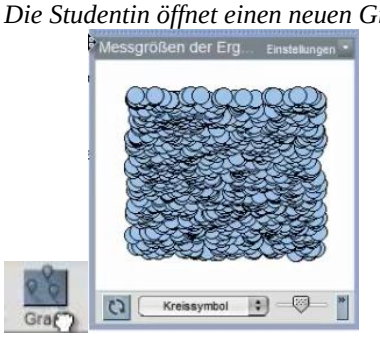
In dieser Episode wird auf Softwareebene auf Anweisung der Interviewerin die Animation beim Messgrößensammeln ausgeschaltet und die zur Zufallsmaschine gehörigen Objekte werden minimiert. Die Studentinnen führen die Anweisungen aus, die Kommunikation findet rein auf Softwareebene statt.

Episode 9-6: Geschwindigkeit des Messgrößensammelns als Schwierigkeit (Z. 298-324)

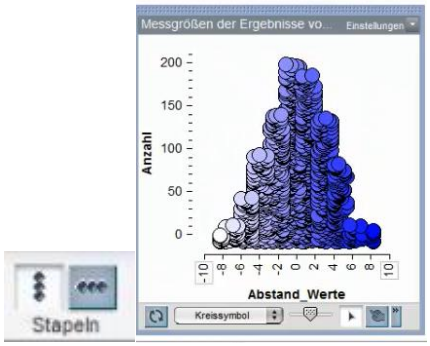
Zu Beginn dieser Episode wird das Messgrößensammeln gestartet, allerdings scheint die Software nichts zu machen. Die Interviewerin findet heraus, dass in der Zufallsmaschine noch „mittlere Geschwindigkeit“ eingestellt war. Nachdem dies auf „am schnellsten“ gestellt wurde, geht der Sammelvorgang zügig voran. Die Kommunikation beschränkt sich wie bereits zuvor auf die technische Seite der Software. Da der Sammelvorgang zwischendrin, nachdem zwei Messgrößen gesammelt waren, abgebrochen wurde, liegen am Ende 1002 Messgrößen vor, was von Alisa kommentiert wird mit „Ok. Jetzt hat er 1002 Ergebnisse, aber das ist ja nicht schlimm.“ (324).

Episode 9-7: Referenzverteilung erstellen (Z. 325-334)

325 M: Und jetzt? (...) Jetzt bin ich verwirrt. Jetzt können wir die Punkte doch eigentlich auch noch mal in einen Plot ziehen, oder?
326 I: Probier doch.
327 A: Also einen Graph, ne?
328 M: Schön. Ja.
329 *Die Studentin öffnet einen neuen Graph.*



330 A: So. Und dann/.
331 M: / Jetzt ordnen wir die mal. Und stapeln die irgendwie.
332 *Die Studentin zieht die Daten auseinander und stapelt sie.*



333 M: Müssen die so noch geordnet werden?
334 A: Nein.

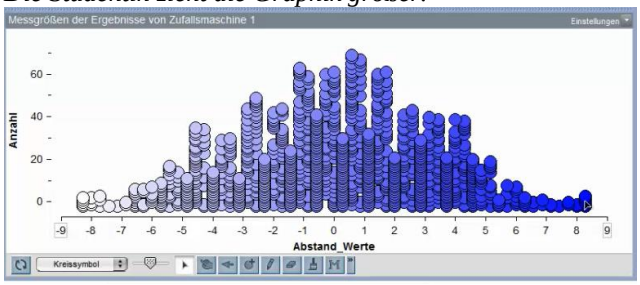
Nachdem die Messgrößen gesammelt sind, drückt Mandy direkt aus, dass sie verwirrt sei (325), allerdings wird nicht klar, worauf sie sich hier bezieht. Direkt anschließend werden die gesammelten Messgrößen von Mandy als Punkte interpretiert (bevor sie als solche in einem Graph dargestellt sind) und somit in die Softwaresprache übertragen (325). Möglicherweise ist sie verwirrt, wie sie mit den gesammelten Messgrößen weiter arbeiten soll, trotzdem schlägt sie vor, eine Graphik zu erstellen. Vielleicht hat sie als Strategie für das Arbeiten mit TinkerPlots verinnerlicht, zu gegebenen Daten eine Graphik zu erstellen, und hat hier die Hoffnung, dass ihr durch das Darstellen der Werte in einer Graphik wieder einfällt, wie sie diese auswerten kann. Von der Interviewerin wird sie durch „Probier doch.“ (326) dazu ermuntert, ohne dass hiermit eine konkrete Anweisung verbunden ist und Alisa führt die Aktion in TinkerPlots durch (327-329). Nachdem die Werte ungeordnet in einem Graphen vorliegen, beginnt Alisa: „So. Und dann/“ (330), als würde sie die weiteren Aktivitäten, die sie nun durchführen möchte, zuerst be- oder zumindest aussprechen wollen. Sie wird jedoch von Mandy unterbrochen, die vorschlägt, zu ordnen und zu stapeln (331) und sich somit einer TinkerPlots-Sprache bedient. Die Werte werden vollständig separiert und gestapelt (332), aber nicht geordnet, was an dieser Stelle auch keine sinnvolle weitere Aktivität wäre. Mandy fragt direkt nach: „Müssen die so noch geordnet werden?“ (333), was Alisa mit „Nein“ (334) beantwortet. Mit Blick auf die in (332) erstellte Graphik scheint somit Alisa eher klar zu sein, wie es weiter gehen soll (was auch durch ihre spätere Aussage in (337) bestätigt wird). Kommunikation auf Statistik- oder Kontextebene hat während dieser Episode nicht stattgefunden.

Interaktionseinheit 10: Ermitteln des P-Werts (Z. 335-399)

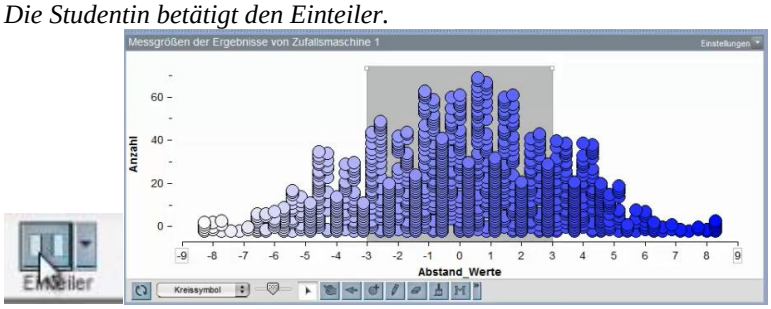
Die Interaktionseinheit, in der der P-Wert ermittelt wird, wird in vier Episoden gegliedert, von denen jedoch nur die erste und die letzte direkt mit dem Ermitteln des P-Werts zu tun haben. In der zweiten Episode werden erste, knappe Schlussfolgerungen gezogen und in der dritten Episode beschäftigen sich Mandy und Alisa mit dem Ausfüllen des Randomisierungstestschemas zum Punkt *4) Teststatistik*. In der letzten Episode dieser Einheit geht es um das Ausfüllen des Schemas für den Punkt *5) P-Wert*.

Episode 10-1: Ermitteln des P-Werts in TinkerPlots (Z. 335-356)

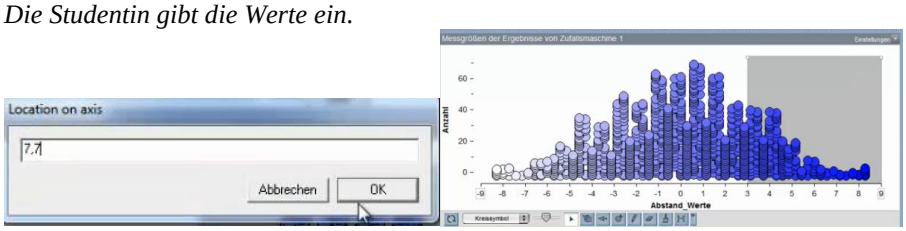
335 M: Gut. Und was war jetzt unsere ursprüngliche Differenz? 7,7?
336 *Die Studentin zieht die Graphik größer.*



337 A: Genau, das war dieser typische Verlauf.
338 M: Das (unv.)
339 I: Einteiler.
340 M: Einteiler heißt es. Ach da.
341 *Die Studentin betätigt den Einteiler.*

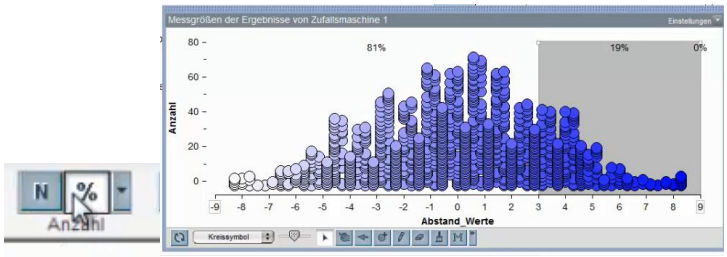


342 M: Von 7,7 bis Ende offen.
343 A: Warte mal. 7,7 und Ende ganz nach da?
344 M: Ja.
345 *Die Studentin gibt die Werte ein.*



346 M: Und dann Prozentzahl anzeigen.
347 A: Anzahl der Einteilung. Nee.
348 *Das Fenster „Anzahl der Einteilungen“ wird geöffnet und direkt wieder geschlossen.*

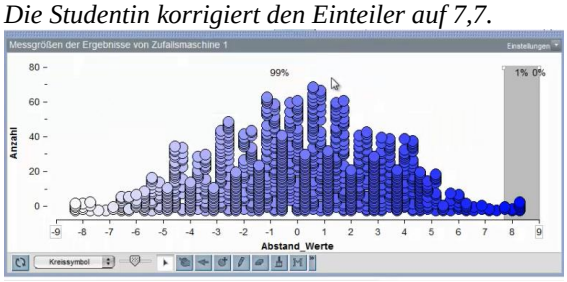
349 M: Einfach auf Prozente.
350 A: Ach da.
351 *Die Studentin klickt auf das Symbol für Prozente und diese werden angezeigt.*



352 A: 19 Prozent.

353 M: (ungläubig) 19 Prozent? (.) Aber jetzt ist der irgendwie wieder verrutscht. Guck mal, weil das/.

354 *Die Studentin korrigiert den Einteiler auf 7,7.*



355 A: 1 Prozent.

356 M: Das hört sich schon ein bisschen besser an.

Mandy eröffnet diese Episode mit der Frage „Und was war jetzt unsere ursprüngliche Differenz? 7,7?“ (335). Somit kann interpretiert werden, dass sie weiß, dass sie die simulierten Werte in Bezug zum ursprünglich beobachteten Wert setzen muss. Alisa übernimmt den Turn: „Genau, das war dieser typische Verlauf.“ (336), wobei nicht klar ist, ob sich das „Genau“ zu Beginn ihres Turns als Antwort auf Mandys Frage bezieht oder ob dies eine Bewertung der Verteilung ist, die sie sieht. „Das war dieser typische Verlauf“, deutet daraufhin, dass sie eine bestimmte Erwartungshaltung hat, die sie durch die vorliegende Verteilungsform bestätigt sieht. Die im Baustein Inferenzstatistik erstellten Referenzverteilungen sahen ähnlich aus, auf diese bezieht sich Alisa hier vermutlich. Alternativ könnte sie auch den Verlauf des bisherigen Lösungsprozesses meinen, allerdings ist dies mit Blick auf die Platzierung dieser Äußerung direkt nach dem Erstellen der Referenzverteilung eher unwahrscheinlich. Mandy scheint daraufhin auf den Einteiler hinzuweisen (338), dies kann jedoch aufgrund der leisen Sprechweise nicht rekonstruiert werden, allerdings scheint die Interviewerin dies in der Situation ebenso gedeutet zu haben, denn sie wirft „Einteiler.“ (339) als einzelnes Wort ein. Mandys Aussage in (340) bestätigt die Interpretation, dass sie in (338) nach der Bezeichnung gesucht hat. Der Einteiler wird angeklickt und Mandy gibt die Anweisung zur Bedienung „Von 7,7 bis Ende offen“ (342), was sich sicherlich auf das Verschieben der grauen Fläche des Einteilers bezieht. Alisa fragt noch einmal genau nach „Warte mal. 7,7 und Ende ganz nach da?“ (343), anscheinend ist sie nicht ganz so sicher, an welche Stelle der Einteiler platziert werden soll. Bis hierher hat keine von Beiden darüber gesprochen, dass mit Hilfe des Einteilers der P-Wert ermittelt werden soll und welcher Anteil (alle Werte ab 7,7 und größer) dazu in Betracht gezogen werden muss. Mandy ist

sich jedoch sicher, denn sie antwortet mit „Ja.“ (344), und der Wert 7,7 wird für das untere Ende des Einteilers eingegeben. Da jedoch dadurch der untere Wert über dem bisherigen oberen Wert des Einteilers liegt, verschiebt sich der graue Bereich und misst nun alles ab dem Wert 3, was von den beiden Studentinnen jedoch zunächst nicht bemerkt wird. Mandy gibt weiterhin Anweisungen („Prozentzahl anzeigen“ (346)), die von Alisa nach kurzem Suchen nach dem richtigen Button ausgeführt werden (347-351). Am Ende nennt Alisa die durch den Einteiler ermittelte Prozentzahl „19 Prozent“ (352), ohne dies zu kommentieren. Mandy ergreift daraufhin den Turn und fragt in einer ungläubigen Stimmlage: „19 Prozent?“ (353), so als würde sie diesen Wert anzweifeln. Nach einer kurzen Pause, in der sie vermutlich noch einmal auf die Referenzverteilung schaut, meint sie: „Aber jetzt ist der irgendwie wieder verrutscht. Guck mal, weil das.“ (353). Sie meint hier den Einteiler, der wie oben beschrieben nicht ab dem eingegebenen Wert 7,7 misst, sondern bereits ab 3. Ohne den Satz zu Ende zu führen, wird der Einteiler auf 7,7 korrigiert (354) und erneut nennt Alisa den nun angezeigten Wert: „1 Prozent“ (355). Eine Bewertung hierzu nimmt Mandy vor: „Das hört sich schon ein bisschen besser an.“ (356). Offenbar hat sie auch zur Größe des P-Werts, der, nebenbei bemerkt, immer noch nicht so bezeichnet wurde, wie bereits zuvor zur Form der Referenzverteilung, eine Vorstellung, mit der der nun vorliegende Wert in Einklang steht.

Episode 10-3: Erste Schlussfolgerungen aus dem P-Wert (Z. 357-361)

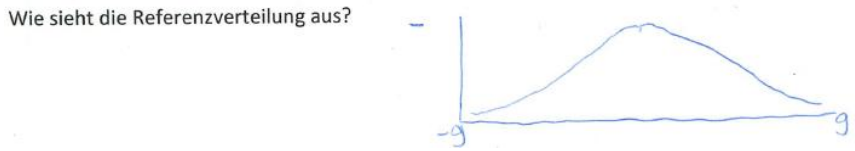
357	A: Genau und dann ist es das schon mit dem P-Wert mit der Signifikanz? Wenn die unter/.
358	M: /10 Prozent ist. Dann ist gut. (...)
359	A&M: (Lachen)
360	M: Wenn wir nicht in der Medizin sind.
361	A: Ja, dann nicht.

Nachdem der P-Wert in der vorigen Episode mit 1% ermittelt wurde, bezieht sich nun Alisa zum ersten Mal sprachlich darauf „Genau und dann ist es das schon mit dem P-Wert mit der Signifikanz? Wenn die unter/.“ (357). Sie nennt hier auf statistischer Ebene den P-Wert und verbindet Signifikanz damit, möglicherweise hervorgerufen durch das Randomisierungstestschema, in welchem *P-Wert* und *Signifikanz* als Stichwörter genannt sind. Alisa scheint zu versuchen, sich an Inhalte aus dem Baustein Inferenzstatistik zu erinnern, hier an die Grenze, ab der ein P-Wert als signifikant eingestuft wurde, wird jedoch von Mandy unterbrochen, die diese Grenze parat hat und direkt eine Einschätzung abgibt: „10 Prozent. Dann ist gut (…).“ (358). Nach einem kurzen gemeinsamen Auflachen fügt Mandy noch hinzu: „Wenn wir nicht in der Medizin sind.“ (360), weil sie sich wohl erinnert, dass im Baustein Inferenzstatistik die Grenze von 10 Prozent als nicht fest vorgeschrieben dargestellt wurde und dass in der Medizin niedrigere Grenzen, wie beispielsweise 0,01 oder 0,001 üblich seien. Weitere Schlussfolgerungen werden hier jedoch nicht gezogen, auch im Kontext werden keine Schlussfolgerungen an dieser Stelle formuliert.

Episode 10-3: Ausfüllen des Schemas zur Teststatistik (Z. 362-386)

362 M: Ok. Jetzt müssen wir das nur noch in Worte fassen.
363 A: Also welcher Wert soll überprüft werden? Der Durchschnittswert.
364 I: Ein bisschen genauer.
365 M: Der Durchschnittswert der neuen Vergleichsgruppen.
366 I: Was habt ihr denn gemacht?
367 M: Die Differenz der Durchschnittswerte der beiden Gruppen.
368 A: Die Differenz der Durchschnittswerte der beiden Gruppen. Der?
369 *Es wird notiert:*

370 M: Der neuen Gruppen einfach? Weil es waren ja jetzt im Prinzip nicht nur 2 Gruppen, sondern/.
371 A:/ ja.
372 M: Der neuen Gruppen einfach.
373 A: Ja. Das muss halt/ dass es immer 2 sind, die man vergleicht.
374 M: Also kann man das einfach sagen der beiden Gruppen? (..) Ok.
375 A: Gut.
376 M: Die beiden neuen Gruppen.
377 A: Ja die sind ja alle neu. Ja aber wir vergleichen ja auch das mit dem da, oder (*zeigt auf das Aufgabenblatt*)? Welcher Wert soll überprüft werden?
378 M: Nur der beiden Gruppen.
379 A: Ja, weil das/. Ok. Wie sieht die Referenzverteilung aus?
380 I: Da machst du einfach eine grobe Skizze. Einfach eine grobe Achseneinteilung. Dass, wenn man das nachher noch anguckt, man eine grobe Idee hat. Dass man dann sieht, wie sieht sie denn aus, von wo bis wo.
381 M: Ok. Minus 9 bis 9 ungefähr.
382 A: Ja und dann zur 0 hin hoch und dann runter wieder.
383 M: Zur Mitte hin hoch (unv.)
384 A: Ja so ungefähr.
385 *Es wird skizziert:*



386 M: Sieht gut aus. Muss noch irgendwas dabei?

In dieser Zwischenepisode beschäftigen sich Mandy und Alisa mit dem Ausfüllen des Schemas, rückgreifend, da dies noch nicht geschehen ist, mit der Teststatistik. Offensichtlich haben sie das Gefühl, mit dem gerade ermittelten P-Wert von einem Prozent eine Art Abschluss der Arbeiten in TinkerPlots gefunden zu haben, so dass sie sich nun dem Schema zuwenden. Alisa schlägt für die Teststatistik, deren auf dem Schema notierte Frage sie noch einmal vorliest, „Der Durchschnittswert.“ (363) vor. Hier übernimmt die Interviewerin den Turn und greift mit „Ein bisschen genauer.“ (364) ein. Mandy versucht daraufhin zu präzisieren: „Der Durchschnittswert der neuen Vergleichsgruppen.“ (365), woraufhin erneut die Interviewerin den Turn übernimmt und die Frage stellt „Was habt ihr denn gemacht?“ (366). Daraufhin entwickelt sich das Gespräch zwischen Mandy und Alisa weiter. Mandy wird nun noch präziser und nennt

„Die Differenz der Durchschnittswerte der beiden Gruppen“ (367), was so als Teststatistik verschriftlicht wird. Es folgt eine Diskussion darüber, ob es „der beiden Gruppen“ oder „der neuen Gruppen“ heißen soll (370-378). Als Argument für „die beiden Gruppen“ liefert Alisa, „dass es immer zwei sind, die man vergleicht“ (373) und hat damit im Blick, dass die Gruppenzusammensetzung immer verschieden sein kann. Anschließend gehen sie über zu der Frage, wie die Referenzverteilung aussehe (379). Hier gibt die Interviewerin den Hinweis, dass eine grobe Skizze gemacht werden solle (380). Daraufhin ergreift Mandy den Turn und schlägt als Grenzen der X-Achse -9 bis 9 vor (381), was von Alisa ergänzt wird mit „Ja und dann zur 0 hin hoch und dann wieder runter.“ (381). So wird es von den Beiden auf dem Randomsierungstestschema umgesetzt (383-385).

Episode 10-4: Ausfüllen des Schemas für den P-Wert (Z. 387-399)

387 A: Ok. Wie groß ist der P-Wert?
388 M: 1 Prozent.
389 A: Und dann das Ergebnis ist signifikant.
390 M: 99 Prozent.
391 A: Nee. Signifikant auf dem 1 Prozent-Niveau.
392 M: So rum stimmt. Man durfte das dann so rum machen. Reicht das, wenn man P-Wert hinschreibt oder sollen wir noch zur Signifikanz da jetzt noch was schreiben?
393 A: Ja ist signifikant, oder?
394 M: Oder sollen wir das erst bei Schlussfolgerungen schreiben. Dass es da offenbar ein signifikanter Wert ist?
395 A: Aber das wissen wir ja jetzt schon. Dass er signifikant ist. Also/.
396 I: Es ist definitiv ein signifikanter Wert. Kannst du einfach signifikant hinschreiben.
397 A: Ok.
398 *Es wird notiert:*

Alisa leitet die Episode ein mit der Frage „Wie groß ist der P-Wert?“ (387). Vermutlich nimmt sie die Anregung des Schemas auf und liest die dort gestellte Frage vor, wodurch sie den Turn an Mandy übergibt. Diese nimmt ihn an und antwortet kurz: „1 Prozent.“ (388), was dem entspricht, was die Beiden vorher in TinkerPlots ermittelt haben. Eine Ausformulierung, welche Wahrscheinlichkeit der P-Wert beschreibt, wie im Baustein Inferenzstatistik gelernt, findet nicht statt. Alisa übernimmt direkt wieder den Turn und sagt: „Und dann ist das Ergebnis signifikant.“ (389), womit sie sich vermutlich auf das nächste Stichwort des Randomisierungstestschemas bezieht und daran erinnert, dass in der Veranstaltung die Norm gesetzt wurde, dass P-Werte kleiner gleich zehn Prozent als signifikant betrachtet werden. Mandy versucht zu ergänzen „99 Prozent.“ (390), was interpretiert werden kann entweder als Wahrscheinlichkeitsangabe, die sie zu machen versucht, in dem Sinne, dass der P-Wert mit einer Wahrscheinlichkeit von 99% signifikant ist, was eine falsche Einschätzung darstellen würde; oder es kann interpretiert werden, dass sie einfach nur den zweiten Prozentwert vorliest, der im nicht grau markierten

Bereich des Einteilers angezeigt wird. Gegen diese Interpretation spricht jedoch der Zusammenhang, in dem die Aussage hervorgebracht wird. Alisa greift hier ein und sagt „Nee. Signifikant auf dem ein Prozent Niveau.“ (391), womit sie Mandys Aussage so interpretiert, dass diese eine Aussage über das Niveau des P-Werts machen wollte und diese korrigiert. Alisas Interpretation ist hier richtig, so wurde dies gelernt. Mandy stimmt dem im nächsten Turn zu: „So rum stimmt. Man durfte das dann so rum machen…“ (393), womit sie indirekt zugibt, dass sie selbst eine Fehlinterpretation geliefert hat und die Prozentangaben für die Bewertung des P-Werts verwechselt hat. Weiter fragt sie in diesem Turn, ob es reiche, den P-Wert aufzuschreiben, oder ob noch etwas zur Signifikanz geschrieben werden solle. Hierauf übernimmt wieder Alisa den Turn und fragt zurück: „Ja ist signifikant, oder?“ (394), in dem Wissen, dass sie dies gerade herausgearbeitet hat. Hier scheint eine Diskrepanz zwischen den Beiden zu herrschen, denn Alisa interpretiert hier den P-Wert korrekt bezüglich der Signifikanz, Mandy scheint ihr jedoch nicht folgen zu können, wie ihre Fragen und Aussagen in dieser Episode zeigen. Als nächstes fragt Mandy „Oder sollen wir das erst bei Schlussfolgerungen schreiben. Dass es da offenbar ein signifikanter Wert ist?“ (395). In diesem Sinn ist die Aussage über Signifikanz keine Schlussfolgerung, weil es hierzu eine feste Grenze gibt, ab der ein P-Wert im Baustein als signifikant kategorisiert wurde. Auch das Wort „offenbar“ könnte in die Richtung gedeutet werden, dass Mandy mit dem Konzept Signifikanz nicht ganz sicher ist. Alisa antwortet dementsprechend „Aber das wissen wir ja jetzt schon. Dass er signifikant ist. Also.“ (396). Um die Diskussion abzukürzen, übernimmt nun die Interviewerin den Turn und bestätigt Alisa:„Es ist definitiv ein signifikanter Wert. Kannst du einfach signifikant hinschreiben.“ (397), woraufhin dies in das Randomisierungstestschema notiert wird. Beide Teilnehmerinnen gehen nicht darauf ein, dass es sich durch die Simulation um einen Schätzwert für den P-Wert handelt, obwohl dies im Baustein Inferenzstatistik thematisiert wurde. Auch ein Bezug zum Kontext wird hier nicht hergestellt, somit findet keine Kommunikation auf der Kontextebene statt.

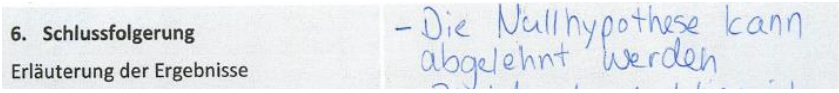
Interaktionseinheit 11: Schlussfolgerungen ziehen (Z. 399-436)

In der elften Interaktionseinheit werden Schlussfolgerungen von Mandy und Alisa gezogen. Dabei orientieren sich die Beiden an den Stichworten des Randomisierungstestschemas und daraus ergibt sich die Gliederung dieser Einheit in drei Episoden. In der ersten Episode werden Schlussfolgerungen auf statistischer Ebene gezogen ohne Bezug zum Kontext, in der zweiten Episode werden kausale Schlussfolgerungen gezogen und direkt auf die Kontextebene gewechselt und in der dritten Episode diskutieren Mandy und Alisa über Schlussfolgerungen bezogen auf die Population.

Episode 11-1: Schlussfolgerungen auf statistischer Ebene (Z. 399-408)

399 M: Ok. (...) Schlussfolgerungen zu.

400 A: Erläuterung der Ergebnisse.
401 M: Ja. Damit können wir jetzt die Nullhypothese zu 99-prozentiger Sicherheit ablehnen.
402 A: Nee 95/ nee das war was anderes.
403 M: Richtig.
404 I: Das Problem ist, eine Wahrscheinlichkeitsangabe kannst du da gar nicht machen. Du kannst erst mal sagen du kannst sie ablehnen. Und dann können wir uns überlegen was heißt das denn eigentlich, was ihr da habt?
405 A: Ok also erst mal würde ich schreiben wir lehnen die ab.
406 M: Also die Nullhypothese/.
407 M&A: /Kann abgelehnt werden.
408 *Es wird notiert:*



Mandy leitet die Interaktionseinheit mit „Schlussfolgerungen zu.“ (399) ein und Alisa liest die ersten Stichwörter des Randomisierungstestschemas vor: „Erläuterung der Ergebnisse.“ (400). Dann übernimmt wieder Mandy den Turn und schließt auf Statistikebene „Damit können wir jetzt die Nullhypothese zu 99 prozentiger Sicherheit ablehnen.“ (401). Hier zeigt sich, wie in der Episode zuvor, die Fehlvorstellung von Mandy, dass sie versucht, eine Hypothesenwahrscheinlichkeit anzugeben, denn im Umkehrschluss würde ihre Aussage bedeuten, dass sie interpretiert, dass die Nullhypothese zu 99 Prozent falsch ist. Positiv kann interpretiert werden, dass sie richtig schließt, dass die Nullhypothese abgelehnt werden kann. Alisa scheint das Problem zu erkennen, denn sie reagiert darauf mit: „Nee 95/ nee das war was anderes.“ (402). Vermutlich hat die Redewendung „99-prozentige Sicherheit“ von Mandy im Turn zuvor bei ihr die Erinnerung wachgerufen an die sprachliche Norm, die in der Veranstaltung zum $1/\sqrt{n}$- Gesetz gelegt wurde, mit *die tatsächlich gesuchte Wahrscheinlichkeit liegt mit 95 prozentiger Sicherheit im Intervall* $[h_n\text{-}1/\sqrt{n};\ h_n\text{+}1/\sqrt{n}]$ *um die simulierte Häufigkeit*, was jedoch einem anderen Inhaltsbereich zuzuordnen ist. Sie korrigiert sich selbst mit „nee das war was anderes“, was zeigt, dass sie diese Formulierung nicht unbedingt exakt parat hat, aber zumindest weiß, dass sie nicht zur Interpretation eines P-Werts gehört. Mandy bestätigt in (403), was gedeutet werden kann, dass ihr vielleicht bewusst wird, dass ihre in (401) gelieferte Schlussfolgerung nicht korrekt ist, oder dass sich die Formulierung „mit 95-prozentiger Sicherheit“ auf etwas anderes als den hier vorliegenden Inhalt bezieht. Die Interviewerin ergreift hiernach den Turn und stellt richtig, dass eine Wahrscheinlichkeitsangabe nicht gemacht werden könne, sondern nur gesagt werden könne, dass die Nullhypothese abgelehnt werde. Sie stellt die Frage, was das heiße, was hier vorliege (404). Alisa antwortet daraufhin, dass sie die ablehnen würde (405) und vermeidet so richtigerweise jede Angabe einer Wahrscheinlichkeit, mit der sie diese Schlussfolgerung zieht. Mandy präzisiert: „Also die Nullhypothese.“ (406). Und dies wird korrekt notiert als „Die Nullhypothese kann abgelehnt werden“ (408).

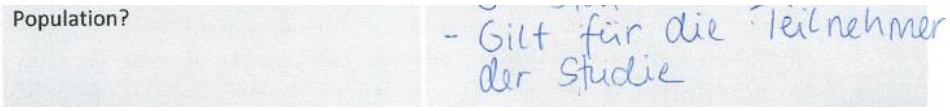
Episode 11-2: Kausale Schlussfolgerungen (Z. 409-421)

409 A: Kausal. Irgendwas mit Grund.
410 M: Das war, wodrauf schließen wir denn, woran liegt’s denn?
411 A: Dass wir das ablehnen, oder?
412 M: Nee was wir beobachtet haben. Gibt es einen kausalen Zusammenhang?
413 A: Ja, dass das Fischöl mehr bringt.
414 M: Ja offenbar dass es doch am Fischöl liegt.
415 A: Effektiver ist.
416 M: Ähm (...) Blutdruckreduktion liegt am Fischöl.
417 A: Ja liegt am/ wie hatten wir das hier genannt? An der Art der Öl-Diät.
418 M: Ok. (...) Dann wäre es abhängig von der Art der Öl-Diät?
419 A: Ja. (...)
420 *Es wird notiert:*

Alisa wendet sich dem nächsten Stichwort des Schemas zu: „Kausal. Irgendwas mit Grund.“ (409) und ihr scheint nicht ganz klar zu sein, worauf sich das Stichwort Kausal bezieht. Mandy übernimmt daraufhin den Turn und sagt erläuternd: „Das war, wodrauf schließen wir denn, woran liegt’s denn?“ (410), was eine gute Umschreibung ist für die Art der Schlussfolgerung, die hier gezogen werden soll. Alisa fragt noch einmal nach: „Dass wir das ablehnen, oder?“ (411), als sei ihr nicht ganz klar, worauf sie sich beziehen soll. Mandy entgegnet „Nee was wir beobachtet haben. Gibt es einen kausalen Zusammenhang?“ (412) und formuliert nun mit „Gibt es einen kausalen Zusammenhang?“ genau die richtige Frage, die beantwortet werden soll. Alisa beantwortet diese Frage direkt auf Kontextebene mit „Ja, dass das Fischöl mehr bringt.“ (413). Sie geht also nicht mehr auf die Ebene ein, auf der die beiden bisher zu klären versucht haben, welche Frage hier zu beantworten ist oder welche Art Schlussfolgerung hier gezogen werden soll, sondern hat offenbar die von Mandy in (412) formulierte Frage als korrekt angenommen und beantwortet diese nun inhaltlich. Alisa zieht den Schluss, dass „Fischöl mehr bringt“ und Mandy bestätigt das mit „Ja offenbar dass es doch am Fischöl liegt.“ (414). Das Wort „doch“ lässt hier die Interpretation zu, dass Mandy möglicherweise daran denkt, dass die Nullhypothese, also die zufällige Zuordnung zu den Gruppen, als Erklärung im Hintergrund steht, unter der simuliert wurde und die hier als wahr vorausgesetzt wurde und meint nun, dass es nicht an dieser zufälligen Zuordnung, sondern „doch“ am Fischöl liegt. Von (415) bis (420) suchen Mandy und Alisa gemeinsam nach der Formulierung, wie sie das aufschreiben können und einigen sich auf „Blutdruckreduktion ist abhängig von der Art der Öl-Diät“ (421). Interessant ist hier die Wortwahl „abhängig“, die von Mandy in (418) ins Spiel gebracht wurde als Frage „Dann wäre es abhängig von der Art der Öl-Diät?“, was eine gute Umformulierung darstellt, weil die Abhängigkeit der Wirkung von der im ursprünglichen Experiment durchgeführten Behandlung aufgenommen wird.

Episode 11-3: Schlussfolgerungen bezüglich der Population (Z. 421-437)

421 M: Population war das mit den Teilnehmern (unv.)
422 I: Genau die Frage ist, für wen gilt das?
423 A: Also hier jetzt für Männer.
424 M: Mit hohem Blutdruck.
425 A: Mit hohem/.
426 M:/ hier steht Hochdruck.
427 I: Druckfehler.
428 A: Also männliche Bluthochdruckpatienten. Dann müsste man ja theoretisch noch sagen wo. Also eher geographisch. Aber das ist ja/.
429 I: Genau also entweder sagt man. Für diese 14 Leute scheint das so zu sein. Und die Frage ist: Kann man das noch WEITER verallgemeinern und dann wird es halt immer schwierig. Weil wir hier von Freiwilligen reden und nicht von einer Zufallsstichprobe oder irgendwas. Deshalb immer diese Frage. Population für wen gilt das?
430 A: Ok. Für die Teilnehmer dieser Studie.
431 M: Ja. Reicht wirklich für die Teilnehmer der Studie oder GILT für die Teilnehmer der Studie?
432 A: Ich würde gilt dazu.
433 M: Gilt.
434 *Es wird notiert:*



435 M: Ok?
436 A: Ja.
437 *********Ende der selbständigen Bearbeitung**********

Mandy greift nun das letzte Stichwort des Randomisierungstestschemas auf „Population war das mit den Teilnehmern.“ (421). Die Interviewerin hat das als Frage an sie aufgefasst, denn sie antwortet direkt „Genau die Frage ist, für wen gilt das?“ (422), womit sie direkt präzisiert, dass Mandy und Alisa in den Blick nehmen sollen, für wen ihre Schlussfolgerungen gelten. Alisa antwortet daraufhin auf Kontextebene kurz: „Also hier jetzt für Männer.“ (423), was zeigt, dass sie das Design des Experiments im Kopf hat, da sie nicht erst im Aufgabentext nach dieser Information suchen muss. Sie schränkt somit in einem ersten Schritt die Schlussfolgerungen auf die Gruppe der Männer ein. Mandy schränkt dies im nächsten Turn noch weiter ein: „Mit hohem Blutdruck.“ (424), womit sie Alisas Aussage quasi fortführt und somit bestätigt. Alisa formuliert schließlich: „Also männliche Bluthochdruckpatienten.“ (428) und versucht weiter einzuschränken: „Dann müsste man ja theoretisch noch sagen wo. Also eher geographisch. Aber das ist ja.“ Dies scheint die Erinnerung an die letzte Vorlesung zu sein, in der thematisiert wurde, auf welche Personen die Ergebnisse eines Randomisierungstests bezogen werden dürfen und ein Teilaspekt war hier eine Auswahl von Personen für ein Experiment nach geographischen Gesichtspunkten, was jedoch auf die vorliegende Situation nicht zutrifft. Die Interviewerin übernimmt den Turn und stellt heraus, dass es zunächst für diese 14 Personen gilt und weiterhin, dass eine Verallgemeinerung über die Stichprobe hinaus schwierig sei, wenn es sich nicht um eine Zufallsstichprobe handele. Sie endet mit der Frage „Population für wen gilt das?“ (429), was von Alisa beantwortet wird mit

„Ok. Für die Teilnehmer dieser Studie.“ (430), was an dieser Stelle jedoch nicht als eigenständige Interpretation gewertet werden kann, weil es durch die Interviewerin im Turn zuvor quasi vorgegeben wurde. Somit notieren die Beiden (431-434) „Gilt für die Teilnehmer der Studie“ und beenden damit die Durchführung des Randomisierungstests.

Interaktionseinheit 12: Zusammenhang zwischen den formulierten Hypothesen und den möglichen Erklärungen (Z. 437-441)

438	I: Prima. Dann gehen wir das jetzt nochmal durch anhand von ein paar Fragen, die ich habe.
439	A: Ok.
440	I: So fangen wir nochmal vorne an. Könnt ihr bitte nochmal die Forschungshypothese und die Nullhypothese vorlesen, die ihr gewählt habt?
441	A: Die Forschungshypothese ist: Die Fischöldiät ist effektiver um den Blutdruck zu senken als die normale Öl-Diät. Und die Nullhypothese ist: Es spielt keine Rolle welche Öl-Diät gemacht wird, um den Blutdruck zu senken, sondern die Ergebnisse liegen an der zufälligen Gruppeneinteilung.
442	I: Genau da habt ihr jeweils schon eine Erklärung mit drin. Wir hatten ja am Anfang überlegt, welche Erklärungen könnten denn dahinter stecken? Und diese zwei finden sich da genau wieder. Einmal die Wirksamkeit des Fischöls als Erklärung in der Forschungshypothese und als Erklärung dass es an der zufälligen Einteilung der Teilnehmer auf den Gruppen liegt in der Nullhypothese. (..) …

In dieser Interaktionseinheit beginnt der dritte Teil des Interviews entlang der Leitfragen. Nachdem die Forschungs- und die Nullhypothese, die von Mandy und Alisa notiert wurden, vorgelesen wurden (441), geht die Interviewerin selbst auf die Frage A1 des Leitfadens ein und formuliert die Erklärungen, die jeweils hinter den beiden Hypothesen stehen, da diese ihrer Meinung nach direkt in die Formulierungen gesteckt wurden.

Interaktionseinheit 13: Beschreiben der Zufallsmaschine (Z. 442-487)

In der dreizehnten Interaktionseinheit wird im Interview beschrieben, was in der Zufallsmaschine in TinkerPlots geschieht und es wird nach einzelnen Einstellungen gefragt. Dazu wird die Einheit in vier Episoden gegliedert. In der ersten Episode wird die händische Simulation reflektiert, die Mandy und Alisa durchgeführt haben. In der zweiten Episode erklären die beiden Teilnehmerinnen die Zufallsmaschine. In der dritten Episode wird noch einmal auf das erfolgte Umbenennen der Merkmale eingegangen und in der vierten Episode wird die Bedeutung der zufälligen Zuordnung noch einmal aufgegriffen.

Episode 13-1: Reflektieren der händischen Simulation (Z. 442-454)

442	I: … (..) Dann habt ihr erst TinkerPlots ausprobiert und seid dann doch auf die Zettel zurückgegangen. (...) Wie seid ihr damit zurechtgekommen?
443	A: Mit den Zetteln?
444	I: Mhm (bejahend).
445	A: Eigentlich relativ gut.
446	M: Ja. Das mit den Zetteln das war ja/.
447	A: / Ja es ist ja auch relativ einfach, weil man keine technischen Sachen dabei hat. Man muss die

Zahlen/ oder die Werte einfach nur aufschreiben und dann in die zwei Gruppen einteilen.
448 I: Da war euch also auch klar, was ihr zu machen habt?
449 M: Ja. Wie war das. Das hatten wir kurz vergessen, oder?
450 A: Die Differenz.
451 M: Die Differenz. Dass wir die berechnen mussten. Aber ich denk/ obwohl du hast mich doch darauf aufmerksam gemacht, oder?
452 A: Ja.
453 M: Also eigentlich war jeder Schritt klar.
454 A: Würde ich auch sagen.

Der Eindruck, den die Beiden vermitteln, ist, wie sie in (445) und (446) ausdrücken, dass sie mit der händischen Simulation gut zurecht gekommen sind. Alisa nennt als Grund in (447), „weil man keine technischen Sachen dabei hat“ und führt den Prozess aus, der durchgeführt werden muss als „Man muss die Zahlen oder die Werte einfach nur aufschreiben und dann in die zwei Gruppen einteilen.“, was zwar eine verkürzte Darstellung ist, da sie beispielsweise außen vorlässt, dass zunächst angenommen werden muss, dass die Werte unabhängig vom zuvor eingenommen Öl betrachtet werden müssen und dass die Einteilung in die zwei neuen Gruppen zufällig erfolgen muss, aber dennoch richtig zusammengefasst ist. Sie stellen noch fest, dass sie kurz vergessen hatten, dass die Differenz zu berechnen war, kommen aber zu dem Schluss, „Also eigentlich war jeder Schritt klar.“ (454). An dieser Stelle ist jedoch keinem aufgefallen, dass die händische Simulation nur einmal durchgeführt wurde und das Eintragen der Differenz der Mittelwerte der neuen Gruppen in einen Graph nicht erfolgt ist. Dennoch passt die Selbsteinschätzung der Teilnehmerinnen mit der in dieser Analyse gemachten zusammen, dass die händische Simulation gut gelang.

Episode 13-2: Erklären der Zufallsmaschine (Z. 455-469)

455 I: Mhm (bejahend). (...) So dann kommen wir noch mal auf die Zufallsmaschine. Kannst du die nochmal aufmachen, die ist jetzt wahrscheinlich irgendwo klein versteckt.
456 *Die Studentin vergrößert die Zufallsmaschine.*

457 I: Oh sehr gut. Gleich gefunden. Ok könnt ihr bitte nochmal erklären, was in der Zufallsmaschine genau passiert?
458 A: Ja also erst mal hat man die erste Urne. Da sind alle Werte drin, die wir haben.
459 I: Was für Werte?
460 A: Von der normalen Öl-Diät und von der Fischöl-Diät. Also von den Teilnehmern die Werte. Und dann gibt es eine zweite Urne, wo die Gruppeneinteilung von A und B mit drin sind mit dem Stapelbutton.

461 I: Mit dem Stapelbauteil. Genau.
462 A: Ja dann hat man die Durchgänge 14, weil es 14 Werte sind. (..) 2 Ziehungen, weil das lässt sich nicht ändern.
463 I: Warum sind es denn zwei Ziehungen?
464 A: Wegen der 2 Urnen.
465 I: Genau aber was passiert denn?
466 A: Also achso. Ein Wert wird rausgezogen aus der ersten Urne und dann zufällig in A oder B eingeteilt. Ja das sind zwei Ziehungen.
467 I: Genau. Und was kriegt ihr dann hinterher raus?
468 M: Ja zufällig neue eingeteilte Gruppen. Erst mal. Ja.
469 A: Ja.

Mit Bezug zum Kontext erläutert Alisa die beiden Bauteile der Zufallsmaschine (458) und (461). Dabei beschreibt sie, welche Elemente sich jeweils mit Bezug zur Softwareebene in dem Bauteil befinden („Alle Werte von der normalen Öl Diät und von der Fischöldiät. … Und wo die Gruppeneinteilung von A und B mit drin sind“). Sie erklärt, dass 14 Durchgänge benötigt werden, weil es 14 Werte gebe (462), wobei sie nicht darauf eingeht, dass die ursprünglichen Gruppengrößen reproduziert werden sollen. Dass zwei Ziehungen gemacht werden, begründet sie damit, dass sich diese Einstellung nicht ändern lässt, was auf der technischen Ebene korrekt ist, auf der statistischen Ebene allerdings die wichtige Bedeutung hat, dass jeweils einem Blutdruckwert eine neue Gruppe zugeordnet wird. Zu der Ziehungsanzahl gibt es dann Nachfragen der Interviewerin (463, 465 und 467), die erst in (468) in einer Erklärung von Mandy enden, die auf statistischer Ebene operiert: „Ja zufällig neue eingeteilte Gruppen.“ Insgesamt erscheint die Beschreibung der Zufallsmaschine auf einer technischen Ebene zu geschehen, das, was gesehen ist, wird beschrieben, aber darüber hinausgehenden Erläuterungen, beziehungsweise Erläuterungen auf der statistischen oder der kontextuellen Ebene werden nicht, oder nur aufgrund stetiger Nachfragen durch die Interviewerin formuliert.

Episode 11-3: Reflektieren des Umbenennens der Merkmale (Z. 470-475)

470 I: Ok. Ihr habt die Merkmale umbenannt, richtig?
471 A: Genau. In Werte und neue Gruppen.
472 I: Ja. Warum?
473 A: Damit es ein bisschen anschaulicher ist.
474 I: Mhm (bejahend).
475 M: Ja.

Auf die Nachfrage der Interviewerin, warum die Merkmale umbenannt wurden (472), lautet Alisas Antwort: „Damit es ein bisschen anschaulicher ist.“ (473). Auch wenn dies hier nur kurz erklärt wird, wird aus der entsprechenden Episode deutlich, dass das Umbenennen Schwierigkeiten vorgebeugt hat.

Episode 11-4: Klären der Bedeutung des zufälligen Zuordnens (Z. 476-487)

476 I: In Ordnung. Und (..) Was bedeutet das hier jetzt, dass man die Werte zufällig auf neue Gruppen aufteilt?
477 M: Naja, wenn wir jetzt in der Nullhypothese sagen, dass es eigentlich egal ist zu welcher Gruppe man ursprünglich gehörte, dann hätten wir ja/ also dann/ (..) / dann wird es ja auch keine Rol-

le spielen, wenn wir die einfach mischen und in neue Gruppen aufteilen, weil dann, wenn wir jetzt weitergehen. Die Differenz sich ja im Prinzip nicht großartig unterscheiden musste, sondern im Endeffekt immer ähnlich groß sein dürfte. Hm.

478 I: Du machst so hmm?

479 M: Ja das war jetzt so ein bisschen weit ausgeholt. Ich hab das Gefühl es ist noch nicht ganz auf den Punkt gebracht. Aber es ist was ich meine.

480 I: Mhm (bejahend). Also wir waren bei der Frage was es bedeutet, dass man die Werte zufällig auf zwei neue Gruppen aufteilt und du hast gesagt/ also was ich daraus interpretieren würde. Was in eurer Nullhypothese steckt ist, dass es egal ist, ob man Fischöl oder normales Öl genommen hat. Man hat einfach die Werte und guckt, was passiert, wenn ich die jetzt zufällig zwei Gruppen zuordne. Welche Differenzen von Durchschnittswerten, von Mittelwerten bekomme ich dann. Ja?

481 M: Ja.

482 I: Ja das zu formulieren ist immer ein bisschen schwierig. Die Frage ist: Schwebt das so ungefähr in euren Köpfen?

483 A: Ja man geht immer nach der Nullhypothese, richtig?

484 I: Was hast du denn in die Zufallsmaschine reingesteckt?

485 A: Ja. Ja wegen der zufälligen Einteilung. Also das hab ich ja da gemacht und deswegen oder wie? Geht man nach der Nullhypothese.

486 I: Ja?

487 A:Ja.

Diese Episode ist geprägt von der Frage der Interviewerin, was es bedeute, dass die Werte zufällig auf neue Gruppen aufgeteilt werden (476). Auf statistischer Ebene argumentiert Mandy mit der Nullhypothese, in der ausgedrückt wird, „dass es eigentlich egal ist zu welcher Gruppe man ursprünglich gehörte…“ (477) und weiter mit „dann wird es ja auch keine Rolle spielen, wenn wir die einfach mischen und in neue Gruppen aufteilen…“. Somit beschreibt sie gut das, was in der Nullhypothese angenommen wird und wie dies für den Randomisierungstest zugrunde gelegt wird. Von der Interviewerin wird dies noch einmal einschließlich der Teststatistik formuliert (480) und Mandy stimmt dem zu (481). Auch Alisa stellt noch einmal fest, „man geht immer nach der Nullhypothese, richtig?“ (483), was ihr nicht so ganz sicher erscheint, da sie fragend mit „richtig?“ endet. Die Interviewerin begegnet ihr mit einer Gegenfrage: „Was hast du denn in die Zufallsmaschine reingesteckt?“ (486), woraufhin Alisa nun erneut die zufällige Einteilung nennt und mit „Geht man nach der Nullhypothese“ (485) vermutlich meint, dass man die Nullhypothese zugrunde legt, beziehungsweise als wahr voraussetzt.

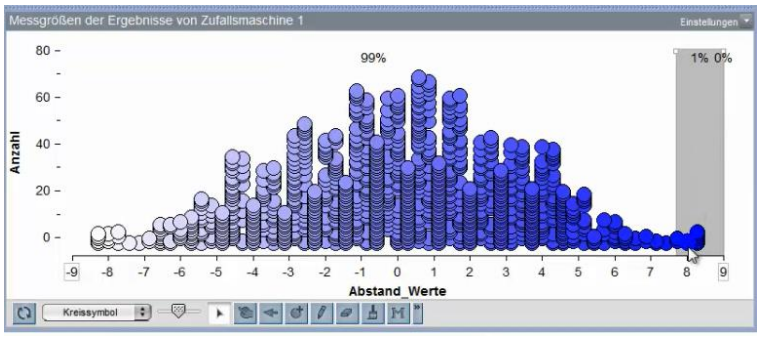
Interaktionseinheit 14: Verständnis über einzelne Punkte der Referenzverteilung (Z. 488-512)

Den Fragen des Leitfadens unter A3 bezogen auf die Referenzverteilung wird in der 14. Interaktionseinheit nachgegangen. Die Einheit wird in zwei Episoden gegliedert. In der ersten wird erläutert, was einzelne Punkte der Referenzverteilung bedeuten, dies können Mandy und Alisa direkt nennen mit Bezug auf die Teststatistik und wird deshalb hier nicht berichtet (488-495). In der zweiten Episode wird der Frage A3d nachgegangen, ob ein Fall mit dem Wert 8 der Referenzverteilung überraschend sei. Hieran schließt sich

ein längerer Dialog an, in welchem erarbeitet wird, was der P-Wert von einem Prozent bedeutet. Hierzu der folgende Transkriptausschnitt (496-512).

496 I: Wenn ihr jetzt dahinten guckt in dem grauen Bereich, der mit dem Einteiler markiert ist. Wenn ihr da so einen Wert nehmt, sagen wir mal irgendwo bei der 8. Was ist da passiert?

497 *Die Studentin markiert einen solchen Wert.*



498 M: Ja wieder das gleiche. Also bei irgendeiner zufälligen Zuordnung war die Differenz halt eben 8 groß.

499 I: Und wie beurteilt ihr das in Bezug auf die Nullhypothese? (...) Ist das überraschend, dieser Punkt bei der 8?

500 M: Ist das überraschend, ja. (...) Ich weiß es ehrlich gesagt gerade nicht.

501 A: Ich weiß nicht, ob man das irgendwie so sieht, dass es überraschend wäre, wenn es mehr von diesen/ in diesem Bereich gäbe. Wenn dieser Einteiler einen größeren Bereich umfassen würde. Aber an sich, wenn es eine zufällige Einteilung ist, muss es ja genauso drin stecken wie die Null zum Beispiel. Oder wie im Minusbereich oder so.

502 I: Mhm (bejahend). (...) Und was bedeutet das in Bezug auf die Daten, die hier im Experiment erhoben wurden? (...) Du guckst mich so ganz fragend an.

503 A: Ja ich weiß nicht/ ich hab das nicht ganz verstanden.

504 M: Also (..) das bedeutet ja jetzt erst mal, wenn wir uns nur diesen 1%- Bereich angucken, dass wir nur in einem Prozent unserer ganzen Neueinteilung eine Differenz haben die größer ist als 7,7. Was das jetzt aber wirklich für unsere Schlussfolgerungen bedeutet. Also da fehlt mir immer so der Zwischenschritt muss ich sagen.

505 I: Ok. Können wir ja gemeinsam erarbeiten, aber das war schon mal ganz wichtig. Das heißt was du gerade gesagt hast: In einem Prozent der Fälle ist es bei dieser zufälligen Zuordnung passiert, dass man einen Wert wie den beobachteten von 7,7 oder noch größer bekommen hat.

506 A: Mhm (bejahend).

507 I: Das heißt in einem Prozent der Fälle kann das durch die zufällige Zuordnung geschehen. Das man so einen Wert wie hier bekommt.

508 M: Jetzt hab ich auch glaube ich. Weil je größer die Differenz ist, desto besser ist es ja eigentlich für die Studie. Weil dann ja. Jetzt hab ich auch warum es wichtig ist. Also glaub ich. Ich versuche es nochmal in Worte zu fassen. Also je größer die Differenz zwischen den beiden Gruppen ist/.

509 I: /Zwischen den Ursprünglichen/.

510 M:/ Genau. Zwischen den ursprünglichen Gruppen, desto valider ist ja quasi erst mal das Ergebnis. Und wenn wir jetzt sagen das kommt quasi so, wenn wir die Gruppen mischen nie vor, außer in einem Prozent der Fälle. Dann bedeutet das, dass die Nullhypothese halt so nicht gültig ist. Ja.

511 I: Genau die ist einfach nicht mehr plausibel, wenn ich mir das angucke. In einem Prozent der Fälle kann es aber trotzdem sein. Das heißt ihr habt euch jetzt dafür entschieden die Nullhypothese abzulehnen. In einem Prozent der Fälle hättet ihr euch falsch entschieden. Denn es kann sein, dass auch bei einer zufälligen Einteilung ein solches Ergebnis vorkommt. Aber es ist so selten. Nur in einem Prozent der Fälle. Und das findet ihr zu selten deswegen lehnt ihr das als Begründung ab. Das steckt so dahinter was ihr hier gemacht habt.

512 A: Ja.

Zunächst wird von Mandy erklärt, dass ein Fall über der Acht bedeutet, dass „bei irgendeiner zufälligen Zuordnung war die Differenz halt eben acht groß.“ (498). Danach schließt die Interviewerin die Frage an, ob dieser Punkt bei der Acht überraschend sei (499). Mandy übernimmt den Turn, zögert jedoch bei der Antwort und formuliert, dass es ihr schwer fällt zu antworten (500). Daraufhin übernimmt Alisa den Turn und konstruiert die fiktive Situation, wenn es mehr Fälle im Bereich ab 7,7 gäbe. Mit „wenn dieser Einteiler einen größeren Bereich umfassen würde“ (501) ist vermutlich nicht gemeint, dass der Bereich vergrößert wird, also der Anteil der Fälle ab einem niedrigeren Wert bestimmt werden sollte, sondern eher, dass fiktiv die Verteilung anders aussähe und deshalb in den jetzigen Bereich des Einteilers mehr Fälle fallen würden. Prinzipiell scheint sie jedoch nicht überrascht zu sein, denn es müsse „ja genauso drin stecken wie die Null zum Beispiel. Oder wie im Minusbereich oder so.“ (501). Die Interviewerin stellt hierauf eine erneute Frage (502), die von Alisa nicht verstanden wird (503) und zu der Mandy eine Antwort versucht (504). Sie erläutert in ihrem Turn, was der „1 Prozent Bereich“ bedeute, nämlich „dass wir nur in einem Prozent unserer ganzen Neueinteilung eine Differenz haben die größer ist als 7,7…“ (504). Dies ist bis auf den kleinen Fehler, dass der Bereich auch die Werte umfasst, die genau 7,7 sind, eine gute Erklärung des P-Werts und zeigt, dass sie dessen Bedeutung wiedergeben kann. Allerdings formuliert sie am Ende ihres Turns noch Schwierigkeiten, die sie damit hat, das nun auf die Schlussfolgerungen zu beziehen und nennt dies einen „Zwischenschritt“, der ihr fehle. Die Interviewerin übernimmt den Turn und bietet an, dies gemeinsam zu erarbeiten. Sie formuliert noch einmal aus, was der P-Wert hier umfasst (505) und bezieht das auf die Erklärung der zufälligen Zuordnung (507). Daraufhin ergreift Mandy den Turn und sagt „… Weil je größer die Differenz ist, desto besser ist es ja eigentlich für die Studie. Weil dann ja. … je größer die Differenz zwischen den beiden Gruppen ist“ (508) und „… desto valider ist ja quasi erst mal das Ergebnis. Und wenn wir jetzt sagen das kommt quasi so, wenn wir die Gruppen mischen nie vor, außer in einem Prozent der Fälle. Dann bedeutet das, dass die Nullhypothese halt so nicht gültig ist. Ja.“ (510). Aus der Art, wie Mandy hier spricht, lässt sich interpretieren, dass sie hier einen grundlegenden Gedankengang entwickelt hat, der ihr Verständnis bezeugt, dass sie nun aufgebaut hat. Die Interviewerin geht darauf noch einmal ein und führt das weiter, denn sie ergänzt, dass durch das Ablehnen der Nullhypothese in der vorliegenden Situation in einem Prozent aller Fälle eine falsche Entscheidung getroffen würde und führt noch einmal den Gedankengang aus, der im Idealfall zur Ablehnung der Nullhypothese geführt hat (511). Somit ist in der gemeinsamen Interaktion von Mandy, Alisa und der Interviewerin hier im Gespräch die Bedeutung des P-Werts von einem Prozent hervorgebracht worden.

Interaktionseinheit 15: Schlussfolgerungen ziehen (Z. 513-538)

In der letzten Interaktionseinheit des Interviews wird den drei Fragen A6, A7 und A9 in drei Episoden nachgegangen. Da bereits ausführlich über den P-Wert von 1% gesprochen wurde, wurde die Frage A8 im Interview ausgelassen.

Episode 15-1: Hat Fischöl eine blutdrucksenkende Wirkung? (Z. 513-515)

513	I: Da gehen wir jetzt auch gleich nochmal hin. Jetzt kommen wir nämlich zu dem Bereich Schlussfolgern. Das ist ja eigentlich das Wichtige. Da haben wir das gerade schon/ was bedeutete es, dass hier ein signifikantes Ergebnis mit einem P-Wert von einem Prozent erhalten wurde? (..) Das haben wir gerade besprochen. Dass wir nämlich nur ein Prozent der Fälle haben, wo unter der zufälligen Einteilung ein solches Ergebnis kommt. (..) So, was sagt ihr denn jetzt. Lässt sich jetzt sagen, dass Fischöl eine blutdrucksenkende Wirkung bei den Versuchspersonen hat?
514	A: Ich würde sagen bei den 14 Personen, die wir hier haben: Ja.
515	M: Ja.

Auf die Frage der Interviewerin, ob Fischöl eine blutdrucksenkende Wirkung bei den Versuchspersonen habe (513), antworten beiden Teilnehmerinnen mit „Ja.“ (514-515), wobei Alisa es auf die 14 Versuchspersonen einschränkt. Beide geben hier keinen Anlass für weitere Diskussionen und somit folgt direkt die nächste Episode.

Episode 15-2: Sicherheit bezüglich der Schlussfolgerungen (Z. 516-529)

516	I: Wie sicher seid ihr euch damit?
517	M: Das können wir glaube ich nicht sagen. Also wir haben ja jetzt nur die Nullhypothese schon mal ausgeschlossen. Also dass es nicht an der zufälligen Einteilung liegt/.
518	A: / Aber falsch ist sie ja trotzdem nicht.
519	M: Ja.
520	I: Warum nicht?
521	A: Weil es immer noch/ weil ein Prozent der (..) also weil zu einem Prozent das noch zutrifft/.
522	M:/ möglich ist, dass es stimmt.
523	A: Ja.
524	I: Wenn ihr so eine Einschätzung macht: Sehr sicher, ein bisschen sicher, ziemlich unsicher?
525	A: Ich würde sagen, wie wir das in der Vorlesung hatten, dass der P-Wert bis 10 Prozent sein kann, dann würde ich sagen sind wir uns sehr sicher, oder?
526	M: Ja. Sehr sicher.
527	A: Wenn das ein Prozent hier ist.
528	I: Mhm (bejahend). Weil das eben nur ein Prozent ist?
529	A: Ja.

Auf die Frage der Interviewerin, wie sicher die Beiden bezüglich der vorigen Schlussfolgerung, dass Fischöl eine blutdrucksenkende Wirkung habe, seien (516), antwortet zunächst Mandy: „Das können wir glaube ich nicht sagen. Also wir haben ja jetzt nur die Nullhypothese schon mal ausgeschlossen…“ (517). Es lässt sich interpretieren, dass sie möglicherweise vorsichtig geworden ist, eine Wahrscheinlichkeitsangabe bezüglich ihrer Sicherheit zu machen und deshalb sagt, dass sie dies nicht angeben könne. Sie formuliert korrekt, dass die zufällige Einteilung als Erklärung abgelehnt wurde, worauf ihr Alisa ins Wort fällt mit „Aber falsch ist sie ja trotzdem nicht“ (518). Sie bezieht sich vermutlich auf die vorherige Diskussion zum P-Wert von einem Prozent, durch den die

Nullhypothese nicht als falsch betrachtet werden kann. Dies formuliert sie auf eine entsprechende Nachfrage der Interviewerin in (512): „Weil es immer noch/weil ein Prozent der (..) also weil zu einem Prozent das noch zutrifft.“, und verhaspelt sich sprachlich. Mandy übernimmt den Turn und fügt an „möglich ist, dass es stimmt.“ (522). Somit haben die beiden eine korrekte Erklärung geliefert, dass die Nullhypothese abgelehnt werden kann, jedoch nicht als komplett falsch ausgeschlossen werden darf mit Bezug auf den P-Wert von einem Prozent. Die Interviewerin kommt nun noch einmal auf die Frage zurück und gibt eine mögliche Kategorisierung vor für die Einschätzung der Sicherheit bezüglich dieser Schlussfolgerung (524). Alisa antwortet und nennt die in der Vorlesung genannte Grenze von 10 Prozent als Begründung, weshalb sie sich bei einem P-Wert von einem Prozent sehr sicher ist bezüglich der Schlussfolgerung (525). Mandy bestätigt dies (526) und die Nachfrage der Interviewerin, weil der P-Wert nur ein Prozent sei (528) wird von Alisa bejaht (529). Damit endet das Interview.

11.4.2 TinkerPlots-Benutzung zur Simulation von Mandy und Alisa

Aus den vorherigen Analysen wird in diesem Abschnitt entlang den in Kapitel 10.4.4 vorgestellten Kategorien analysiert, welche der notwendigen Schritte in TinkerPlots in den Interaktionseinheiten 8, 9 und 10 von Mandy und Alisa selbständig, mit leichter Intervention oder mit starker Intervention durch die Interviewerin durchgeführt werden.

Tabelle 40: Notwendige Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Mandy und Alisa

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen der Zufallsmaschine	-	Ja	-
Vorhandene Kugeln der Zufallsmaschine löschen	-	-	Ja
Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren	-	-	Ja
Ein zweites Bauteil der Zufallsmaschine hinzufügen	-	-	Ja
Zweites Bauteil füllen	-	-	Ja
Die Einstellung ohne Zurücklegen bei beiden Bauteilen auswählen	-	-	Ja
Durchgangsanzahl festlegen	-	-	Ja

Mandy und Alisa formulieren vorab Schwierigkeiten, die sie bei der Benutzung von TinkerPlots zur Simulation erwarten. Diese bestätigen sich im weiteren Verlauf. Sie führen zunächst eine händische Simulation durch, hierzu arbeiten sie selbständig ohne Intervention durch die Interviewerin. Bei der Übertragung der händischen Simulation in die Software benötigen sie auf Softwareebene viele Hilfestellungen durch die Interviewerin. Auf statistischer Ebene scheint der Prozess klarer zu sein, in der Sprache der Software hingegen weniger. Bei nahezu allen erforderlichen Schritten benötigen sie genaue Anweisungen für die Bedienung, die sie dann jedoch flüssig ausführen können.

Tabelle 41: Optionale Schritte in TinkerPlots zur Erstellung der Zufallsmaschine von Mandy und Alisa

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einstellen der Geschwindigkeit	-	-	Ja
Umbenennen der Merkmale Merkm1 und Merkm2	-	Ja	-

Die Geschwindigkeit wird von Mandy und Alisa nicht verändert, was beim Messgrößensammeln später zu Problemen führt und von der Interviewerin behoben wird.

Tabelle 42: Notwendige Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Mandy und Alisa

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen einer Graphik zur Simulation	-	Ja	-
Das erste Merkmal auswählen und darstellen	Ja	-	-
Das zweite Merkmal auswählen und in der Graphik passend darstellen	Ja	-	-
Die Mittelwerte einblenden	Ja	-	-
Differenz der Mittelwerte mit Hilfe der Linealfunktion messen	-	-	Ja
Messgröße definieren	-	-	Ja
Anzahl der zu sammelnden Messgrößen eintragen	-	Ja	-
Messgrößen sammeln	-	Ja	-
Erstellen einer Graphik zu den gesammelten Messgrößen	-	Ja	-
Referenzverteilung darstellen	Ja	-	-

Bezüglich des Arbeitens mit den simulierten Ergebnissen sind die beiden Teilnehmerinnen wesentlich sicherer als beim Arbeiten mit der Zufallsmaschine. Sie benötigen den Hinweis, dass eine Graphik gebraucht wird, um die Ergebnisse darzustellen, danach können sie selbständig weiterarbeiten. Die Bedienung des Lineals stellt eine Schwierigkeit dar, weshalb starke Interventionen durch die Interviewerin vonnöten sind, ebenso wie für das Definieren der Messgröße. Beim Erstellen der Graphik für die Referenzverteilung wird lediglich eine leichte Intervention durch eine Nachfrage an die Interviewerin provoziert, zum Darstellen der Referenzverteilung wird keine Hilfe benötigt.

Tabelle 43: Optionale Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung von Mandy und Alisa

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Fälle in der Graphik stapeln	Ja	-	-
Beschleunigen des Messgrößensammelns	-	-	Ja

Wie nach der Aufgabenanalyse erwartet, wird starke Hilfe benötigt, um den Messgrößensammelvorgang zu beschleunigen.

Tabelle 44: Notwendige Schritte in TinkerPlots zur Ermittlung des P-Werts von Mandy und Alisa

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einteiler verwenden	Ja	-	-
Prozente einblenden	Ja	-	-

Die Bedienung der Software, um den P-Wert zu finden, gelingt Mandy und Alisa gut.

Mandy und Alisa führen zunächst eine händische Simulation durch, deren Prozess sie nur mit Hilfe der Interviewerin in die Software übersetzen können. Das Bedienen und Einstellen der Zufallsmaschine stellt für sie eine Schwierigkeit dar. Anschließend können sie sukzessive selbständiger arbeiten und ihre Vorhaben in der Software umsetzen.

11.4.3 Fallzusammenfassung Mandy und Alisa

Bis hierher wurde der Bearbeitungsprozess von Mandy und Alisa ausführlich dargelegt und interpretiert. Hier wird nun die Analyse zusammengefasst im Hinblick auf besondere Auffälligkeiten (Voigt 1984).

Das Konzept der zufälligen Aufteilung der Versuchspersonen auf die Gruppen erklären Mandy und Alisa zu Beginn des Interviews kurz und knapp mit Bezug auf das Merkmal Haarfarbe, nach welchem die Gruppen hätten eingeteilt werden können, auf das bei einer solchen Aufteilung jedoch geschlossen werden müsste (Interaktionseinheit 2). Hier scheint ihnen das Konzept für das Design des Experiments klar zu sein, in der nächsten Interaktionseinheit wird es jedoch nicht als eine mögliche Erklärung von den Beiden genannt, auch auf Nachfrage der Interviewerin wird es nicht von den Beiden als eine mögliche Erklärung formuliert, da sie sich darauf versteifen, dass bei einem solch großen beobachteten Unterschied zwischen den Gruppen als mögliche Erklärungen nur die Wirksamkeit des Fischöls oder unbekannte konfundierende Variablen in Frage kommen. Für die Formulierung der Nullhypothese stellt sich dies jedoch nicht als Schwierigkeit heraus, denn hier formulieren sie die Nullhypothese mit explizitem Bezug auf die zufällige Aufteilung. Im weiteren Verlauf der Fallstudie wird deutlich, dass sie das Konzept und dessen Bedeutung für den Randomisierungstest verstanden haben. Sie validieren die von ihnen erstellte Zufallsmaschine im Hinblick darauf, ob sie die Werte tatsächlich zufällig auf die beiden neuen Gruppen aufteilt und auch bei der späteren Erläuterung der Zufallsmaschine wird deutlich, dass sie den Bezug zwischen der zufälligen Aufteilung und der Nullhypothese und ihrer Bedeutung für die Simulation herstellen. Auch wenn die zufällige Aufteilung nicht als mögliche Erklärung für die beobachteten Unterschiede in Interaktionseinheit 3 von Mandy und Alisa genannt wird, so scheinen sie das Konzept trotzdem verstanden zu haben und es an dieser Stelle möglicherweise nur nicht genannt zu haben, weil es ihnen als mögliche Erklärung aufgrund des großen beobachteten Unterschieds zwischen den Gruppen zu unplausibel erschien.

Null- und Forschungshypothese werden von Mandy und Alisa ohne Schwierigkeiten formuliert.

Für die Schlussfolgerungen orientieren sich Mandy und Alisa an den Stichworten des Randomisierungstestschemas und formulieren diese sowohl auf statistischer Ebene als auch im Kontext. Sie versuchen zunächst irrtümlich aus dem erhaltenen P-Wert auf eine Hypothesenwahrscheinlichkeit zu schließen, korrigieren sich aber selbst und bekommen auch noch Hinweise hierzu durch die Interviewerin. Daraufhin formulieren sie nur kurz, dass sie die Nullhypothese ablehnen und gehen dazu über, deshalb auf die Wirksamkeit des Fischöls zu schließen. Die Reichweite dieser Schlussfolgerung schränken sie mit Hilfe durch die Interviewerin auf die 14 Versuchspersonen ein, die am Experiment teilgenommen haben. Somit haben sie zwar sehr knapp, aber vollständig, alle Aspekte an Schlussfolgerungen korrekt in Betracht gezogen, die im Baustein Inferenzstatistik gelernt wurden.

Eine auffällige Episode dieses Paares ist die durchgeführte händische Simulation mit den Zetteln. Da Mandy und Alisa sehr unsicher bezüglich der Simulation in TinkerPlots sind, gehen sie auf den Vorschlag der Interviewerin ein, die Simulation zunächst händisch durchzuführen. Dabei erinnern sie sich offenbar an das Vorgehen im Seminar, denn hierzu findet recht wenig planerische Kommunikation statt und sie führen die einzelnen Schritte nacheinander aus. Sie notieren alle Blutdruckwerte auf einzelne Zettel, mischen und teilen diese auf zwei neue Gruppen auf, die sie „A“ und „B“ nennen. Die Berechnung der jeweiligen Mittelwerte nimmt etwas Zeit in Anspruch, letztlich wird die Teststatistik jedoch korrekt ermittelt. Allerdings ist auffällig, dass sie die händische Simulation nur für eine Neuaufteilung der Werte durchführen und somit nur einen Wert für die Teststatistik erhalten. Eine Auswertung bzw. Einschätzung des ursprünglichen Unterschieds mit dem durch die händische Simulation erhaltenen Wert geschieht nicht, sie gehen direkt dazu über, die zufällige Aufteilung der Werte auf zwei Gruppen mit der Software zu modellieren. Sie können nun genau beschreiben, was sie machen wollen, bei der Umsetzung in der Software benötigen sie dennoch viel Unterstützung durch die Interviewerin (vgl. 11.4.2). Da sie zu Beginn der Interaktionseinheit 8 jedoch überhaupt keinen Ansatz hatten, um weiter zu arbeiten, können sie nun zumindest den Prozess erklären, den sie simulieren möchten und auch die Bedeutung der einzelnen Schritte immer wieder erläutern. Die tatsächlich durchgeführte händische Neuzuordnung der Werte auf zwei neue Gruppen mit Hilfe von Zetteln scheint Mandy und Alisa für den weiteren Bearbeitungsprozess demnach stark unterstützt zu haben.

11.5 Vergleichende Analyse in Bezug auf die Unterforschungsfragen

Die durchgeführte Interaktionsanalyse offenbart einige Gemeinsamkeiten, aber auch deutliche Unterschiede zwischen den Bearbeitungsprozessen der Paare. Diese werden in

diesem Abschnitt im Hinblick auf die Unterforschungsfragen dargestellt. Die Beantwortung der Forschungsfragen und Folgerungen für die Lerneinheit zur Inferenzstatistik und allgemein für die Mathematikdidaktik finden sich im abschließenden Kapitel 12.

11.5.1 Welche Problemstellen treten auf? Welche davon sind unerwartet? Welche Gründe lassen sich identifizieren für die unerwarteten Problemstellen?

Insgesamt werden sieben Problemstellen aus den durchgeführten Paaranalysen in den Kapiteln 11.2., 11.3 und 11.4 identifiziert. Allerdings ergeben sich diese erst aus der Analyse aller drei Paare, denn nicht jede Schwierigkeit tritt bei jedem Paar auf. Es werden folgende Problemstellen identifiziert: *mögliche Erklärungen für beobachtete Unterschiede finden*, *Übertragen der Nullhypothese in die Zufallsmaschine in TinkerPlots*, *Messgrößensammeln beschleunigen*, *Referenzverteilung im Randomisierungstestschema skizzieren*, *Versuch der Angabe einer Hypothesenwahrscheinlichkeit*, *Schlussfolgerungen ziehen* und *Verknüpfung der Kontext- mit der Statistik- und der Softwareebene* identifiziert, die in unterschiedlichem Ausmaß bei allen Paaren auftreten.

Diese werden nun vorgestellt und es wird versucht, Gründe für deren Auftreten zu identifizieren.

Problemstelle „mögliche Erklärungen für beobachtete Unterschiede finden“

Bei dieser Problemstelle haben die Teilnehmer Schwierigkeiten, beide möglichen Erklärungen für die beobachteten Unterschiede zu finden. Die eine Erklärung, dass Fischöl wirksam ist, wird von allen genannt, aber die zweite Erklärung, dass die beobachteten Unterschiede durch die zufällige Aufteilung der Versuchspersonen auf die Gruppen des Experiments erklärt werden können, wird nur von zwei der drei Paare gefunden. Laura und Fabia können diese Erklärung nicht selbständig formulieren, selbst mit Hilfe der Interviewerin gelingt ihnen das nicht, so dass die Erklärung schließlich von der Interviewerin vorgegeben werden muss.

Diese Problemstelle findet sich ähnlich in den Ergebnissen bei Pfannkuch et al. (2015). Die Autoren berichten davon, dass ihre Studienteilnehmer nach der Lerneinheit Schwierigkeiten haben, die Wirksamkeit der Behandlung als mögliche Erklärung zu nennen. Im Unterschied birgt in der vorliegenden Studie die Erklärung über die zufällige Zuordnung für die Teilnehmer eine Schwierigkeit.

Dieser schon fast gegensätzliche Effekt von zwei unabhängigen Studien lässt darauf schließen, dass das Konzept der möglichen Erklärungen für sich eine Schwierigkeit darstellt. Da dies ein entscheidendes Konzept für den Randomisierungstest ist, da es die zwei möglichen Deutungsalternativen des beobachteten Unterschieds aufzeigt, sollte es sorgfältig in einer Lerneinheit adressiert werden und gibt somit einen ersten Hinweis auf ein mögliches Redesign (vgl. Kapitel 12.2.2.6).

Problemstelle „Übertragen der Nullhypothese in die Zufallsmaschine in TinkerPlots“

Die nächste Problemstelle, die identifiziert wird, ist das Modellieren der Nullhypothese in der Zufallsmaschine. Hier muss jedoch unterschieden werden zwischen dem Erklären des Modellierens ohne Bezug zur Software und der konkreten Übertragung in die Software. Zwei der drei Paare äußern hinsichtlich der allgemeinen Bedienung der Software zu erwartende Schwierigkeiten, alle drei Paare können jedoch den Prozess des Randomisierungstests allgemein gut erklären. Erst bei der konkreten Benutzung der Software und hier speziell bei der Erstellung der Zufallsmaschine treten Schwierigkeiten auf. Für die Erstellung der Zufallsmaschine treten bei allen drei Paaren Schwierigkeiten auf, jedoch an unterschiedlichen Stellen, die hier zusammengefasst werden zu einer Problemstelle.

Mandy und Alisa sind eins der Paare, die hier vorab Schwierigkeiten äußern, sie behelfen sich zunächst damit, sich den Prozess des Randomisierens durch eine händische Simulation in Erinnerung zu rufen mit dem erklärten Ziel, danach die Software (wieder) besser bedienen zu können. Bei der anschließenden Übertragung des durchgeführten händischen Vorgangs in die Software verbleiben jedoch trotzdem Schwierigkeiten für dieses Paar.

Details zu den einzelnen Aspekten dieser Problemstelle finden sich in Kapitel 11.5.5, in welchem die Softwarebenutzung durch die Teilnehmer genau analysiert wird und die einzelnen Schwierigkeiten in Bezug auf jede Einstellung dargelegt werden. Im Wesentlichen lassen sich zwei Stellen herauskristallisieren, bei denen alle drei Paare Schwierigkeiten haben. Die erste Stelle ist das Einfügen der Werte in das Bauteil Box. Dies hat etwas mit der Bedienung der Software zu tun und es ist erstaunlich, dass hierzu alle Paare Hilfe benötigen, da dieser Schritt mit einem einfachen Kopier- oder Abtippvorgang gelöst werden könnte (vgl. 10.4.4). Die zweite Stelle hingegen konnte man eher erwarten, da hier inhaltlich das Konzept der zufälligen Zuordnung angewendet werden muss, was bereits als mögliche Schwierigkeit in der a priori Analyse antizipiert wurde. Konkret handelt es sich bei dieser Problemstelle um die Schwierigkeit, die zufällige Zuteilung der Werte auf zwei neue Gruppen zu modellieren, auf Softwareebene also ein zweites Bauteil mit den Gruppenetiketten zu erstellen.

Die Problemstelle liegt jedoch nicht grundsätzlich im Verstehen des Vorgangs des Randomisierungstests, wie vor allem das Planen aller Paare im Vorfeld der Simulation zeigt. Vielmehr lassen sich Gründe in der nicht abgeschlossenen instrumentellen Genese der Paare hinsichtlich der *instrumented action schemes* vermuten. Vor allem die Tatsache, dass die Schwierigkeiten bezüglich vieler unterschiedlicher Aspekte der Zufallsmaschine auftreten und hierbei auch kaum ein Muster zu erkennen ist (vgl. Tabelle 46), lässt diese Vermutung zu. Auf der anderen Seite ist es auch denkbar, dass für den Vorgang des (Re-)randomisierens, also des Zuordnens der Werte auf zwei neue Gruppen, in

der Lerneinheit die drei Ebenen nicht genügend miteinander in Verbindung gebracht wurden und deshalb eine Problemstelle für alle Studienteilnehmer darstellt.

Problemstelle „Messgrößensammeln beschleunigen“

Mit dieser Problemstelle ist gemeint, dass alle Paare Hilfe benötigten bei der Beschleunigung des Messgrößenvorgangs und nicht, dass die Definition der Messgröße oder der dahinterstehenden Teststatistik eine Problemstelle darstellt. Diese Problemstelle wurde jedoch in der a priori Analyse erwartet. Da es sich um eine Problemstelle handelt, die rein auf technischer Ebene liegt, lässt sich hier erneut vermuten, dass die instrumentelle Genese der Teilnehmer nicht genügend ausgebildet war und nicht, dass das Konzept der Teststatistik, das mit dem Messgrößensammeln zusammenhängt, eine Schwierigkeit darstellt.

Problemstelle „Referenzverteilung im Randomisierungstestschema skizzieren“

Die handschriftlich anzufertigende Skizze der Referenzverteilung für das Randomisierungstestschema wird als nächste Problemstelle identifiziert. Alle Teilnehmer fertigten eine grobe Skizze der Referenzverteilung an, in der teilweise die Achsenbeschriftungen fehlten, vor allem aber immer der beobachtete Wert nicht eingezeichnet wurde, wodurch keine qualitative Einschätzung desselben aufgrund der Skizze möglich war. Von den Teilnehmern wurde dies nicht als Problemstelle empfunden, erst die nachträgliche Analyse der Transkripte offenbarte diese Problemstelle.

Diese Problemstelle wurde a priori nicht erwartet. Als Grund hierfür lässt sich vermuten, dass der „Sinn“ der Skizze nicht erfasst wurde, und deshalb dieser Punkt des Randomisierungstestschemas nur „abgearbeitet“ wurde. Auch hieraus ergeben sich Hinweise für ein mögliches Redesign der Lernumgebung (vgl. Kapitel 12.2.2.6).

Problemstelle „Versuch der Angabe einer Hypothesenwahrscheinlichkeit“

Es ist aus der Literatur bekannt, dass die Interpretation eines P-Werts Schwierigkeiten birgt. Eine solche Schwierigkeit ist hier aufgetreten, nämlich mit Hilfe des P-Werts eine Angabe zur Wahrscheinlichkeit der Nullhypothese oder auch der Forschungshypothese machen zu wollen. Rebecca und Selina thematisierten dies von sich aus und im Gespräch entwickelten sie gemeinsam, dass eine solche Angabe nicht möglich ist. Im Interviewteil zu diesem Thema ist jedoch auch bei ihnen, wie bei den anderen beiden Paaren, ein Rückfall in diesen Fehlschluss zu entdecken.

Ein Grund hierfür mag darin liegen, dass es sich bei der Interpretation eines P-Werts nach einer solch kurzen Lerneinheit um fragiles Wissen handelt und dass dies intensiv in einer Lerneinheit thematisiert werden muss.

Problemstelle „Schlussfolgerungen“

Die letzte Problemstelle, die identifiziert werden kann, sind mögliche Schlussfolgerungen, die aus den Ergebnissen des Randomisierungstests gezogen werden können. Von allen Paaren werden die Schlussfolgerungen nur äußerst knapp gezogen, obwohl mögliche Schlussfolgerungen ausführlich in der Lerneinheit thematisiert wurden und im Randomisierungstestschema einige Hinweise dazu zu finden sind. Dass mit einem P-Wert von 1% die Nullhypothese abgelehnt werden kann, können alle drei Paare formulieren, die Schwierigkeit liegt darin, darüber hinausgehende Schlussfolgerungen zu machen. Bezüglich der Forschungshypothese werden mögliche Schlussfolgerungen erst im Interviewteil thematisiert und auch hier eher zögerlich und nur knapp. Wie z. B. bei Madden (2008) stellt der konkrete Bezug zum Kontext hier eine Schwierigkeit dar.

Daraus abgeleitet ergibt sich die folgende Problemstelle, die auch als Erklärung dienen kann.

Problemstelle „Verknüpfung der Kontext- mit der Statistik- und der Softwareebene“

Vor allem in den Interaktionseinheiten, in denen Software verwendet wird, adressieren die Gesprächsteilnehmer in hohem Maße im Gespräch die Software, ziehen hierzu teilweise statistische Begriffe mit ein, beziehen dies jedoch nur äußerst selten auf den Kontext. Da sich gerade bei der Bedienung der Software jedoch Schwierigkeiten ergeben, wird diese nur gering ausgeprägte sprachliche Verknüpfung der drei Ebenen als weitere Problemstelle identifziert. Gleiches lässt sich beobachten, wenn beispielsweise über die Referenzverteilung oder den erhaltenen P-Wert gesprochen wird. Hier werden statistische Begriffe verwendet, die jedoch nicht mit dem konkreten Kontext in Verbindung gebracht werden. An mehreren Stellen in den einzelnen Paaranalysen ist aufgefallen, dass die Paare einige Schwierigkeiten durch den Bezug zum Kontext selbst überwinden konnten. Gerade auch für Aspekte der Software wird deshalb empfohlen, durch den Kontextbezug Erklärungsansätze zu finden. Ein konkretes Beispiel ist hier das Umbenennen der Merkmale der Zufallsmaschine. Hierüber haben alle Paare diskutiert und auch, wenn die Dikussion nicht im Umbenennen der Merkmale geendet hat, so hat sie doch zum Verständnis des Konzepts der zufälligen Zuordnung beigetragen.

Ein Grund, der hypothetisch für einen mangelnden Bezug vor allem zur Kontextebene verantwortlich gemacht werden kann, liegt im Design der gesamten Veranstaltung. Elemente des Designs beinhalten eine Verknüpfung der Kontext- mit der Statistik- und der Softwareebene. Allerdings soll die Software gewissermaßen als Mittel zum Zweck verwendet werden. Der Aufbau der instrumentellen Genese und das Verstehen der statistischen Konzepte benötigt jedoch einen hohen Anteil der Lernzeit (auch schon vor der Lerneinheit Inferenzstatistik) und so besteht das Risiko, dass sich die Studierenden auf die Software und die Statistikebene fokussieren und sprachlich sowie inhaltlich die

Kontextebene vernachlässigen. Eine Ursache hierfür kann darin vermutet werden, dass im Gegensatz zu vielen anderen Beispielen der gesamten Veranstaltung (auch in den anderen Bausteinen Datenanalyse, Kombinatorik und Wahrscheinlichkeitsrechnung) wesentlich weniger mit konkretem Bezug zum Kontext gearbeitet wird als dies in der vorliegenden Lerneinheit der Fall ist und deshalb von den Studierenden „gelernt“ wurde, den Kontext zugunsten der Statistikebene und der Softwareebene zu vernachlässigen. Auch hierin lassen sich Hinweise auf ein mögliches Redesign entdecken.

Dies sind die Problemstellen, die aus den Paaranalysen generalisiert werden konnten. Lokal lassen sich weitere Schwierigkeiten bei einzelnen Paaren entdecken, diese lassen sich jedoch nicht über alle Paare hinweg generalisieren.

11.5.2 Wie stark sind die Phasen der Gesprächsführung ausgeprägt? Wie gestalten die Paare die Kommunikation untereinander? Was bedeutet das in Bezug auf die kollaborativen Bearbeitungsprozesse?

In diesem Abschnitt werden Antworten gegeben auf die Unterforschungsfragen, die sich auf die Kommunikationsprozesse beziehen.

Es müssen zunächst durch den Aufbau des Interviews zwei Kommunikationssituationen unterschieden werden. Erstens gibt es die Situation, in welcher die Interviewerin Fragen stellt, die von den Teilnehmern beantwortet werden sollen. Dies ist im ersten und letzten Teil des Interviews der Fall und umfasst die Interaktionseinheiten 1 bis 4 und 12 bis 15. Das Gespräch ist in diesen Einheiten geprägt vom gleichberechtigten Dialog zwischen den drei Gesprächsteilnehmern und vor allem beide Studierende sollen Gelegenheit haben, sich hier zu Wort zu melden. Im mittleren Teil des Interviews (Interaktionseinheiten 5 bis 11) führen die Teilnehmer möglichst selbständig den Randomisierungstest durch, hier geschieht in erster Linie Interaktion zwischen den Teilnehmern und eine Gesprächsübernahme durch die Interviewerin stellt eine Intervention dar. Der Kommunikationsablauf ist hier selbstbestimmt durch die Teilnehmer.

Wie stark sind die Phasen der Gesprächsführung ausgeprägt?

Auf gesprächsanalytischer Makroebene zieht sich durch alle Interaktionseinheiten und Episoden während des selbständigen Durchführens des Randomisierungstests durch die Teilnehmer, dass es höchstens eine minimale Eröffnungsphase gibt, die maximal einen Turn umfasst. In vielen Fällen ist die Eröffnungsphase durch das Nennen des jeweiligen Stichworts des Randomisierungstestschemas gekennzeichnet, z. B. Mandy: „Ok. Schlussfolgerungen.“ (400) als Eröffnung der Interaktionseinheit 11, in der Schlussfolgerungen gezogen werden. Häufig erfolgt sogar ein direkter inhaltlicher Einstieg in einem Turn, aus dem die zweite Sprecherin schließen muss, dass nun das nächste Konzept adressiert wird (z. B. Laura (44): „Randomisierung ist ja zufällige Einteilung.“ als Eröffnungsphase der Episode, in der zu den beobachteten Daten diskutiert wird, ob im

Experiment eine Randomisierung der Teilnehmer auf die Gruppen stattgefunden hat). Das Ende einer Einheit ist häufig gekennzeichnet durch das Notieren der Ergebnisse im Randomisierungstestschema, ohne dass verbal das Ende einer Einheit eingeleitet wird. Aus Außensicht könnte man denken, dass solch kurze Überleitungs- bzw. Eröffnungsphasen und kein bewusstes Beenden von Gesprächseinheiten zu Problemen führt, allerdings ist dies bei keinem der drei Paare als Schwierigkeit aufgefallen. Die Gesprächslenkung und die Adressierung des nächsten Konzepts ergeben sich bei allen Paaren als natürlich im Kommunikationsprozess zwischen den Teilnehmern und bedürfen keines Aushandelns darüber, ob eine Gesprächs- bzw. Inhaltseinheit nun abgeschlossen ist oder welcher Schritt als nächstes durchzuführen ist. Dies gilt sowohl für die aufeinander eingespielten Paare (Laura und Fabia und Mandy und Alisa) als auch für das einander unbekannte Paar (Rebecca und Selina).

Wie gestalten die Paare die Kommunikation untereinander?

Es ist auf mikroanalytischer Gesprächsebene auffällig häufig festzustellen, dass alle sechs Teilnehmer in einem Turn wenige vollständige Sätze formulieren, sondern häufig Sätze anfangen, die entweder gar nicht oder von der Partnerin beendet werden. Dies ist in besonderem Maße während der Durchführung des Randomisierungstests der Fall, welcher in den Interaktionseinheiten 5 bis 11 stattfindet. Diese Art der Kommunikation zeigt ein hohes Maß an Interaktion zwischen den Paaren und lässt sich dahingehend interpretieren, dass sie den Gedankengängen der jeweiligen anderen aufmerksam folgen und ähnliche Gedanken haben, sonst könnten und würden sie die Sätze nicht in diesem Sinne vervollständigen oder ergänzen. Häufig finden sich zudem von der ersten Sprecherin, die einen Satz begonnen hat, bezüglich der Formulierungen im nächsten Turn durch die Partnerin im wiederum darauf folgenden Turn bestätigende Floskeln, die anzeigen, dass die erfolgte Ergänzung akzeptiert wird. Ein Beispiel hierfür ist der folgende Transkriptausschnitt zur Interaktionseinheit 5 des Paares Mandy und Alisa, in welchem Mandy zunächst die Interaktionseinheit eröffnet (23), Alisa daraufhin versucht, eine Formulierung der Forschungsfrage zu finden (24), die erst von Mandy ergänzt wird (25), woraufhin Alisa weiterformuliert (26) und dies schließlich von Mandy bestätigt wird (28).

23 M: Ok. Welche Frage soll beantwortet werden?
24 A: Ist Fischöl oder bringt Fischöl mehr oder senkt Fischöl den Blutdruck/.
25 M: /Effektiver als/.
26 A: /Normales Öl?
27 *Die Studentinnen notieren:*

28 M: Ok.

Dies ist ein typisches Beispiel der ablaufenden Kommunikation zwischen den Paaren, das sich an vielen Stellen strukturell ähnlich findet. Auf diese Art wird sehr häufig ge-

meinsames Wissen hervorgebracht. Mit „Art“ ist hier gemeint, dass schnelle Turnwechsel zwischen den Interaktanten stattfinden, bei denen nur einzelne Bestandteile einer Formulierung in einem Turn genannt werden. Diese Art der Kommunikation ist bei allen Paaren zu beobachten, wenn es um das Finden von Formulierungen für das Ausfüllen des Randomisierungstestschemas geht. Einerseits zeigt sich hier, dass das Formulieren des entsprechenden statistischen Konzepts zunächst eine Hürde darstellt, da nicht von einer Teilnehmerin allein für beide zufriedenstellend formuliert wird, andererseits zeigt dies aber auch, dass im gemeinsamen Interaktionsprozess eine zufriedenstellende (und bis auf eine Ausnahme auch eine korrekte) Formulierung des jeweiligen Konzepts gefunden wird. Das Paar Rebecca und Selina fällt hier am stärksten aus dem Rahmen, denn diese Beiden formulieren häufig bereits im ersten Turn vollständig eine Erklärung. Hierzu lassen sich zwei Hypothesen formulieren. Entweder lässt sich die vollständige Kommunikation von Rebecca und Selina damit erklären, dass diese beiden im Gegensatz zu den anderen Paaren miteinander nicht vertraut sind und deshalb versuchen, „vollständiger“ zu kommunizieren. Oder es lässt sich damit erklären, dass bei diesen beiden das Wissen und Verständnis zu den einzelnen Konzepten sehr gut ausgeprägt ist, so dass sie nicht erst nach richtigen Formulierungen suchen müssen, sondern diese direkt hervorbringen können.

Auch das Sprecherwechselsystem funktioniert sehr gut, es gibt, wenn überhaupt nur sehr kurze Überlappungen der Gesprächsbeiträge, das Rederecht ist somit fast immer eindeutig, wenn auch stillschweigend, geklärt. Dies zeigt, dass keins der Paare konfliktreiche Diskussionen führt, sondern in weiten Teilen Einigkeit über die Gesprächsinhalte herrscht.

Was bedeutet die bisherigen Erkenntnise in Bezug auf die kollaborativen Bearbeitungsprozesse?

Die Kommunikationsprozesse der hier analysierten Paare sind also einerseits geprägt durch ein hohes Verständnis der Interaktanten füreinander, das sich darin ausdrückt, dass es keine Schwierigkeiten gibt, dem Gesprächsverlauf zu folgen, auch wenn ein Wechsel zu einem neuen Thema nur durch sehr knappe Marker in der Eröffnungsphase angezeigt wird. Andererseits sind die Kommunikationsprozesse geprägt von kurzen Turns, in denen häufig nur einzelne Worte oder Satzfragmente geäußert werden, die jedoch wechselseitig zu vollständigen Sätzen und zu vollständigen Argumentationen von den Sprecherinnen ergänzt werden. Eine dritte Seite ist die nahezu konfliktfreie Gesprächsführung, womit gemeint ist, dass eine starke Fokussierung auf die Inhalte zu erkennen ist und kaum andere Themen (beispielsweise das Organisieren von Gesprächsbeiträgen) aufgegriffen werden.

Eine der wenigen Arbeiten, die ebenfalls kollaborative Bearbeitungsprozesse mathematischer Aufgaben in Kleingruppen unter Gesichtspunkten der Gesprächsführung unter-

suchen, ist die von Höck (2015). In dieser Studie werden Gesprächsprozesse zwischen Grundschülern analysiert, u. a. im Hinblick auf Kommunikationsprozesse. Bei Höck (2015, 217) beispielsweise sind Eröffnungsphasen ebenfalls häufig durch aufgabenbezogene Beiträge gekennzeichnet und gehen somit wie in der vorliegenden Analyse über reine Eröffnungsphasen hinaus. Im Unterschied zu Höck (2015) wurden in der vorliegenden Studie keine konfliktreichen Gesprächsteile ausgemacht, was sich möglicherweise mit dem Alter und der Reife der jeweiligen Teilnehmer erklären lässt. Der Übergang zwischen den einzelnen Gesprächsphasen wird von Höck (2015) als unproblematisch beschrieben, was auch für die vorliegende Studie zutrifft.

Aus den ähnlichen Ergebnissen der Analyse von Kommunikationsprozessen in Kleingruppen zwischen sehr unterschiedlichen Teilnehmern (Höck (2015) analysiert Dritt- und Viertklässler, in der vorliegenden Studie werden Studierende analysiert), lässt sich eine Theorie über kollaborative (mathematische) Kommunikationsprozesse in Kleingruppen entwickeln. Dies findet sich in Form von Merkmalen für kollaborative Kommunikationsprozesse im nächsten Abschnitt.

11.5.3 Arbeiten die Teilnehmer der Studie kollaborativ zusammen? An welchen Merkmalen lässt sich die Kollaboration erkennen?

Vor allem in den wechselseitigen Ergänzungen der Gesprächsbeiträge zwischen den Paaren zeigt sich die inhaltliche Vernetzung der eigenen Beiträge mit denen des Partners und somit eine starke Kollaboration der Teilnehmer untereinander. Je tiefer die Teilnehmer inhaltlich in ein Konzept einsteigen, desto mehr Interaktion ist zwischen ihnen zu beobachten. Dies scheint umso stärker zu gelingen, wenn sich die beiden Partner bezüglich ihres Wissensstands etwa auf Augenhöhe zu befinden scheinen, wie beispielsweise Rebecca und Selina. Somit ergeben sich zwei Arten von Merkmalen, an denen Kollaboration zu erkennen ist. Es gibt strukturelle Merkmale, die sich auf die Gesprächsführung beziehen, und inhaltliche Merkmale, die sich auf die jeweiligen Inhalte einer Interaktionseinheit beziehen, die Kollaboration erkennen lassen.

Strukturelle Merkmale für kollaborative Bearbeitungsprozesse sind:

- Stabile Eröffnungsphasen, die konfliktfrei eine neue Interaktionseinheit und damit ein neues Thema beginnen,
- Stabile Durchführungsphasen in Abgrenzung zu Eröffnungs- und Beendigungsphasen von Interaktionseinheiten, in denen ohne nennenswerte Abschweifungen über ein Thema kommuniziert wird,
- Geteilte Beendigungsphasen, in denen das „Endprodukt“ zu einem Thema als gemeinsam ausgehandelt erklärt wird,
- Häufige Turnwechsel, für die folgende Situationen unterschieden werden können:

  - Satz- oder Gedankenfortführung durch den jeweiligen Partner,
  - Gemeinsames Formulieren bestimmter Sachverhalte.

Inhaltsbezogene Merkmale, die eine Kollaboration erkennen lassen, sind:

- Fokussierung auf ein bestimmtes Thema innerhalb einer Interaktionseinheit,
- Inhaltliche Turnübernahme zur Fortführung oder Ergänzung eines vorherigen Sprecherbeitrags,
- Hervorbringen von als gemeinsam geteilt geltenden Bedeutungen,
- Hervorbringen von gemeinsam formulierten Argumentationen.

Gekennzeichnet sind die Beiträge der Paare darüber hinaus durch die wechselseitige Anerkennung der individuellen Beiträge. Die einzelnen Merkmale können dabei unterschiedlich ausgeprägt sein, so ist beispielsweise die Kommunikation von Rebecca und Selina eher durch vollständige Satzbildung in einzelnen Turns geprägt, was zu längeren Turns führt und die Kommunikation der anderen beiden Paare ist häufig durch unvollständige Sätze geprägt, was zu kurzen und teilweise mehr Turns in einer Einheit führt. Vollständige oder unvollständige Sätze in einem Turn können somit nicht als Merkmal kollaborativer Bearbeitungsprozesse abgeleitet werden.

11.5.4 Können die Teilnehmer die verschiedenen, in Kapitel 6.4 identifizierten Konzepte (einen Unterschied in den beobachteten Daten identifizieren, Nullhypothese aufstellen, etc.), die zur Argumentation eines Randomisierungstests gehören, verwenden und erläutern? Welche Schwierigkeiten treten dabei auf?

Im Folgenden werden Ähnlichkeiten und Unterschiede der drei Paare hinsichtlich der für einen Randomisierungstest benötigten Konzepte (vgl. Kapitel 6.4) ausgeführt. Dabei werden auch die beiden Konzepte *Zufällige Aufteilung* und *Mögliche Erklärungen* aufgenommen, wobei zu beachten ist, dass diese im Interviewgespräch mit der Interviewerin erarbeitet wurden und es hierzu keine Verschriftlichungen im Randomisierungstestschema gibt.

Konzept „Zufällige Aufteilung der Versuchspersonen auf die Gruppen“

Dieses Konzept können alle drei Paare im Gespräch erläutern. Näheres findet sich in der entsprechenden Interaktionseinheit der einzelnen Paare und in den Fallzusammenfassungen (11.2.3, 11.3.3 und 11.4.3). Bei Fabia und Laura ist auffällig, dass sie im Gegensatz zu den anderen beiden Paaren zunächst die Perspektive einer Person einnehmen, die an dem Experiment teilnimmt. Die anderen beiden Paare nehmen die „Forscherperspektive“ ein. Fabia und Laura haben Schwierigkeiten zu Beginn des Interviews, das Konzept der zufälligen Aufteilung zu erläutern, was möglicherweise auf die unterschiedliche Perspektive zurückgeführt werden kann.

Alle drei Paare können Merkmale nennen (Haarfarbe, Geschlecht, Sport, Alter), nach denen die Gruppen hätten eingeteilt werden können, wenn keine zufällige Einteilung vorgenommen worden wäre und können auch argumentieren, dass dadurch dann jeweils ein Schluss auf das entsprechende Merkmal und nicht auf die durchgeführte Behandlung hätte geschehen müssen. Dies passt zu den Forschungsergebnissen von Pfannkuch et al. (2015), in deren Studie Studierende ähnlich argumentierten. Im Laufe der Interviews adressieren alle drei Paare immer wieder dieses Konzept, wobei auffällig ist, dass Fabia und Laura dies nicht während der Formulierung der Nullhypothese adressieren, diese aber trotzdem korrekt aufstellen und verwenden.

Konzept „Mögliche Erklärungen für die beobachteten Unterschiede“

Von allen drei Paaren wird problemlos die Erklärung genannt, die der Forschungshypothese zugrunde liegt. Die der Nullhypothese zugrunde liegende Erklärung wird von Rebecca und Selina und von Mandy und Alisa ebenso problemlos genannt, Laura und Fabia können diese Erklärung nicht selbst formulieren. Rebecca und Selina greifen in ihren Ausführungen direkt auf die beiden Hypothesen vor. Die dritte Erklärung, dass es an konfundierenden Variablen liegt, wird weniger klar formuliert, dies kann aber auch daran liegen, dass ja bereits zuvor die zufällige Aufteilung der Versuchspersonen auf die Gruppen thematisiert wurde und dies eventuell von den Teilnehmern als Voraussetzung angesehen und nicht in Frage gestellt wird. Die dritte Erklärung wird von allen drei Paaren mit dem Verhalten der Versuchspersonen in Verbindung gebracht, dass diese sich anders verhalten, weil sie sich in einer Studie befänden. In allen drei Interviews greift die Interviewerin aktiv in die Erklärungen ein, um diese zusammenzufassen oder sogar weiter auszuführen, da die von den Teilnehmern im gemeinsamen Gespräch hervorgebrachten Erklärungen teilweise unvollständig sind oder unklar formuliert werden. Auch dies ist in Einklang mit den Ergebnissen von Pfannkuch et al. (2015), bei denen ca. 31% der Teilnehmer die Wirksamkeit der Behandlung nicht als mögliche Erklärung (in einem Nachtest) nannten.

Konzept „Forschungsfrage“

Rebecca und Selina formulieren die Forschungsfrage auf Anhieb, ohne weitere Diskussion. Fabia und Laura diskutieren über verschiedene Formulierungen, weshalb bei ihnen dies sechs Turns umfasst. Mandy und Alisa kommen einander ergänzend auf eine Formulierung. Die drei von den Paaren formulierten Forschungsfragen lauten:

Rebecca und Selina:	*Ist Fischöl wirksam, um eine Blutdrucksenkung herbeizuführen?*
Fabia und Laura:	*Sorgt die Fischöl-Diät für eine höhere Blutdrucksenkung als die andere Diät?*
Mandy und Alisa:	*Senkt Fischöl den Blutdruck effektiver als normales Öl?*

Alle drei gefundenen Forschungsfragen sind im Kontext formuliert und stellen geeignete Forschungsfragen dar. Rebecca und Selina haben ihre Formulierung am mühelosesten

gefunden, eine Schwierigkeit hat dies für keins der Paare bedeutet. Eher ist hierdurch allen Teilnehmern der Fokus der Untersuchung klar geworden.

Konzept „Beobachtete Daten – Randomisierung“

Bezüglich der erfolgten Randomisierung der Daten des Experiments diskutieren Rebecca und Selina am ausführlichsten die Aufteilung der Personen zu den Gruppen und stellen mögliche Szenarien auf, wie diese Aufteilung stattgefunden haben könnte. Laura und Fabia stellen schlicht fest, dass Randomisierung zufällige Einteilung bedeutet und diskutieren nicht, ob diese in Bezug auf das Experiment stattgefunden hat, notieren es jedoch so. Mandy und Alisa schließlich stellen den Mittelweg zwischen den beiden anderen Paaren dar, sie handeln kurz mit Bezug zum Kontext aus, dass eine Randomisierung im ursprünglichen Experiment stattfand und ergänzen jedoch fälschlicherweise, dass zusätzlich zur zufälligen Aufteilung eine zufällige Auswahl der Teilnehmer stattfand. Folgende Formulierungen wurden notiert, wobei nur die von den ersten beiden Paaren richtig ist.

Rebecca und Selina:	*Zufällige Zuordnung*
Fabia und Laura:	*Zufällige Einteilung von Freiwilligen*
Mandy und Alisa:	*Zufällige Auswahl und Zuteilung der Teilnehmer in die Gruppen*

Durch das Wort „Freiwilligen“ in der Verschriftlichung von Fabia und Laura lässt sich bei ihnen ein Bezug zum Kontext erkennen, der in der Kommunikation nicht hergestellt wurde, ebenso bei Mandy und Alisa. Rebecca und Selina operieren hier auf der statistischen Ebene ohne Bezug zum Kontext, allerdings hatten sie sich in der zugehörigen Gesprächsepisode in die Lage der Forscher versetzt und somit den stärksten Bezug zum Kontext hergestellt.

Konzept „Beobachtete Daten – Welcher Unterschied lässt sich feststellen?“

Die beiden Paare Rebecca und Selina und Mandy und Alisa konzentrieren sich darauf, bezüglich des Unterschieds der beiden Gruppen den Unterschied im arithmetischen Mittel der beiden Gruppen zu betrachten und daraus zu schließen, dass die Fischölgruppe in Bezug auf die Blutdrucksenkung besser abschneidet als die andere Gruppe. Fabia und Laura diskutieren an dieser Stelle zunächst, ob sie den Median oder das arithmetische Mittel als Vergleich nutzen wollen und benötigen auch zur Formulierung wesentlich länger. Folgende Formulierungen werden von den drei Paaren notiert:

Rebecca und Selina:	*Blutdrucksenkung bei Fischöl-Gruppe im ar. Mittel 7,7mm QS größer*
Fabia und Laura:	*Differenz der Mittelwerte beträgt 7,7* *→ Höhere Blutdruckreduktion bei Fischöl-Diät*
Mandy und Alisa:	*Die Fischöl-Gruppe hat im Durchschnitt eine höhere Blutdruckreduktion als die normale Öl*

Alle drei Formulierungen sind richtig und im Kontext verschriftlicht. Rebecca und Selina und Fabia und Laura quantifizieren den Unterschied, und alle drei Paare formulieren

den Unterschied zwischen den beiden Gruppen, jedoch nur Mandy und Alisa nennen die Vergleichsgruppe. Schwierigkeiten treten hier in der Form auf, dass sich Fabia und Laura zunächst unsicher sind, welchen Mittelwert sie nutzen sollen und die abgebrochenen Sätze zeigen, dass ihnen eine passende Formulierung nicht leicht fällt. Bei den anderen beiden Paaren werden die Formulierungen in einander ergänzenden Turns gemeinsam gefunden. Bei Rebecca und Selina lässt sich feststellen, dass diese hier bereits auf einen Kausalzusammenhang schließen, denn sie formulieren, dass Fischöl wirksamer ist, um den Blutdruck zu senken, anstatt sich auf den gefundenen Unterschied zu konzentrieren.

Konzept „Nullmodell – Forschungshypothese“

Die Formulierung der Forschungshypothese stellt bei allen drei Paaren die kürzeste Episode dar und nur Fabia und Laura benötigen hier mehrere Turns, in denen sie explizit den Bezug zur Forschungsfrage herstellen, um eine Formulierung zu finden. Folgende Formulierungen werden gewählt:

Rebecca und Selina:	*Fischöl ist hilfreich zur Blutdrucksenkung*
Fabia und Laura:	*Die Fischöl-Diät führt zu einer höheren Blutdruckreduktion als die Normale Diät*
Mandy und Alisa:	*Die Fischöl-Diät ist effektiver um den Blutdruck zu senken als die normale Öl-Diät*

Schwierigkeiten, die typisch sind bei der Formulierung der Null- und Forschungshypothese, insbesondere die Verwechslung von Null- und Alternativhypothese lassen sich bei diesen drei Paaren nicht entdecken. Alle drei Forschungshypothesen sind im Kontext formuliert, auffällig ist, dass die beiden Paare Fabia und Laura und Mandy und Alisa die Forschungshypothese vergleichend zwischen den beiden Ölbehandlungen formulieren, wohingegen Rebecca und Selina sich rein auf die Wirkung des Fischöls beziehen. Beide Formulierungen sind jedoch als richtig zu bewerten.

Konzept „Nullmodell – Nullhypothese“

Im Vergleich zum Formulieren der Forschungshypothese dauert das der Nullhypothese bei Fabia und Laura deutlich länger, bei den anderen beiden Paaren geht es ähnlich schnell. Wieder zeigt sich ein Unterschied in den Formulierungen der drei Paare:

Rebecca und Selina:	*Zufällige Gruppenzuordnung als Ursache für beobachtete Unterschiede*
Fabia und Laura:	*Es gibt keinen Unterschied auf die Blutdruckreduktion in der Auswirkung zwischen beiden Diäten*
Mandy und Alisa:	*Es spielt keine Rolle, welche Öl-Diät gemacht wird um den Blutdruck zu senken sondern die Ergebnisse liegen an der zufälligen Gruppeneinteilung*

Nur Rebecca und Selina notieren die Nullhypothese ohne weitere Formulierungsalternativen und beziehen die zufällige Gruppenzuordnung in die Formulierung mit ein. Mandy und Alisa haben die zufällige Zuordnung besprochen, nehmen diese jedoch erst nach einer Intervention durch die Interviewerin mit in die schriftliche Formulierung der Null-

hypothese auf. Generell benötigen sie viele Turns, in denen sie wechselseitig Sprachbausteine vorschlagen und ergänzen, um die Formulierung der Forschungshypothese zu erhalten. Dabei kommt es jedoch nicht zu Diskussionen in dem Sinne, dass sie Vorschläge als falsch abweisen, sondern sie ergänzen und präzisieren nur gegenseitig ihre Formulierungen. Fabia und Laura formulieren wie bereits zuvor echte Alternativen wechselseitig, ohne jedoch die von der jeweils anderen vorgeschlagene Formulierung zu kommentieren. Alle drei Paare formulieren die Nullhypothese auf Kontextebene, wobei Rebeccas und Selinas Formulierung auch in anderen Kontexten funktionieren würde und somit den geringsten Bezug zum Kontext erkennen lässt. Sowohl Rebecca und Selina als auch Mandy und Alisa nehmen in einzelnen Episoden während der Simulation Bezug zu dem in der Nullhypothese formulierten Sachverhalt und setzen ihn explizit als wahr voraus. Im Gegensatz zu den von Biehler et al. (2015) berichteten Schwierigkeiten zur Formulierung der Nullhypothese lassen sich hier keine Schwierigkeiten erkennen.

Konzept „Simulationsmodell“

Bezüglich dieses Konzepts unterscheiden sich die drei Paare deutlich. Allen drei Paaren ist direkt klar, dass sie die in der Nullhypothese beschriebene Situation modellieren müssen. Allerdings benötigen Rebecca und Selina viel Hilfe der Interviewerin, um die Simulation in TinkerPlots umzusetzen. Fabia und Laura beherrschen dies am Besten und benötigen weniger Hilfe durch die Interviewerin (vgl. Tabelle 46), deshalb benötigen sie am wenigsten Zeit in dieser Interaktionseinheit (vgl. Abbildung 199). Mandy und Alisa sind das einzige Paar, das die Randomisierung zunächst händisch vornimmt, anschließend benötigen sie trotzdem noch viel Hilfe durch die Interviewerin bei der Umsetzung der Simulation, dadurch benötigen sie am längsten. Die Beschreibung des Simulationsmodells nehmen alle drei Paare nach der Simulation im Randomisierungstestschema vor, hier zeigen sich wieder große Gemeinsamkeiten.

Rebecca und Selina:	*5000 mal zufällige Zuordnung zu den zwei neuen Gruppen A & B*
Fabia und Laura:	*Mit Hilfe der Zufallsmaschine von TinkerPlots*
	Zufällige Einteilung der Werte in zwei Gruppen
Mandy und Alisa:	*Zufällige neue Einteilung aller Werte in zwei Gruppen*

Über die jeweiligen Formulierungen diskutiert kein Paar lange, Rebecca und Selina formulieren die Anzahl der Wiederholungen und die Benennung der neuen Gruppen mit hinein, Fabia und Laura erwähnen explizit TinkerPlots als Hilfsmittel, vermutlich in Abgrenzung zur händischen Simulation. Beim Zusammenfassen für das Randomisierungstestschema ergaben sich keine Schwierigkeiten.

Konzept „Teststatistik“

Die Teststatistik wird von allen drei Paaren auf statistischer Ebene und auf Softwareebene richtig identifiziert, auch wenn sie teilweise Unterstützung auf Softwareebene be-

nötigen (vgl. Tabelle 48). Bezüglich der Formulierung gibt es bei Rebecca und Selina und bei Fabia und Laura keine Alternativvorschläge, Mandy und Alisa bekommen leichte Interventionen durch die Interviewerin, da diese beiden zunächst nur den „Durchschnittswert“ (363) als Teststatistik angeben, woraufhin sie recht lange untereinander diskutieren, welche sprachlichen Aspekte sie aufnehmen möchten.

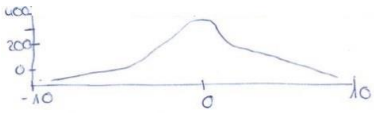
Rebecca und Selina:	*Differenz der ar. Mittel der neuen Gruppen A & B → dazu dann den Ursprungswert (7,7) in Bezug setzen*
Fabia und Laura:	*Differenz der Mittelwerte*
Mandy und Alisa:	*Die Differenz der Durchschnittswerte der beiden Gruppen*

Rebecca und Selina formulieren die meisten Details und geben sich selbst gleich eine Handlungsanweisung, wie mit der Teststatistik umzugehen ist. Die Unsicherheit von Mandy und Alisa beim Formulieren der Teststatistik zeigt sich deutlich darin, dass sie nicht wissen, welche sprachlichen Elemente (z. B. „die beiden neuen Gruppen“) sie aufnehmen sollen. Dies müssen sie erst miteinander aushandeln. Das Konzept scheint ihnen verständlich genug zu sein, um es in TinkerPlots umzusetzen, allerdings nicht genug, um es direkt formulieren zu können.

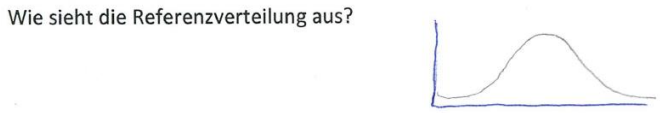
Konzept „Referenzverteilung“

Die Referenzverteilung in TinkerPlots zu erstellen, bereitet keinem der Paare Schwierigkeiten. Auch in dem letzten Teil des Interviews wird bei allen Paaren deutlich, dass sie sowohl die Verteilung als Ganzes, als auch einzelne Punkte der Verteilung korrekt erläutern können. Die Referenzverteilung in das Randomisierungstestschema zu übertragen ist der Schritt, der allen drei Paaren Schwierigkeiten bereitet und für den alle drei Paare leichte bis starke Intervention durch die Interviewerin benötigen. Alle bekommen den Hinweis, dass eine Skizze genüge, woraufhin sie sich in den Gruppen unterschiedlich lang darauf einigen, von wo bis wo die Achsen gezeichnet werden sollen. Folgende Referenzverteilungen werden skizziert:

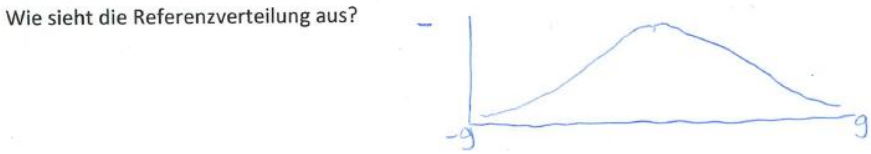
Rebecca und Selina:



Fabia und Laura:



Mandy und Alisa:



Fabia und Laura skizzieren ohne Achsenbeschriftung, Rebecca und Selina nehmen bei beiden Achsen eine Einteilung der Werte vor, Mandy und Alisa nehmen eine grobe Einteilung der X-Achse vor. Rebecca und Selina bemühen sich, die Verteilung der simu-

lierten Werte nachzuzeichnen, die anderen beiden Paare idealisieren stärker Richtung Normalverteilung, obwohl dies in der Veranstaltung nicht thematisiert wurde. Keins der Paare zeichnet den beobachteten Wert ein, um qualitativ dessen Lage zu dokumentieren, dies sollte für das Design der Lerneinheit in das Randomisierungstestschema aufgenommen werden (vgl. Kapitel 12.2.2.6).

Konzept „P-Wert“

Rebecca und Selina ermitteln den P-Wert in TinkerPlots mit Hilfe des Einteilers ohne Schwierigkeiten und notieren ihn kommentarlos. Bei den anderen beiden Paaren ergibt sich zunächst durch eine ungünstige Bedienung der Software ein markierter Anteil von ca. 17%, der jedoch von beiden Paaren korrekt korrigiert werden kann. Beim Aufschreiben ergibt sich für Mandy und Alisa die Schwierigkeit, dass sie zunächst richtig interpretieren, dass der P-Wert signifikant ist, sie jedoch nicht wissen, ob sie dies beim P-Wert oder bei den Schlussfolgerungen eintragen sollen. Schließlich einigen sie sich darauf, dass sie es beim P-Wert notieren, so wie auch die anderen beiden Paare.

Rebecca und Selina:	*1%*
	Signifikantes Ergebnis → P-Wert kleiner als 10%
Fabia und Laura:	*1%*
	Ja
Mandy und Alisa:	*P-Wert = 1%*
	Signifikant

Keins der Paare geht darauf ein, dass es sich beim erhaltenen P-Wert um einen geschätzten Wert handelt, obwohl dies im Baustein Inferenzstatistik thematisiert wurde. Allerdings hat auch keins der Paare Schwierigkeiten damit, den richtigen Bereich des Einteilers in TinkerPlots als P-Wert zu identifizieren, so wie dies z. B. bei Frischemeier (2017) beobachtet wurde und als mögliche Schwierigkeit erwartet wurde.

Konzept „Schlussfolgerungen – statistisch“

Rebecca und Selina unterhalten sich über die Schlussfolgerungen, die sie ziehen können und über Aussagen, die sie nicht treffen dürfen, wie beispielsweise, dass etwas bewiesen ist. Hier findet Kommunikation in ganzen Sätzen statt und ein richtiger Austausch über und somit eine Annäherung an gemeinsam hervorgebrachte Schlussfolgerungen. Gleichzeitig schweifen sie in der zugehörigen Interaktionseinheit ab, denn sie suchen den Dialog mit der Interviewerin darüber, wie positiv sie den Inhalt des Bausteins Inferenzstatistik empfunden haben und wie viel kritischer sie nun beim Lesen von Studienergebnissen sind. Fabia und Laura ziehen ihre Schlussfolgerungen in nur zwei Turns und diese werden von ihnen falsch formuliert, denn sie formulieren, dass die Nullhypothese nur zu einer Wahrscheinlichkeit von einem Prozent stimme (303). Anschließend verschriftlichen sie kommentarlos jedoch richtig. Mandy und Alisa schließlich machen ähnlich wie Fabia und Laura eine Aussage darüber, dass die Nullhypothese mit

99prozentiger Sicherheit abgelehnt werden könne (401), revidieren dies jedoch selbständig.

Rebecca und Selina:	*Forschungshypothese kann angenommen & Nullhypothese kann abgelehnt werden, allerdings unter Restunsicherheit*
Fabia und Laura:	*Wir lehnen die Nullhypothese ab und nehmen die Forschungshypothese an bei einem Wert von p=1%*
Mandy und Alisa:	*Die Nullhypothese kann abgelehnt werden*

In den gewählten Formulierungen zeigen sich Unterschiede hinsichtlich gewisser Aspekte. Rebecca und Selina betonen die Unsicherheit, unter der ihre Entscheidung fällt, hierüber haben sie auch zuvor schon kommuniziert. Fabia und Laura präzisieren das Niveau, auf welchem sie die Nullhypothese ablehnen und Mandy und Alisa schließlich formulieren nur sehr allgemein.

Konzept „Schlussfolgerungen – kausal“

Auf das Stichwort „kausal“ des Randomisierungstestschemas geht nur das Paar Mandy und Alisa ein, die anderen beiden Paare sprechen nicht über mögliche kausale Schlussfolgerungen. Mandy und Alisa handeln zunächst sprachlich aus, was genau unter dem Stichwort zu verstehen ist und nähern sich der endgültigen Formulierung mit mehreren Formulierungsvorschlägen.

Rebecca und Selina:	-
Fabia und Laura:	-
Mandy und Alisa:	*Blutdruckreduktion ist abhängig von der Art der Öl-Diät*

In ihren Formulierungsalternativen führen Mandy und Alisa auch „Blutdruckreduktion liegt am Fischöl“ (416) an, was ebenso gut wie die notierte Formulierung wäre. Dass die anderen beiden Paare dieses Stichwort nicht adressieren, lässt sich kaum erklären, allerdings wenden sich Fabia und Laura auch im nachfolgenden Interview diesem Konzept nicht wirklich zu, sondern kommunizieren eher auf statistischer oder auf Softwareebene. Möglicherweise ist ihnen am wenigsten klar, welche kausalen Schlussfolgerungen gezogen werden können und noch weniger, wie diese zu formulieren sind. Rebecca und Selina widmen sich im nachfolgenden Interviewteil dieser Frage ausführlich und zeigen dort, dass sie dieses Konzept verstanden haben und formulieren können.

Konzept „Schlussfolgerungen – Population“

Auch bezüglich der Population ziehen Fabia und Laura keine Schlussfolgerungen. Mandy und Alisa thematisieren dies am ausführlichsten, allerdings wieder nur in sehr kurzen Turns, in denen einzelne Sprachbausteine eingeworfen werden. Die Interviewerin steigt mit einer ausführlichen Intervention in die entsprechende Episode ein und hilft somit bei der Formulierung. Rebecca und Selina gehen auf diesen Punkt in der selbständigen Bearbeitung ein und kommunizieren ähnlich wie Mandy und Alisa in kurzen Turns.

Rebecca und Selina:	*Lässt sich nicht verallgemeinern (nur 14 Teilnehmer)*
Fabia und Laura:	-
Mandy und Alisa:	*Gilt für die Teilnehmer der Studie*

Dieses Konzept scheint allen drei Paaren Schwierigkeiten zu bereiten. Mandy und Alisa formulieren die Schlussfolgerung zwar korrekt, jedoch bekamen sie hierzu starke Intervention durch die Interviewerin, da sie es nicht selbst formulieren konnten. Bei Rebecca und Selina ist nicht klar, ob sie die Schlussfolgerung ziehen, dass für die 14 Personen des Experiments der Schluss gezogen werden darf, dass Fischöl wirksam ist um den Blutdruck zu senken. Dass dies auch für diese beiden ein unklares Konzept bleibt zeigt auch der nachfolgende Interviewteil. Fabia und Laura ziehen die entsprechende Schlussfolgerung gar nicht, auch im nachfolgenden Interview zeigen diese beiden hierzu kaum ergiebige Redebeiträge, so dass geschlossen wird, dass in Verbindung mit kausalen Schlussfolgerungen diese beiden Konzepte am schwierigsten für die Teilnehmer waren.

11.5.5 In welcher Weise gelingt es den Teilnehmern, die drei Ebenen Kontext, Statistik und Software aufeinander zu beziehen?

In 11.5.1 wurde bereits die Verknüpfung der drei Ebenen Kontext, Statistik und Software als Problemstelle identifiziert. In diesem Abschnitt ist in Tabelle 45 dargestellt, welche Ebenen die Teilnehmer sprachlich adressieren, wenn sie über die neun Konzepte sprechen und sich diese erarbeiten.

Tabelle 45: Übersicht über die Ebenen, auf denen die Paare während der Durchführung des Randomisierungstests kommunizieren

Konzept	Kontextebene	Statistikebene	Softwareebene
Zufällige Aufteilung	Rebecca & Selina	Rebecca & Selina	-
	Fabia & Laura	Fabia & Laura	-
	Mandy & Alisa	Mandy & Alisa	-
Mögliche Erklärungen	Rebecca & Selina	Rebecca & Selina	-
	Fabia & Laura	Fabia & Laura	-
	Mandy & Alisa	Mandy & Alisa	-
Forschungsfragen	Rebecca & Selina	-	-
	Fabia & Laura	-	-
	Mandy & Alisa	-	-
Beobachtete Daten	Rebecca & Selina	Rebecca & Selina	-
	Fabia & Laura	Fabia & Laura	-
	Mandy & Alisa	Mandy & Alisa	-
Nullmodell	Rebecca & Selina	Rebecca & Selina	-
	Fabia & Laura	Fabia & Laura	-
	Mandy & Alisa	Mandy & Alisa	-
Simulationsmodell	Rebecca & Selina	Rebecca & Selina	Rebecca & Selina
	Fabia & Laura	Fabia & Laura	Fabia & Laura
	Mandy & Alisa	Mandy & Alisa	Mandy & Alisa
Teststatistik/ Referenzverteilung	-	Rebecca & Selina	Rebecca & Selina
	-	Fabia & Laura	Fabia & Laura
	-	Mandy & Alisa	Mandy & Alisa

P-Wert	-	Rebecca & Selina	Rebecca & Selina
	-	Fabia & Laura	Fabia & Laura
	-	Mandy & Alisa	Mandy & Alisa
Schlussfolgerungen	-	Rebecca & Selina	-
	-	Fabia & Laura	-
	Mandy & Alisa	Mandy & Alisa	-

Es lässt sich ein deutliches Muster bei allen drei Paaren erkennen, von dem es nur eine Abweichung gibt. Die beiden Konzepte *Zufällige Aufteilung* und *Mögliche Erklärungen* werden sowohl auf der Kontextebene als auch auf der Statistikebene adressiert. Dies entspricht den Erwartungen nach der Analyse der Aufgabe, noch dazu, weil diese beiden Konzepte während der Interviewsituation gemeinsam mit der Interviewerin diskutiert wurden, von der durch Nachfragen entsprechende Erläuterungen auf beiden Ebenen eingefordert wurden. Dabei findet ein steter Wechsel in der Kommunikation zwischen den beiden Ebenen statt, auch innerhalb eines Turns. Ein Bezug zur Softwareebene findet nicht statt, ist jedoch auch nicht nötig, da es um allgemeine Erläuterungen für einen Randomisierungstest geht.

Über die Forschungsfragen wird von allen drei Paaren nur in einer kurzen Interaktionseinheit kommuniziert, und abgesehen von dem Wort „Forschungsfrage“ findet die Kommunikation hierzu rein auf Kontextebene statt, allerdings wird auch im Wesentlichen von allen nur die Formulierung einer Forschungsfrage erarbeitet.

Um über das Konzept beobachtete Daten zu kommunizieren, zu dem das Design des Experiments gehört und der Unterschied, der zwischen beiden Gruppen beobachtet wurde, nutzen wieder alle Paare die Kontext- und die Statistikebene, die miteinander verwoben werden.

Bezüglich des Nullmodells formulieren alle Paare die Nullhypothese auf Kontextebene mit Bezug zur Statistikebene. Die von Rebecca und Selina formulierte Nullhypothese ist durch einen geringen Kontextbezug beliebig anwendbar auf vergleichbare Situationen, dies ist eine mögliche Erklärung, weshalb sie am Ende des Bearbeitungsprozesses keine Schlussfolgerungen im Kontext ziehen. Allerdings ziehen auch Fabia und Laura solche Schlussfolgerungen nicht, obwohl sie die Nullhypothese mit deutlicherem Bezug zum Kontext formuliert haben (vgl. Abschnitt 11.5.2). Einzig Mandy und Alisa, die ebenfalls die Nullhypothese im Kontext formuliert haben, ziehen am Ende auch Schlussfolgerungen auf Kontextebene, allerdings lassen sie sich dabei durch das entsprechende Stichwort des Randomisierungstestschemas leiten, so dass nicht geschlossen werden kann, dass ihre Formulierung der Nullhypothese sie dazu geführt hat.

Während des Erstellens des Simulationsmodells in Interaktionseinheit 7 finden die Gespräche auf allen drei Ebenen statt, die erneut miteinander verbunden werden. Gerade für die Bedienung der Software ist zu beachten, dass der konkrete Bezug zum Kontext hilfreich ist.

Ein interessanter Aspekt ist, dass fast die gesamte Kommunikation in den Interaktionseinheiten 9 und 10, in denen mit TinkerPlots gearbeitet wird, ausschließlich auf Statistik- und Softwareebene stattfindet. Diese Ebenen sind hierbei stark verzahnt in der Kommunikation, allerdings komplett losgelöst von einem Bezug zum Kontext.

Schlussfolgerungen schließlich werden von allen Paaren auf Statistikebene formuliert, nur Mandy und Alisa beziehen hier die Kontextebene von sich aus mit ein. Insgesamt ist während der selbständigen Durchführung des Randomisierungstests die hierzu gehörige Interaktionseinheit vergleichsweise kurz. Sie erscheint in Abbildung 199 bei Rebecca und Selina lediglich sehr lang, weil hierin der Exkurs zu den Inhalten des Bausteins Inferenzstatistik fällt.

Während der Arbeit mit der Software TinkerPlots wäre möglicherweise ein Bezug zum Kontext hilfreich, um sich klar zu machen, was wie mit der Software abgebildet werden muss. Allerdings lässt sich an den vorgestellten Kommunikationen auch erkennen, dass eine Trennung der Ebenen, wie sie von Noll und Kirin (2017) auf Grundlage des Schemas von Biehler et al. (2015) interpretiert wird, nicht sinnvoll ist, sondern vielmehr eine bewusste Verknüpfung aller drei Ebenen angestrebt werden sollte.

11.5.6 Können die Teilnehmer die Software TinkerPlots als Instrument einsetzen, um einen Randomisierungstest durchzuführen? An welchen Stellen bei der Benutzung der Software TinkerPlots benötigen die Studierenden Hilfe?

Aus Tabelle 45 lässt sich bezüglich der Verwendung der Software TinkerPlots erkennen, dass über diese, wie nach der Aufgabenanalyse erwartet, in den drei Interaktionseinheiten 8, 9 und 10 kommuniziert und sie in diesen Einheiten auch eingesetzt wird. Es ist zu beobachten, dass es kein Zurückspringen in der Softwarebenutzung gibt, alle drei Paare arbeiten linear entlang des Randomisierungstestschemas. Das lässt sich dadurch erklären, dass auftretende Schwierigkeiten direkt, ggf. mit Hilfe der Interviewerin, gelöst werden. Deshalb ist kein Rückgriff auf ein bereits benutztes Objekt in TinkerPlots nötig. Schematisch lässt sich der Weg der Benutzung von TinkerPlots zur Durchführung eines Randomisierungstests aller drei Paare wie in Abbildung 210 darstellen.

Alle drei Paare beginnen mit dem Erstellen der Zufallsmaschine ((1) in Abbildung 210), lassen die Tabelle der simulierten Ergebnisse anzeigen ((2) in Abbildung 210) und gehen dazu über, die simulierten Ergebnisse in einer Graphik darzustellen ((3) in Abbildung 210), in der sie die Messgröße definieren, die in einer Tabelle gesammelt wird ((4) in Abbildung 210). Anschließend wird die Referenzverteilung erstellt ((5) in Abbildung 210). Dabei geschehen die Einstellungen in den jeweiligen Objekten durchaus in unterschiedlicher Reihenfolge, z. B. wird erst die Durchgangsanzahl in der Zufallsmaschine angepasst, bevor das zweite Bauteil gefüllt wird oder umgekehrt, aber der Weg von Objekt (1) zu Objekt (5) ist bei allen der gleiche.

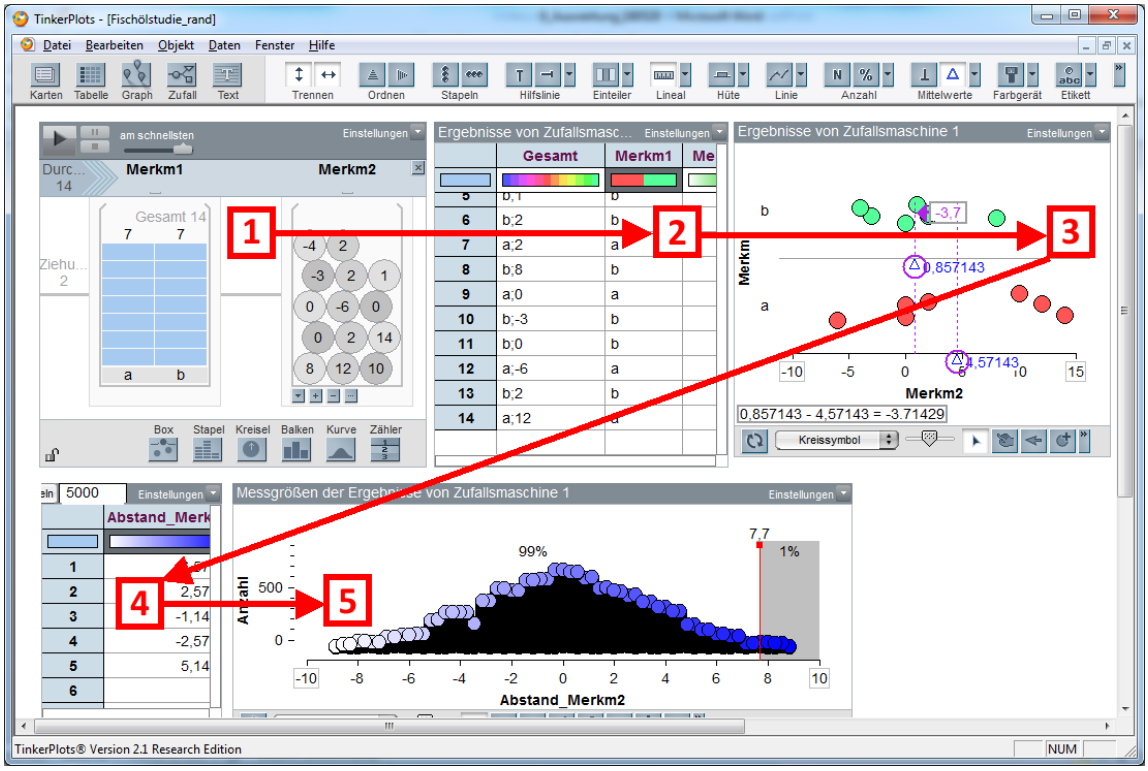


Abbildung 210: Benutzungsweg durch TinkerPlots beim Durchführen des Randomisierungstests

Wie bereits in den Analysen der Einzelpaare ausführlich dargelegt wurde, ergaben sich bei der Benutzung der Software auf Softwareebene einige Schwierigkeiten, an denen die Interviewerin interveniert hat. Ausgelöst wurde eine solche Intervention entweder durch direktes Fragen der Teilnehmer nach konkreter Hilfe, oder durch eigeninitiierte Turnübernahme der Interviewerin im Anschluss an längere Pausen, in denen die Teilnehmer nicht selbständig weitergearbeitet haben. Dabei wurden Hilfestellungen auf zwei Ebenen gegeben, entweder wurde eine leichte Intervention gegeben oder eine starke. Wie in Kapitel 10.4.4 ausgeführt, ist eine leichte Intervention gekennzeichnet durch einen allgemeinen Impuls der Interviewerin, durch den noch keine konkrete Handlungsanweisung für die Bedienung von TinkerPlots gegeben wird. Eine starke Intervention ist geprägt von genau solch direkten Handlungsanweisungen. Tabelle 46 bis Tabelle 50 geben eine Übersicht über alle Paare und deren TinkerPlots-Benutzung in Bezug auf die stärkste benötigte Intervention für eine in TinkerPlots auszuführende Aktion.

Tabelle 46: Notwendige Schritte in TinkerPlots zur Erstellung der Zufallsmaschine

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen der Zufallsmaschine	Fabia & Laura	Mandy & Alisa	Rebecca & Selina
Vorhandene Kugeln der Zufallsmaschine löschen	Fabia & Laura	-	Rebecca & Selina Mandy & Alisa

Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren	-	-	Rebecca & Selina Fabia & Laura Mandy & Alisa
Ein zweites Bauteil der Zufallsmaschine hinzufügen	Rebecca & Selina Fabia & Laura	-	Mandy & Alisa
Zweites Bauteil erstellen (für die neue Gruppenzuteilung)	-	-	Rebecca & Selina Fabia & Laura Mandy & Alisa
Die Einstellung ohne Zurücklegen bei beiden Bauteilen auswählen	-	Fabia & Laura	Rebecca & Selina Mandy & Alisa
Durchgangsanzahl festlegen	Fabia & Laura	Rebecca & Selina	Mandy & Alisa

Für das Erstellen der Zufallsmaschine benötigen alle Paare starke Interventionen, wie in Tabelle 46 zu sehen. Besonders auffällig ist, dass Mandy und Alisa für jeden Schritt Interventionen benötigen und bis auf eine Ausnahme immer direkte Anweisungen durch die Interviewerin. Diesem Paar hat für die Softwarebedienung offenbar die zuvor durchgeführte händische Simulation nicht geholfen. Auch Rebecca und Selina benötigen für fast alle Schritte direkte Anweisungen. Vor allem für den Schritt *Alle Werte des Merkmals Blutdruckreduktion in ein Bauteil kopieren* benötigen alle drei Paare konkrete Anweisungen, dies mit den Tastaturkürzeln STRG+C und STRG+V durchzuführen. Ebenso stellt der Schritt *Zweites Bauteil erstellen (für die neue Gruppenzuteilung)*, in welchem die neuen Gruppen abgebildet werden, für alle drei Paare eine Schwierigkeit dar. Die Interaktionsanalyse hat gezeigt, dass die Hilfen hierzu sowohl auf Softwareebene, als auch teilweise auf statistischer Ebene benötigt werden. Es lässt sich sagen, dass ohne Hilfe keins der Paare die Zufallsmaschine hätte erstellen können, jedoch wurden die einzelnen Schritte währenddessen oder im letzten Interviewteil von den Paaren gut erklärt und mit der statistischen Ebene verknüpft, wie die Interaktionsanalyse gezeigt hat. Bezüglich des Schritts *Durchgangsanzahl festlegen* lässt sich vermuten, dass dies einen semantischen Konflikt ausgelöst hat bei Rebecca und Selina und bei Mandy und Alisa, die versucht haben, über die Durchgangsanzahl die Wiederholungsanzahl der Simulationen (z. B. n = 1000) zu realisieren, anstatt die Stichprobengröße des Experiments zu reproduzieren. Dies mag auf Softwareebene daran liegen, dass für einen Randomisierungstest die Simulationsmethode der sequentiellen Simulation (vgl. Kapitel 3.5.2.2) benötigt wird, die in den Veranstaltungen nur sehr knapp behandelt wurde und somit nicht klar zu sein scheint, oder daran, dass auf Statistikebene die Modellierung der in der Nullhypothese angenommenen Situation nicht klar genug ist, um dies entsprechend in der Software umzusetzen.

Tabelle 47: Optionale Schritte in TinkerPlots zur Erstellung der Zufallsmaschine

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einstellen der Geschwindigkeit	Rebecca & Selina Fabia & Laura	-	Mandy & Alisa
Umbenennen der Merkmale Merkm1 und Merkm2	Rebecca & Selina Fabia & Laura	Mandy & Alisa	-

Zwei der drei Paare konnten die optionalen Schritte zur Erstellung der Zufallsmaschine ohne Hilfe durchführen, Mandy und Alisa benötigen auch hierzu leichte bzw. starke Intervention. Im Unterschied zu den Ergebnissen von Noll und Kirin (2017) hat sich das Nicht-Umbenennen der Merkmale hier nicht als Schwierigkeit herausgestellt, allerdings hat es zumindest für das Paar Rebecca und Selina einen hervorragenden Diskussionsanlass geboten, um das Konzept der zufälligen Aufteilung zu durchdringen.

Tabelle 48: Notwendige Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Erstellen einer Graphik zur Simulation	Fabia & Laura	Mandy & Alisa	Rebecca & Selina
Das erste Merkmal auswählen und darstellen	Fabia & Laura Mandy & Alisa	-	Rebecca & Selina
Das zweite Merkmal auswählen und in der Graphik passend darstellen	Rebecca & Selina Mandy & Alisa	Fabia & Laura	-
Die Mittelwerte einblenden	Rebecca & Selina Fabia & Laura Mandy & Alisa	-	-
Differenz der Mittelwerte mit Hilfe der Linealfunktion messen	Fabia & Laura	-	Rebecca & Selina Mandy & Alisa
Messgröße definieren	-	-	Rebecca & Selina Fabia & Laura Mandy & Alisa
Anzahl der zu sammelnden Messgrößen eintragen	Fabia & Laura	Rebecca & Selina Mandy & Alisa	-
Messgrößen sammeln	Rebecca & Selina Fabia & Laura	Mandy & Alisa	-
Erstellen einer Graphik zu den gesammelten Messgrößen	Rebecca & Selina Fabia & Laura	Mandy & Alisa	-
Referenzverteilung darstellen	Rebecca & Selina Fabia & Laura Mandy & Alisa	-	-

Die in Tabelle 48 gezeigten Schritte umfassen das Erstellen einer geeigneten Graphik zur Darstellung der simulierten Ergebnisse zur Definition der Teststatistik als Messgröße bis hin zur Erzeugung der Referenzverteilung. Hier zeigt sich, dass die hierzu benötigten Schritte von den Teilnehmern häufig ohne Interventionen durch die Interviewerin durchgeführt werden können. Starthilfe für diesen Abschnitt benötigten Rebecca und Selina, bei Mandy und Alisa genügte ein kurzer Hinweis und Fabia und Laura konnten ohne Hilfe beginnen. Bei der Benutzung der Linealfunktion ergaben sich für zwei der drei Paare größere Schwierigkeiten, dies war jedoch zu erwarten, da mit diesem Werkzeug nur in einer Seminarsitzung gearbeitet wurde. Auch das Festlegen der Messgröße barg für alle drei Paare Schwierigkeiten. Alle auftretenden Schwierigkeiten dieser Episoden waren jedoch auf Softwareebene angesiedelt, die Paare konnten immer zufrieden-

stellend erklären, welchen Schritt sie durchführen wollen (zum Beispiel dass sie den Abstand zwischen den Mittelwerten der beiden neuen Gruppen messen und sammeln wollen) und benötigten rein auf Softwareebene Hilfen.

Tabelle 49: Optionale Schritte in TinkerPlots zur Definition der Teststatistik und zur Erzeugung der Referenzverteilung

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Fälle in der Graphik stapeln	Fabia & Laura Mandy & Alisa	Rebecca & Selina	-
Beschleunigen des Messgrößensammelns	-	Fabia & Laura	Rebecca & Selina Mandy & Alisa

Die Fälle in der Graphik zu stapeln ist zur Ermittlung des P-Werts ein nicht notwendiger Zwischenschritt, jedoch sinnvoll. Zwei der drei Paare taten dies selbständig, Rebecca und Selina bekamen ungefragt eine leichte Intervention hierzu. Bezüglich des Beschleunigens des Messgrößensammelvorgangs wurden wie nach der Aufgabenanalyse erwartet starke Interventionen benötigt. Erstaunlich war, dass Fabia und Laura diesen Schritt selbständig begannen und lediglich leichte Interventionen benötigten.

Tabelle 50: Notwendige Schritte in TinkerPlots zur Ermittlung des P-Werts

Schritt	Keine Intervention	Leichte Intervention	Starke Intervention
Einteiler verwenden	Rebecca & Selina Mandy & Alisa	Fabia & Laura	-
Prozente einblenden	Rebecca & Selina Fabia & Laura Mandy & Alisa	-	-

Das Ermitteln des P-Werts mit Hilfe des Einteilers gelang allen drei Paaren mühelos, lediglich Fabia und Laura fragten nach dem zu benutzenden Werkzeug und bekamen hierzu eine leichte Intervention.

Im abschließenden Kapitel 12 werden die Ergebnisse der hier vorliegenden Analysen zusammengefasst und die Forschungsfragen beantwortet.

11.6 Einschränkung und Ausblick

Wie bei Fallstudien üblich, ist auch die vorliegende mit einigen Einschränkungen verbunden. Die explorative Fallstudie wurde mit einer geringen Anzahl an Teilnehmern durchgeführt, die noch dazu einen speziellen fachlichen Hintergrund haben. Es haben sechs Studierende (drei Paare) an der Fallstudie teilgenommen, die sich freiwillig gemeldet haben und als nicht repräsentativ für die Kohorte der Studierenden anzusehen sind, auch nicht für diejenigen Studierenden, die im Sommersemester 2017 die Veranstaltung „Modellieren, Größen, Daten und Zufall I“ und die Veranstaltung „Computergestützte Lernumgebungen zu Modellieren, Größen, Daten und Zufall“ besucht haben.

Diese beiden Veranstaltungen und insbesondere der darin verortete Baustein „Inferenzstatistik mit Randomisierungstests“ stellt darüber hinaus das spezielle Vorwissen dar, das die Teilnehmer der explorativen Fallstudie mitgebracht haben.

Aus diesen Einzelfalluntersuchungen lassen sich durch die detailreichen Analysen jedoch wertvolle Einblicke in die Bearbeitungsprozesse von Studierenden bekommen. Nur eine qualitativ angelegte Studie mit wenigen Teilnehmern ermöglicht einen solch tiefen Einblick. Auch wenn damit kein Anspruch auf unmittelbare Verallgemeinerbarkeit erhoben werden kann, so kann doch ein gewisses Generalisierungspotential angenommen werden. Dies gilt umso stärker, als die vorliegenden Ergebnisse mit denen anderer Studien verglichen werden und daraus Theorien abgeleitet werden können (siehe Kapitel 12).

Durch die vorliegenden Auswertungen liegen qualitative Einschätzungen vor, beispielsweise wie die Teilnehmer die verschiedenen Ebenen in ihre Argumentationen einbeziehen, wie die einzelnen Konzepte adressiert werden oder an welchen Stellen Hilfe bei der TinkerPlotsbenutzung benötigt wird. Eine anschließende Studie, in der quantitative Auswertungen gemacht werden, könnte hier als Folgeprojekt weitere Einblicke liefern.

Somit ergibt sich eine Perspektive für an die Arbeit anschließende Forschung (Kapitel 12.4).

12 Fazit

Simulationen werden als ein Weg angesehen, um einen einfachen Zugang zum statistischen Schließen mit Randomisierungstests zu erhalten (G. Cobb 2007, Edgington 1995, Lock et al. 2014). Mit der vorliegenden Arbeit wurde das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation analysiert und aufgezeigt, wie diese Software in einer universitären Veranstaltung für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung eingesetzt werden kann, um eine Lerneinheit zum statistischen Schließen über Randomisierungstests zu gestalten. Darauf aufbauend wurde eine explorative Studie durchgeführt, wie der Bearbeitungsprozess von Studierenden aussieht, die einen Randomisierungstests mit der Software TinkerPlots durchführen. In diesem Kapitel werden nun die zentralen Ergebnisse der Arbeit zusammengefasst, für die Stochastikdidaktik eingeordnet und ein Ausblick auf weitere Forschung gegeben.

12.1 Ausgangslage

Das Treffen von datenbasierten Entscheidungen unter Unsicherheit gewinnt zunehmend an Bedeutung (AK Stochastik 2003, 21) und stellt eine wichtige Fähigkeit dar, die jeder Erwachsene haben sollte (Garfield & Ben-Zvi 2008a, Ridgway 2016, Schiller & Engel 2018). Dabei muss zunächst die Logik des statistischen Schließens verstanden werden (Pfannkuch et al. 2011, Rossman 2008,), woran sich Fragen nach Erklärungen für beobachtete Unterschiede in Daten häufig natürlich anschließen (Madden 2011, Wild & Pfannkuch 1999). Randomisierungstests werden von vielen internationalen Stochastikdidaktikern als eine Methode angesehen, um in diese beiden Felder einzuführen. Eine elementare Grundlage, um die Einfachheit von Randomisierungstests zu erhalten, sind stochastische Simulationen, für die die Wahl einer geeigneten Software, mit der diese durchgeführt werden können, entscheidend ist (Budgett et al. 2012, Burrill & Biehler 2011, Holcomb, Chance, Rossman & Cobb 2010).

Die Möglichkeit, mit einer bestimmten Software ein bestimmtes Problem zu simulieren, ist von der jeweiligen Simulationskapazität der Software abhängig. Der Prozess des Simulierens, in welchem eine Software als sinnvolles Instrument eingesetzt wird, wird durch die instrumentelle Genese (Rabardel 2002) des Nutzers bestimmt. Diese wiederum ist davon abhängig, inwiefern eine Software bestimmte Nutzungsweisen ermöglicht und inwiefern ein Nutzer die jeweiligen Bedingungen und Beschränkungen einer Software kennt. Eine Werkzeuganalyse, die das Unterstützungspotential zur stochastischen Simulation herausarbeitet, ist für den Prozess der instrumentellen Genese notwendig, damit aus einem Werkzeug ein Instrument wird, das zielgerichtet eingesetzt werden kann.

Um Studierende, die bis dato wenig oder keine Erfahrung mit statistischem Schließen und stochastischen Simulationen haben, in dieses Thema einzuführen, wird eine passende Lernumgebung benötigt. Oftmals findet sich in der fachlichen Ausbildung von Lehramtsstudierenden nur eine einzige Veranstaltung im Bereich der Stochastik. Vor allem in der fachlichen Ausbildung von Studierenden des Grundschullehramts ist das Thema Inferenzstatistik nicht vorhanden, obwohl diese über grundschulnahe Inhalte hinaus (Sill 2018), wie alle anderen Studierenden auch, zu einer gewissen Mündigkeit, und somit Teilhabe an der Demokratie (Engel 2014b) ausgebildet werden sollen. Hier bietet sich die Möglichkeit, eine kurze Lerneinheit in eine bestehende Lehrveranstaltung zu implementieren, um dieses wichtige Thema der Stochastik in die fachliche Ausbildung aufzunehmen.

Die Software TinkerPlots ist eine in vielerlei Hinsicht gut geeignete Software, um sie in einer Lerneinheit zur Inferenzstatistik für Grundschullehramtsstudierende einzusetzen. Das Erlernen von Simulationen unter den drei von Biehler und Maxara (2007) genannten Perspektiven, nämlich um Zufallsexperimente zu repräsentieren, im Wechselspiel mit analytischen Methoden und als Methode eigener Art, wird von TinkerPlots unterstützt und die Software kann sowohl zum Erlernen stochastischer Inhalte genutzt werden als auch als Demonstrationsmedium im Unterricht.

Mit Blick auf den Stand der Forschung finden sich einige Arbeiten, die den Nutzen der Software zum Erlernen stochastischer Inhalte herausstellen (vgl. Kapitel 3.3 und 6.5), allerdings fehlt eine systematische Analyse des Unterstützungspotentials der Software TinkerPlots zur stochastischen Simulation. Einführungen in die Inferenzstatistik über Randomisierungstests finden sich ebenfalls einige wenige in der neueren didaktischen Literatur. Der Bearbeitungsprozess von Lernenden beim Durchführen von Randomisierungstests wird unter einzelnen Fragestellungen in wenigen Studien untersucht, allerdings ist der Gesamtblick auf diesen Prozess in Deutschland bisher unerforscht und stellt eine Forschungslücke dar.

Ziele der Dissertation

Daraus ergaben sich drei Ziele für diese Dissertation, die miteinander verknüpft sind. Zum ersten sollte das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation systematisch in einer Werkzeuganalyse untersucht werden.

Als zweites sollte aufbauend auf der Werkzeuganalyse und dem aktuellen Forschungsstand eine Lerneinheit zur Inferenzstatistik entwickelt werden.

Drittens sollten in einer empirischen Studie die kognitiven Prozesse Studierender analysiert werden, die einen Randomisierungstest unter Verwendung der Software TinkerPlots durchführen, nachdem sie Gelegenheit hatten, sich in einer für sie gestalteten Lerneinheit damit intensiv zu beschäftigen, um die Wirksamkeit der Lerneinheit nach-

zuweisen und generelle Erkenntnisse über Bearbeitungsprozesse in Kleingruppen zu erlangen. Konkret sollte in einer Fallstudie untersucht werden, welche Schwierigkeiten (erwartete und unerwartete) auftreten, welche Faktoren für gelingende kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots identifiziert werden können und welche Erkenntnisse für ein mögliches Redesign der Lerneinheit sich aus den Bearbeitungsprozessen gewinnen lassen. Um diese Fragen zu beantworten, wurde ein qualitatives Design benötigt, das detaillierte Einblicke in die Bearbeitungsprozesse erlaubt.

12.2 Zusammenfassung der Ergebnisse

Der Einsatz von Simulationssoftware als Werkzeug unter besonderer Perspektive der instrumentellen Genese nach Rabardel (2002) und der instrumentellen Orchestrierung nach Trouche (2004) wurde in Kapitel 2 behandelt. Die Werkzeug- und Detailanalyse der Software TinkerPlots zur stochastischen Simulation (Kapitel 3 und 4) bestimmte den ersten Teil dieser Arbeit.

Die neu designte Lerneinheit „Inferenzstatistik mit Randomisierungstests“ und die empirische Fallstudie zum Durchführen von Randomisierungstests unter Verwendung der Software TinkerPlots bestimmten den zweiten Teil der Arbeit. Für diesen zweiten Teil wurden in einem Überblick relevante Forschungsliteratur über Lernende beim Durchführen von Randomisierungstests sowie curriculare Beschreibungen zu Unterrichtseinheiten zur Einführung in Inferenzstatistik über Randomisierungstests aufgearbeitet (Kapitel 5 und 6). Daraus wurden neun Konzepte synthetisiert, die bei der Durchführung eines Randomisierungstests adressiert werden sollten, sowie Designideen für Lernumgebungen zu Randomisierungstests herausgearbeitet.

12.2.1 Zusammenfassung der Ergebnisse der Werkzeuganalyse

Bezüglich der Werkzeuganalyse der Software TinkerPlots lassen sich verschiedene Ergebnisse festhalten.

Ein zentrales Ergebnis ist, dass sich vier Simulationsarten in TinkerPlots unterscheiden lassen. Dies sind die simultane Simulation, die sich in simultan-gebündelt und simultan-hintereinandergeschaltet untergliedern lässt, die sequentielle Simulation, die Simulation als Stichprobenziehung und die Simulation durch Randomisierung. Für alle vier Simulationsarten kann man sich nach dem Simulationsplan aus 2.4 richten, da die Simulationen eine ähnliche Schrittfolge aufweisen. Somit kann ein allgemeines *instrumented action scheme* (Rabardel 2002) per instrumenteller Genese bei Lernenden aufgebaut werden, das durch die Benutzung eines Simulationsplans unterstützt wird.

Das Objekt der Zufallsmaschine als Herzstück des Modellierungsinterfaces bietet viele Möglichkeiten, eine diskrete Ergebnismenge darzustellen. Typische Zufallsgeräte wie

Münze, Würfel oder Urne lassen sich auf verschiedene Weisen realisieren, wobei gerade auch ein Darstellungswechsel von einem zu einem anderen Bauteil lernförderlich sein kann (vgl. Kapitel 3.5). Das Urnenmodell nimmt in der Stochastik eine zentrale Stellung ein (Sedlmeier 1999), und durch das Bauteil „Box“ der Zufallsmaschine können auf dem Urnenmodell basierende mentale Modelle direkt in der Software abgebildet werden, so dass die Visualisierung eines mentalen Modells direkt in dem technischen Werkzeug TinkerPlots geschehen kann, ohne dass eine Übersetzung in eine Programmiersprache oder in Formeln geschehen muss.

Je nach gegebener Problemsituation haben die einzelnen Bauteile der Zufallsmaschine Vorteile, durch die sich Zufallsexperimente verschiedenster Art abbilden lassen und die verschiedene kognitive Vorstellungen des Nutzers bedienen. Sehr benutzerfreundlich ist die formelfreie Konstruktion von Modellierungen, die in besonderer Weise TinkerPlots für die stochastische Simulation als kognitives Werkzeug (Specker 1997) auszeichnet. Die Visualisierung des zufallserzeugenden Prozesses, wie beispielsweise das Drehen des Bauteils „Kreisel“ oder das Durchmischen der Kugeln im Bauteil „Box“, ist eine weitere Stärke des Programms und kann das Lernen nachhaltig unterstützen.

Mit Vorsicht zu behandeln ist die Einstellung „mit Zurücklegen“, die standardmäßig für alle Bauteile aktiv ist. Soll ein Zufallsexperiment „ohne Zurücklegen“ erstellt werden, so muss dies aktiv durch den Nutzer ausgewählt werden, wenn dies nicht geschieht, so wird ggf. keine zu einer gegebenen Situation passende Modellierung erstellt. Problematisch hieran ist, dass dies möglicherweise nicht auffällt, weil trotzdem eine auf Programmierungsebene lauffähige Simulation erstellt wurde.

Das Erstellen mehrerer Bauteilspalten und das Realisieren von Abzweigungen zwischen den Bauteilen lassen Möglichkeiten zu, sowohl unabhängige als auch abhängige mehrstufige Zufallsexperimente ebenso einfach zu konstruieren, wie einstufige Zufallsexperimente oder mehrstufige Zufallsexperimente, die aus mehreren gleichen Teilexperimenten bestehen. Allerdings zeigt sich hier auch ein Nachteil der Software, denn um Zufallsexperimente mit mehreren Abzweigungen zu erstellen, wird ein großer Monitor benötigt, da sich die Elemente der Bauteile nicht dynamisch anpassen und deshalb bei einer zu kleinen Darstellung nicht bedient werden können. Somit sind verschiedene Computer oder Laptops nur bedingt geeignet, um mehrstufige Zufallsexperimente mit TinkerPlots zu modellieren.

Ereignisse und Zufallsgrößen werden in TinkerPlots in der Ereignistabelle als Merkmale oder in Graphen über Messgrößen definiert. Hilfreich sind hierbei vordefinierte Auswertungsmerkmale wie beispielsweise *Summe* oder *Anzahl*, für die keine Formel benutzt werden muss. Hier ist das Unterstützungspotential der Software besonders hoch. Eine besondere Funktion nimmt das Merkmal „Gesamt“ ein, das die simulierten Werte aller Merkmale eines Durchgangs in einer Zelle als Liste protokolliert. Hierdurch sind

Definitionen von verschiedenen Ereignissen und Zufallsgrößen stark vereinfacht und wiederum ist das Unterstützungspotential als hoch einzustufen.

Messgrößen sind ein wichtiges und intuitiv zu bedienendes Konzept in TinkerPlots. Sie werden zwingend benötigt, um sequentielle Simulationen oder Simulationen durch Randomisierungen durchzuführen, kommen dabei allerdings komplett ohne Formeln aus. Durch die leichte Bedienbarkeit lassen sich hiermit intuitiv erste Erfahrungen für Schwankungen eines gegebenen Zufallsexperiments machen.

Die Auswertung der simulierten Daten geschieht in TinkerPlots ausschließlich über Graphen. Hierzu ist das Potential für alle Simulationsarten gleich, eine Analyse desselben stand für diese Arbeit jedoch nicht im Vordergrund.

Das Unterstützungspotential der Software bezüglich Beschriftungen ist ebenfalls hoch, denn in TinkerPlots werden sämtliche nötigen Beschriftungen automatisch durch das Programm vorgegeben. Zusätzlich können sie manuell angepasst werden.

TinkerPlots bietet insgesamt ein hohes Unterstützungspotential für die stochastische Simulation und stellt mit dem Objekt Zufallsmaschine ein Modellierungsinterface vor, das vielfältige Zufallsexperimente abbilden kann und vor allem mentale Modelle des Nutzers visuell und intuitiv unterstützt.

Der Vorteil des hohen Unterstützungspotentials mit vielen Vereinfachungen hat jedoch generell den Nachteil, dass eine inhaltliche Auseinandersetzung mit den Daten, bzw. Variablen nicht in aller Tiefe erfolgen muss. Besonders deutlich wird dies bei der automatischen Beschriftung. Auf der einen Seite können durch die Unterstützung durch eine korrekte Benennung keine Fehler produziert werden, die aus einer falschen Beschriftung resultieren und der Prozess des Simulierens wird durch die automatische Namensvergabe beschleunigt. Auf der anderen Seite besteht dadurch keine Notwendigkeit für den Nutzer, über die inhaltliche Bedeutung der einzelnen Merkmale nachzudenken und hierfür „sprechende“ Namen zu vergeben.

12.2.2 Zusammenfassung der Ergebnisse des zweiten Teils

Der Rahmen des zweiten Teils der Dissertation wurde von vier Forschungsfragen gegeben.

1. Wie kann eine konkrete Lerneinheit zur Inferenzstatistik mit Randomisierungstests für Studierende des Lehramts an Grundschulen mit dem Lernbereich Mathematische Grundbildung konzipiert und gestaltet werden, die die instrumentelle Genese der Lernenden zur Benutzung der Software TinkerPlots für Randomisierungstests anleitet und die die in Kapitel 6 genannten Konzepte und Ideen umsetzt?

In Bezug auf die durchgeführte Fallstudie wurden die folgenden Fragen gestellt.

2. Inwieweit treten die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen auf? Welche unterwarteten Problemstellen können identifiziert werden?
3. Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests können mit der Software TinkerPlots festgestellt werden?
4. Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?

Die Teilnehmer der Fallstudie haben an den beiden Lehrveranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen für Modellieren, Größen, Daten und Zufall I“ im Sommersemester 2017 an der Universität Paderborn teilgenommen, in denen der insgesamt sechs Semesterwochenstunden umfassende Baustein Inferenzstatistik, wie in Kapitel 8.2.2 beschrieben, am Ende des Semesters durchgeführt wurde. Diese Lerneinheit bildete die Grundlage der Teilnehmer. Hier wurden die Teilnehmer in die verschiedenen, zu einem Randomisierungstest gehörigen Konzepte (Kapitel 6.4) eingeführt und lernten, diese anzuwenden, sowohl per händischer Simulation als auch über computergestützte Simulationen mit der Software TinkerPlots.

Unter Laborbedingungen wurde die Studie zur kollaborativen Bearbeitung eines Randomisierungstests durchgeführt, insgesamt haben sechs Studierende jeweils zu zweit an einem halbstrukturierten Leitfadeninterview teilgenommen (Kapitel 9.4) und die Aufgabe (Kapitel 9.2.1) unter Verwendung des Randomisierungstestschemas bearbeitet. Diese Interviews wurden videographiert, zusätzlich wurden die Bildschirmaktivitäten aufgezeichnet und die schriftlichen Bearbeitungen der Teilnehmer eingesammelt. In einem aufwändigen Transkript wurden die Gespräche der Beteiligten, die durch Screenshots festgehaltenen Aktionen am Computer, d. h. insbesondere in der Software TinkerPlots, die Verschriftlichungen im Randomisierungstestschema und die händische Simulation (sofern durchgeführt) in zeitlich passender Reihenfolge zusammengeführt.

Um die Forschungsfragen 2, 3 und 4 zu beantworten, wurden die Transkripte mit interpretativen Methoden (Krummheuer & Naujok 1999) analysiert und durch die funktionale Argumentationsanalyse (Toulmin 1975) ergänzt. Um die Analyse zu strukturieren, wurden Interaktionseinheiten entlang den Abschnitten des Leitfadeninterviews (Kapitel 9.4) und den in Kapitel 6.4 identifizierten Konzepten definiert, dabei wurde das Gesprächsphasenmodell von Deppermann (2008) zur Abgrenzung der Interaktionseinheiten verwendet. Für die Analyse hat sich diese Strukturierung als nützlich und tragbar

herausgestellt. Die Interpretationen der Einheiten wurden mit einem unabhängigen Forscher diskutiert, bis Einigkeit gewonnen wurde.

Entlang der vier Forschungsfragen werden nun die Ergebnisse zusammengefasst.

12.2.2.1 Gestaltung einer Lerneinheit „Inferenzstatistik mit Randomisierungstests“

Auf Grundlage der eigenen Vorarbeiten, der durchgeführten Werkzeuganalyse zur stochastischen Simulation der Software TinkerPlots und dem Forschungsüberblick wurde theoriegeleitet eine Lerneinheit „Inferenzstatistik mit Randomisierungstests“ zur Implementation im Sommersemester 2017 entwickelt. Für diese Lerneinheit sind auf struktureller Ebene und für die konkrete Durchführung Maßnahmen entwickelt worden.

Auf struktureller Ebene wurde die „Kernlogik der Inferenzstatistik“ (G. Cobb 2007) anhand des Dreischritts

1. Nachdenken über die Daten
2. Randomisieren und Wiederholen
3. Schlussfolgerungen ziehen

eingeführt und die neun Konzepte aus Kapitel 6.4 als wesentliche Bestandteile eines Randomisierungstests präsentiert und erläutert. Außerdem wurden nach einem immer gleichen Ablauf (siehe Ergebnisse der Werkzeuganalyse) die tinkerplotsspezifischen *instrumented action schemes* (Rabardel 2002) angeleitet, damit TinkerPlots als Instrument durch Lernende genutzt werden konnte.

Für die konkrete Durchführung wurde sich erstens für ein Eingangsbeispiel mit einem signifikanten P-Wert entschieden (Holcomb, Chance, Rossman, Tietjen, et al. 2010), zweitens entschieden, eine händische Simulation (u. a. Budgett et al. 2012, Rossman et al. 2008, Tintle et al. 2014) in die Lerneinheit aufzunehmen, und drittens das Schema zur Durchführung eines Randomisierungstests als Strukturierungshilfe entwickelt.

Zusammenfassend lässt sich sagen, dass die entwickelte Lerneinheit theoriegeleitet aus dem aktuellen Forschungsstand entwickelt wurde und die Teilnehmer in die Logik der Inferenzstatistik mit Randomisierungstests eingeführt hat. Genauere Erkenntnisse dazu lassen sich aus den folgenden drei Forschungsfragen aus der Analyse der Fallstudie gewinnen, aus denen auch Hinweise für ein mögliches Redesign der Lerneinheit gezogen werden (Kapitel 12.2.2.6).

12.2.2.2 Problemstellen bei der selbständigen Durchführung eines Randomisierungstests

In diesem Abschnitt wird die Forschungsfrage 2 „Inwieweit sind die in der a priori Analyse der Aufgabe zu erwartenden Problemstellen aufgetreten? Welche unerwarteten Problemstellen konnten identifiziert werden?“ beantwortet.

In der a priori Analyse zur Aufgabe der Fallstudie (Kapitel 9.2.2) wurde ein hypothetischer Lösungsweg beschrieben, den Studierende beschreiten können sollten, die an der Lerneinheit „Inferenzstatistik mit Randomisierungstests“ teilgenommen haben. Darin wurden durch die Erfahrungen in der Lerneinheit und durch die Aufarbeitung des Forschungsstands (Kapitel 6) antizipierte Schwierigkeiten beschrieben, die in diesem Prozess auftreten können.

In Kapitel 11.5.1 wurden sieben aufgetretene Problemstellen identifiziert. Diese können drei Problembereichen zugeordnet werden, die teilweise miteinander verbunden sind.

Der erste Problembereich umfasst das inhaltliche Verständnis zu einzelnen Konzepten. Hierzu gehören:

- Problemstelle „mögliche Erklärungen für beobachtete Unterschiede finden“
- Problemstelle „Referenzverteilung im Randomisierungstestschema skizzieren“
- Problemstelle „Versuch der Angabe einer Hypothesenwahrscheinlichkeit“
- Problemstelle „Schlussfolgerungen“

Die erste Problemstelle wurde erwartet, allerdings zeigt die vorliegende Studie Schwierigkeiten der Teilnehmer, die entgegengesetzt zu den von Pfannkuch et al. (2015) sind. Aus den Teilergebnissen dieser beiden Studien lässt sich schließen, dass das Konzept der möglichen Erklärungen Schwierigkeiten für Lernende birgt, weshalb es in einer Lerneinheit noch stärker thematisiert werden sollte.

Die Problemstelle, in der das Skizzieren der Referenzverteilung als Schwierigkeit auftrat, ist bei allen Teilnehmern aufgetreten, scheint jedoch im Design der Lerneinheit zu liegen, in der diesem Aspekt zu wenig Beachtung geschenkt wurde. Da alle Teilnehmer die konkrete Verteilung inklusive dem Bereich, der durch den P-Wert beschrieben wird, in Form eines TinkerPlotsgraphen vor Augen hatten, wurde hier vermutlich lediglich zu wenig Sorgfalt an den Tag gelegt beim Übertragen in das Randomisierungstestschema. Dies lässt nicht den Schluss zu, dass die zugehörigen Konzepte Referenzverteilung und P-Wert nicht verstanden wurden.

Die dritte Problemstelle, in der versucht wurde, mit Hilfe des P-Werts eine Angabe über die Wahrscheinlichkeit der Null- oder der Alternativhypothese zu machen, ist ebenfalls erwartet worden und steht im Einklang mit anderen Studien, in denen ähnliche Schwierigkeiten festgestellt wurden. Auch wenn die Einführung in die Logik des statistischen Schließens mit Randomisierungstests im Vergleich zu anderen Ansätzen als leichter zu verstehen beschrieben wird, so zeigt doch insbesondere diese Problemstelle, dass manche, aus anderen Studien zu Lernumgebungen zum Hypothesentesten bekannte Schwierigkeiten trotzdem auftreten können. Desweiteren wurde von keinem Paar der vorliegenden Studie sprachlich korrekt formuliert, dass es sich bei dem erhaltenen P-Wert um

einen geschätzten P-Wert handelt. Auch dies zeigt, dass das Konzept des P-Werts auch im Kontext von Randomisierungstests Schwierigkeiten birgt.

Bezüglich der Problemstelle „Schlussfolgerungen“ konnte identifiziert werden, dass Schlussfolgerungen vor allem statistisch formuliert wurden, allerdings kaum mit dem konkreten Kontext in Bezug gesetzt wurden. Über ähnliche Ergebnisse in Bezug auf Randomisierungstests wird nach bisherigem Forschungsstand wenig berichtet, die Studien von Madden (2008) zeigen ebenfalls, dass Bezug zum Kontext wenig stattfindet und Pfannkuch (2011) hat herausgefunden, dass der konkrete Kontext den Prozess des Schlussfolgerns allgemein nachhaltig beeinflussen kann. Insofern reihen sich die Ergebnisse der vorliegenden Studie in die bestehende Forschung ein, was ein Hinweis darauf ist, dass der Kontextbezug während des gesamten Prozesses, vor allem für mögliche Schlussfolgerungen, explizit thematisiert werden muss.

Der zweite Problembereich umfasst Schwierigkeiten, die auf technischer Ebene liegen. Hierzu gehören

- Problemstelle „Übertragen der Nullhypothese in die Zufallsmaschine in TinkerPlots“
- Problemstelle „Messgrößen sammeln beschleunigen“

Die designte Lerneinheit sollte die instrumentelle Genese Studierender fördern, um TinkerPlots als Instrument zur Durchführung eines Randomisierungstests einzusetzen. Für die Erstellung der Zufallsmaschine müssen in diesem Prozess viele Einstellungen getätigt werden, die offenbar für die Teilnehmer einige Schwierigkeiten bargen. Desgleichen barg vor allem das Beschleunigen des Messgrößensammelvorgangs Schwierigkeiten, da die hierzu nötigen Einstellungen den Teilnehmern nicht präsent waren. Dies lässt sich einerseits damit erklären, dass die instrumentelle Genese der Teilnehmer nicht weit genug fortgeschritten war, und andererseits mit dem natürlichen Prozess des Vergessens, der innerhalb von acht Wochen eintritt, in denen ein Instrument nicht genutzt wird. Mehr zu diesem Punkt findet sich in 12.2.2.5.

Der dritte Problembereich schließlich betrifft die Verknüpfung der verschiedenen Ebenen.

- Problemstelle „Verknüpfung der Kontext- mit der Statistik- und der Softwareebene“

Vor allem während des Simulierens mit TinkerPlots wurde von den Teilnehmern kaum Bezug zur Kontextebene hergestellt. Dies lässt sich möglicherweise ebenfalls damit erklären, dass generell auf Softwareebne Schwierigkeiten aufgetreten sind, die von den Teilnehmern nicht allein gelöst werden konnten und die deshalb die Kommunikation,

aber auch den Bearbeitungsprozess dominiert haben. Diese Problemstelle bedarf besonderer Aufmerksamkeit und sollte in einer Lerneinheit besonders thematisiert werden, da durch die Verknüpfung der Ebenen möglicherweise andere Schwierigkeiten gelöst oder verringert werden können.

Über diese Problemstellen hinaus wurden weitere Schwierigkeiten nach der a priori Analyse erwartet. Vor allem im Bereich des Aufstellens der Null- und der Forschungshypothese wurden theoriegeleitet Schwierigkeiten erwartet, diese sind jedoch bei keinem Paar aufgetreten. Alle Paare haben die beiden Hypothesen korrekt formuliert und die richtige Hypothese für die Simulation genutzt. Als Erklärung hierzu lässt sich die Verbindung aus aufzustellender Forschungsfrage und Umbenennen der Alternativhypothese in Forschungshypothese heranziehen, durch die diesen Schwierigkeiten vorgebeugt wurde.

Desweiteren wurde das Identifizieren des korrekten Bereichs für den P-Wert als Schwierigkeit erwartet, auch dies ist entgegen den Erwartungen bei keinem Paar aufgetreten. Alle haben (trotz zwischenzeitlicher Verwirrungen durch eine ungünstige Softwarebedienung, durch die ein falscher Bereich als P-Wert angezeigt wurde), den richtigen Bereich als P-Wert identifiziert.

Die aufgetretenen und die nicht aufgetretenen Problemstellen geben wertvolle Hinweise für ein mögliches Redesign der Lerneinheit und werden deshalb in 12.2.2.6 noch einmal aufgegriffen.

Nach der a priori Analyse ist über entstehende kollaborative Bearbeitungs- und Kommunikationsprozesse nichts bekannt, deshalb ergab sich die Notwendigkeit, sich mit diesem Prozess intensiv zu beschäftigen, was durch Forschungsfrage 3 geschah.

12.2.2.3 Faktoren für gelingende, kollaborative Kommunikationsprozesse

Die dritte Forschungsfrage „Welche Faktoren für gelingende, kollaborative Bearbeitungsprozesse der Teilnehmer beim Durchführen eines Randomisierungstests mit der Software TinkerPlots können festgestellt werden?“ wird in diesem und den folgenden beiden Abschnitten beantwortet. Zunächst werden Erkenntnisse zu Kommunikationsprozessen zusammengefasst, anschließend (12.2.2.4) Auffälligkeiten bezüglich der einzelnen Konzepte beschrieben und in 12.2.2.5 die Analysen bezüglich der Softwarenutzung zusammengefasst. An einigen Stellen werden in diesen Abschnitten bereits Hinweise auf ein mögliches Redesign der Lerneinheit (Forschungsfrage 4) gegeben, diese werden in Abschnitt 12.2.2.6 zusammengefasst.

Es wurden in Kapitel 11.5 Faktoren bezüglich der Kommunikationsprozesse herausgearbeitet, die kollaborative Bearbeitungsprozesse unterstützen. Eingebettet in die dritte

Forschungsfrage wurde darüber hinaus analysiert, inwieweit die einzelnen Konzepte korrekt adressiert werden konnten und welche Schwierigkeiten dazu auftraten. Als drittes wurde die Softwarebenutzung der Teilnehmer untersucht, um zu identifizieren, an welchen Stellen konkrete Hilfe nötig war, um daraus weitere Faktoren für gelingende kollaborative Bearbeitungsprozesse zu extrahieren.

Bezüglich der Kommunikationsprozesse gibt es zwei wesentliche Ergebnisse. Erstens ist während der selbständigen Bearbeitung der Aufgabe durch die Paare auffällig, dass auf gesprächsanalytischer Makroebene die Eröffnungsphase einer neuen Interaktionseinheit in fast allen Fällen äußerst kurz ist. Häufig umfasst die Eröffnungsphase nur ein Wort oder maximal einen Turn, die Gesprächssteuerung funktioniert aber trotzdem hervorragend. Es ist keine Stelle während der Gespräche gefunden worden, an denen die Teilnehmer dem Wechsel zu einer neuen Einheit nicht folgen konnten oder es deshalb zu Schwierigkeiten kam. Interaktionsanalysen von Gesprächen, die in eine mathematische Lernsituation eingebettet sind, unter Lernenden mit Fokus auf gesprächsanalytische Makroebene finden sich vor allem in Studien über Gesprächsprozesse zwischen Grundschulkindern. Die in der vorliegenden Arbeit gefundenen Resultate hinsichtlich der Gesprächsphasen decken sich mit Ergebnissen der Studie von Schwarzkopf (2000), der in einer Untersuchung von Unterrichtsgeschehen in einer vierten und fünften Klasse ebenfalls Gesprächsprozesse analysiert hat, in denen die Unterscheidung der Gesprächsphasen von den Beteiligten ebenfalls ohne nennenswerte Schwierigkeiten funktionierte. Auch Höck (2015) hat Gespräche von Dritt- und Viertklässlern hinsichtlich der Gesprächsphasen untersucht, bei den dort untersuchten Gesprächen in Lernpartnerschaften konnte ebenfalls beobachtet werden, dass der Übergang zwischen den Gesprächsphasen unproblematisch war, allerdings waren hier die Eröffnungsphasen teilweise deutlich länger, bis sich die Schüler auf ein neues „Thema“ geeinigt hatten. Ergänzend zu diesen vorliegenden Ergebnissen zu Gesprächen zwischen Lernenden in der Primarstufe lassen sich somit ähnliche Ergebnisse auch für Lernende an der Hochschule finden. Es lässt sich deuten, dass die Gesprächsphasen in kleinen Lerngruppen bezüglich mathematischer Kommunikationsprozesse gut funktionieren, ohne dass diese explizit thematisiert werden müssen.

Ein zweites Ergebnis ist, dass auf mikroanalytischer Gesprächsstruktur während der selbständigen Bearbeitung der Aufgabe durch die Teilnehmer sehr häufig in einzelnen Turns nur unvollständige Sätze oder sogar einzelne Wörter gebildet wurden. Gleichermaßen zeichneten sich die Gespräche jedoch durch ein hohes Verständnis der Teilnehmer füreinander aus, denn Sätze wurden durch diese kurzen Turns häufig gemeinschaftlich über mehrere Turns hinweg gebildet und so Formulierungen für bestimmte Sachverhalte gemeinsam konstruiert. Dies weist daraufhin, dass in den kollaborativen Bearbeitungsprozessen Bedeutung gemeinsam ausgehandelt wird, unabhängig ob sich die Paare untereinander kennen oder nicht, so wie dies von Miller (1986) als Lerntheorie

beschrieben wurde und von Krummheuer und Brandt (2001) für mathematische Bearbeitungsprozesse zwischen Grundschulkindern festgestellt wurde. Die Ergebnisse der vorliegenden Arbeit bestätigen in einem anderen Umfeld das Funktionieren des gemeinsamen Aushandelns von Bedeutung in mathematischen Gesprächsprozessen in kleinen Lerngruppen.

Aus diesen Ergebnissen lassen sich sowohl strukturelle als auch inhaltsbezogene Merkmale identifizieren (Kapitel 11.5.3), die die kollaborativen Bearbeitungsprozesse der Studierenden in der vorliegenden Studie begünstigt haben. Ein Vergleich mit den ausführlichen Analysen von Höck (2015) lässt eine Verallgemeinerung dieser Merkmale über die vorliegenden Inhalte hinaus auf generelle kollaborative (mathematische) Bearbeitungsprozesse zu.

Strukturelle Merkmale für kollaborative Bearbeitungsprozesse sind:

- Stabile Eröffnungsphasen,
- Stabile Durchführungsphasen,
- Geteilte Beendigungsphasen,
- Häufige Turnwechsel, für die folgende Situationen unterschieden werden können:
    - Satz- oder Gedankenfortführung durch den jeweiligen Partner,
    - Gemeinsames Formulieren bestimmter Sachverhalte.

Inhaltsbezogene Merkmale, die eine Kollaboration erkennen lassen, sind:

- Fokussierung auf ein bestimmtes Thema,
- Inhaltliche Turnübernahme zur Fortführung oder Ergänzung eines vorherigen Sprecherbeitrags,
- Hervorbringen von als gemeinsam geteilt geltenden Bedeutungen,
- Hervorbringen von gemeinsam formulierten Argumentationen.

12.2.2.4 Auffälligkeiten hinsichtlich der verschiedenen Konzepte im Hinblick auf ein mögliches Redesign der Lerneinheit

Im Sinne des Design-Based Research (P. Cobb et al. 2003) sollen die Erkenntnisse, die aus der Fallstudie gewonnen werden, für die retrospektive Betrachtung der entwickelten Lerneinheit genutzt werden und Hinweise auf ein mögliches Redesign geben. Da ein wesentliches Gestaltungselement der Lerneinheit die in Kapitel 6.4 identifizierten Konzepte sind, werden hierzu besondere Auffälligkeiten berichtet. Auswertungen bezogen auf jedes Konzept finden sich in Kapitel 11.5.4.

Generell lässt sich festhalten, dass alle neun Konzepte von allen Paaren adressiert wurden und keins ausgelassen wurde. Dies mag daran liegen, dass die Konzepte einerseits

durch die Leitfadenfragen in den Interviewteilen der Studie direkt in den Fokus gestellt wurden und dass des Randomisierungstestschemas eine Adressierung während der Phase der selbständigen Bearbeitung herausforderte. Trotzdem ist es ein gutes Ergebnis, dass keins der Konzepte von den Teilnehmern unbearbeitet blieb.

Ein grundlegendes Konzept beim Randomisierungstest ist das der zufälligen Zuordnung, denn darauf bauen das Design des Tests und die möglichen Schlussfolgerungen, die gezogen werden können, auf. Bei Budgett et al. (2012) haben fünf von zehn Teilnehmern beispielsweise die zufällige Aufteilung der Versuchspersonen auf zwei Gruppen nicht als Erklärung dafür erkannt, dass dadurch bezogen auf die zu untersuchenden Merkmale möglichst gleiche Gruppen erzeugt werden sollen. In der Studie der vorliegenden Arbeit scheint dies weniger problematisch für die Teilnehmer zu sein, denn in der entsprechenden Interaktionseinheit nennen alle Paare Merkmale, nach denen die Gruppen hätten eingeteilt werden können, wodurch jedoch eine Schlussfolgerung bezogen auf diese Merkmale hätte gemacht werden müssen. Bei Pfannkuch et al. (2015) sind ähnliche Argumentationen Studierender beobachtbar.

Mögliche Erklärungen für die beobachteten Unterschiede zu nennen, gelingt allen Paaren. Dabei können zwei der drei Paare beide erwünschten Erklärungen nennen (es liegt an der zufälligen Aufteilung bzw. die Behandlung ist wirksam), ein Paar hat Schwierigkeiten damit, die Erklärung der zufälligen Aufteilung zu finden und benötigt hierzu Hilfe der Interviewerin. Auch im weiteren Interviewverlauf stellt sich dieses Konzept als schwierig für dieses Paar heraus. Ein konträres Ergebnis hierzu findet sich bei Pfannkuch et al. (2015), bei denen knapp ein Drittel der Teilnehmer die Wirksamkeit einer Behandlung nicht als mögliche Erklärung nennen konnten. Beim Unterrichten sollte dieses Konzept deshalb noch stärker in den Fokus gerückt werden und mit Lernenden diskutiert werden.

Generell in der Inferenzstatisik gilt das Aufstellen und Formulieren einer passenden Nullhypothese als schwierig (Castro Sotos et al. 2007, Liu & Thompson 2009, Vallecillos 1999). Bezogen auf Randomisierungstests findet sich dies explizit bei Biehler et al. (2015) und bei Frischemeier (2017). Eine ähnliche Schwierigkeit kann für die vorliegende Studie nicht bestätigt werden. Die Teilnehmer hatten keine Schwierigkeiten, die Nullhypothese zu formulieren und verwechselten diese auch nicht mit der Forschungshypothese. Ein Grund hierfür mag im Aufbau des verwendeten Randomisierungstestschemas liegen, in welchem zuerst die Forschungshypothese und erst danach die Nullhypothese formuliert werden sollte. Durch die Assoziation der Forschungshypothese mit der zuvor aufgestellten Forschungsfrage gelang allen Teilnehmern das Aufstellen einer passenden Forschungshypothese problemlos innerhalb kürzester Zeit und daran anschließend wurde von allen Paaren direkt eine passende Nullhypothese formuliert. Auch im weiteren Verlauf bezogen sich alle drei Paare immer wieder auf das

durch die Nullhypothese ausgedrückte Nullmodell, von dem alle davon ausgingen, dass dies dem Randomisierungstest zugrunde liegen musste und als wahr vorausgesetzt wurde. Dies ist ein starkes Indiz dafür, dass die Abfolge *Forschungsfrage - Forschungshypothese - Nullhypothese* den Bearbeitungsprozess unterstützt und aus diesem Grund ist dies ein wichtiges Element, das im Design beibehalten werden soll.

Die nächste Besonderheit bezüglich eines Konzepts liegt im Auffinden und Interpretieren des P-Werts nach der durchgeführten (korrekten) Simulation der Paare. Frischemeier (2017) berichtet von Schwierigkeiten Studierender, einen P-Wert aufzufinden, dies kann in der vorliegenden Studie nicht bestätigt werden, alle Paare identifizierten den richten Wert als P-Wert. Pfannkuch et al. (2015) berichten von der Tendenz Studierender, die Nullhypothese bei einem nicht signifikanten P-Wert fälschlicherweise anzunehmen und von der Interpretation des P-Werts als Hypothesenwahrscheinlichkeit. Beides lässt sich auch in der vorliegenden Studie wiederfinden. Zwei der drei Paare diskutieren aus sich selbst heraus, dass der P-Wert gerade nicht die Wahrscheinlichkeit des Zutreffens der Nullhypothese angibt, bei einem Paar bleibt jedoch bis zum Ende des Interviews trotz mehrfachem Intervenierens durch die Interviewerin die falsche Vorstellung zum P-Wert vorhanden. Dies zeigt, dass das Konzept des P-Werts auch im Kontext von Randomisierungstests für Lernende seine Schwierigkeiten birgt und gewissenhaft in einer Lerneinheit thematisiert werden muss. Im Sinne des Design-Based Research gibt dies einen Hinweis für die Überarbeitung der Lerneinheit, um das Konzept des P-Werts stärker in den Fokus zu stellen.

Bezüglich des Schlussfolgerns lehnen alle Paare bei einem simulierten P-Wert von ca. 1% die Nullhypothese ab. Dabei ist zu beobachten, dass sie sich strikt am Sprachgebrauch der Signifikanz orientieren, die im Baustein Inferenzstatistik eingeführt wurde. Wie auch in anderen Studien zu beobachten, ist der Bezug auf den Kontext, in welchem Schlussfolgerungen gezogen werden sollen, ein schwieriger Schritt (z. B. Madden 2011, Frischemeier 2017). Nur eins der drei Paare formuliert, dass somit auf die Wirksamkeit der Behandlung geschlossen werden könne. Hingegen formulieren alle drei Paare, dass die Schlussfolgerungen nicht verallgemeinert werden dürfen, sondern nur für die Versuchsteilnehmer gelten. Im Unterschied zu Ergebnissen aus Budgett et al. (2012) formulieren alle Paare jedoch die Unsicherheit, die mit den gezogenen Schlüssen einher geht und zeigen, dass ihnen bewusst ist, dass sie mit dem Randomisierungstest keinen Beweis dafür haben, dass die Behandlung wirksam ist, sondern dass immer ein Restrisiko bleibt, eine falsche Entscheidung getroffen zu haben.

12.2.2.5 TinkerPlots als Instrument beim Randomisierungstesten und auftretende Schwierigkeiten

Keins der Paare konnte den Randomisierungstest komplett ohne Hilfe durchführen. Bei der Benutzung der Software TinkerPlots waren bei allen Paaren für mindestens drei der

in 10.4.4 identifizierten notwendigen Schritte in der Software Hilfe durch die Interviewerin notwendig (vgl. Kapitel 11.5.6).

Trotzdem lässt sich durch die Analyse feststellen, dass die Paare alle Schritte in der Software während der Bearbeitung oder im anschließenden Interviewteil gut erklären konnten. Schwierigkeiten schienen hier weniger im Verständnis der Simulation an sich zu liegen, als vielmehr an der konkreten Bedienung des Werkzeugs. Dies lässt sich mit zwei Faktoren erklären. Zum einen hatten die Teilnehmer während der Lerneinheit der Veranstaltungen nur ein einziges Mal Gelegenheit, einen Randomisierungstest selbständig mit TinkerPlots durchzuführen. Zum anderen fanden die Interviews circa acht Wochen nach Vorlesungsende statt, und in der Zwischenzeit hatten die Teilnehmer nicht weiter mit TinkerPlots gearbeitet. Unter diesen Faktoren ist die Softwarebenutzung der Teilnehmer als weniger gut einzustufen, ihr Vermögen, die einzelnen Schritte zu erklären jedoch als gut, da, wie bereits gesagt, die einzelnen Schritte während der Simulation oder im Nachhinein erklärt wurden und in Zusammenhang mit der Aufgabe gebracht werden konnten. Ein Verständnis über die Simulation scheint vorhanden zu sein, bei der konkreten technischen Umsetzung hingegen ergaben sich Situationen, in denen Hilfe benötigt wurde.

Vor allem beim Kopieren der vorhandenen Werte des Experiments in die Zufallsmaschine von TinkerPlots, beim Erstellen des zweiten Bauteils für die Gruppenzuordnung und beim Messgrößensammeln benötigten alle drei Paare direkte Anweisungen durch die Interviewerin. Das Kopieren der Werte in die Zufallsmaschine und das Messgrößensammeln repräsentieren Schwierigkeiten auf einer technischen Ebene (wie muss TinkerPlots hier bedient werden), wohingegen das Erstellen des zweiten Bauteils für die zufällige Gruppenzuordnung auf Schwierigkeiten im Hinblick auf das Konzept der zufälligen Zuordnung hinweisen kann. Die Metapher des Neugruppierens mit Hilfe von Gruppenetiketten, die zufällig den vorhandenen Werten (Blutdruckreduktionen) zugelost werden, konnte von zwei der Paare nicht selbständig mit der Modellierung in der Zufallsmaschine in Verbindung gebracht werden, auch wenn diese Paare auf statistischer Ebene den Vorgang erklären konnten.

Dies passt zu einem der zentralen Ergebnisse der Studie von Noll und Kirin (2017), die das Umbenennen von Merkmalen in der Simulation mit TinkerPlots als zentral ansehen, weil nur dadurch ihre Teilnehmer erkannten, dass die Werte zufällig auf zwei neue Gruppen zugelost werden. Dies kann zwar für die vorliegende Studie nicht bestätigt werden, allerdings zeigt der Gesprächsausschnitt von Rebecca und Selina zu diesem Thema, dass das Bezeichnen der Merkmale ein zentraler Aspekt ist, der das Verständnis des Konzepts der zufälligen Zuordnung stark fördern kann. Insofern sollte dies in einer Lerneinheit entsprechend thematisiert werden.

Allgemein zeigt sich nach der erfolgten Interaktionsanalyse, dass die *instrumented action schemes* (Rabardel 2002) in Bezug auf die Softwarebenutzung zur Simulation für Randomisierungstests bei allen Paaren recht gut ausgebildet sind, denn alle zeigen immer wieder Gesprächsphasen, in denen sie im Voraus planen, welche Schritte als nächstes in TinkerPlots nötig sind, auch wenn sie für die konkrete Umsetzung teilweise Hilfe durch die Interviewerin benötigen. Eine gute instrumentelle Genese im Hinblick auf die Durchführung eines Randomisierungstests ist also eine wichtige Voraussetzung, damit Lernende Software als Instrument nutzen können und diese als „Mittel zum Zweck" einsetzen können.

Das Randomisierungstestschema wurde von den Paaren vor allem als Ressource im Sinn von Guin et al. (2005) genutzt, um die Aufgabe zu dokumentieren. Dabei stellte das Formulieren für das Schema manchmal eine Herausforderung dar, durch welche die Paare jedoch gezwungen waren, die verschiedenen Konzepte genau zu durchdenken und exakte Formulierungen auszuhandeln. Der Nutzen dieses (oder eines ähnlichen Schemas) ist offensichtlich, weshalb Schemata dieser Art in einer entsprechenden Lerneinheit auf jeden Fall eingefügt werden sollten.

12.2.2.6 Hinweise für ein Redesign der Lerneinheit

Zur Beantwortung der Forschungsfrage 4 „Welche Erkenntnisse aus den Bearbeitungsprozessen der Lernenden in den produktiven und den nicht gelingenden Phasen lassen sich für ein Redesign der Lerneinheit gewinnen (Aufgabenstellung, Anleitung der instrumentellen Genese, unterstützende didaktische Maßnahmen)?" werden die Erkenntnisse der vorherigen Abschnitte zusammengefasst.

- Die generelle Struktur der Lerneinheit hat sich bewährt, die neun Konzepte haben Lernende befähigt, einen Randomisierungstest durchzuführen.
- Das Randomisierungstestschema hat sich als nützliches Werkzeug herausgestellt, um einen Randomisierungstest zu strukturieren und den Teilnehmern eine Orientierung zu bieten. Als Verbesserung sollte das Skizzieren der Referenzverteilung stärker thematisiert und ggf. durch ein weiteres Stichwort (z. B. Einzeichnen des beobachteten Werts) ergänzt werden.
- Die Abfolge aus *Forschungsfrage - Forschungshypothese - Nullhypothese*" hat sich bewährt, um Schwierigkeiten bezüglich der Nullhypothese vorzubeugen.
- Die Aktivität der händischen Simulation hat sich bewährt, um den Prozess des Randomisierens zu verstehen und wieder in Erinnerung zu rufen. Eine solche Aktivität sollte auf jeden Fall in der Lerneinheit beibehalten werden.
- Die instrumentelle Genese von Lernenden im Hinblick auf Randomisierungstests sollte in mehr als einer 20 minütigen Aktivität angeleitet werden, damit die nötigen *instrumented action schemes* ausgebildet werden. Damit verbunden soll-

te eine tiefergehende Verbindung zwischen dem Kontext und allen nötigen Einstellungen der Zufallsmaschine erläutert werden. Die händische Simulation sollte mit der computergestützten Simulation ebenfalls sprachlich stärker verbunden werden, indem bei der Erläuterung darauf Bezug genommen wird.

- Für die Konzepte „mögliche Erklärungen für beobachtete Unterschiede“ und „mögliche Schlussfolgerungen“ sollte mehr Zeit eingeräumt werden, da diese Schwierigkeiten für die Teilnehmer bargen.
- Bezüglich der verwendeten Aufgaben wurden keine besonderen Erkenntnisse gewonnen, so dass diese als geeignet eingeschätzt werden.

12.3 Bedeutung der vorgelegten Studie für die Stochastikdidaktik

Die vorliegende Arbeit leistet einen Erkenntnisbeitrag für die Stochastikdidaktik auf drei Ebenen.

12.3.1 Bedeutung der Werkzeuganalyse

Erstens wurde umfangreich das Unterstützungspotential der Software TinkerPlots zur stochastischen Simulation untersucht. Es wurden verschiedene Simulationsarten identifiziert und gezeigt, inwiefern sich Simulationen intuitiv mit der Zufallsmaschine umsetzen lassen und somit mentale Modelle in der Software ausgedrückt werden können. Somit können Lernumgebungen, Aufgaben und Lösungshorizonte gezielt gestaltet werden und mögliche Schwierigkeiten von Lernenden können besser eingeordnet werden. Dieses Wissen kann in leicht abgewandelter Form auch verwendet werden, um zum Beispiel mit der Weiterentwicklung von TinkerPlots in der frei zugänglichen, webbasierten Software CODAP[49] zu arbeiten, in die zum jetzigen Zeitpunkt (Sommer 2018) eine rudimentäre Version der Zufallsmaschine aus TinkerPlots implementiert ist, die kontinuierlich weiterentwickelt wird. Außerdem kann die vorliegende Werkzeuganalyse auch Anregungen geben, Werkzeuge in anderen Bereichen (z. B. GeoGebra) genauer zu analysieren und zu verstehen.

12.3.2 Bedeutung des Forschungsüberblicks zu Randomisierungstests

Zweitens wurde in dieser Arbeit ein Überblick gegeben über empirische Studien und fachdidaktische Artikel aus der Forschung zum Einführen in die Logik der Inferenzstatistik mit Randomisierungstests. Erkenntnisse aus diesem Überblick sind in die Gestaltung des Bausteins Inferenzstatistik eingeflossen, dessen Lerntrajektorie in Kapitel 8.2.2 beschrieben ist und der eine Einführung in die Logik des inferentiellen Schließens mit Randomisierungstests innerhalb einer kurzen Zeitspanne darstellt. Es wurden Konzepte identifiziert, die in der Stochastikdidaktik verwendet werden können, um in die Logik

[49] http://codap.concord.org/ (abgerufen am 10.06.2018).

des inferentiellen Schließens mit Randomisierungstests einzuführen und eigene Lerneinheiten zu gestalten. Der in Kapitel 8.2.2 beschriebene Baustein Inferenzstatistik kann Anregungen liefern, um ähnliche Lerneinheiten zu konzipieren.

12.3.3 Bedeutung der explorativen Fallstudie

Drittens wurde in einer explorativen Fallstudie untersucht, wie die Bearbeitungsprozesse von Lernenden beim Durchführen eines Randomisierungstests mit TinkerPlots aussehen und wie die Software in diesem Prozess eingesetzt wird.

Die verwendeten interpretativen Methoden haben sich als sehr nützlich herausgestellt, um detailliert in den Bearbeitungsprozess der Lernenden zu blicken. Hierdurch ließen sich Erkenntnisse gewinnen, die andernfalls nicht möglich wären.

Es hat sich gezeigt, dass die sechs Studierenden mit dem Vorwissen aus einer Grundlagenveranstaltung in Stochastik („Modellieren, Größen, Daten und Zufall I“), in welche eine sechs Semesterwochenstunden umfassende Lerneinheit zur Inferenzstatistik implementiert wurde, mit Hilfestellung in Bezug auf die Softwarebenutzung in der Lage waren, einen Randomisierungstest mit der Software TinkerPlots durchzuführen und die Logik der Inferenzstatistik zu verinnerlichen. Um den Prozess des Randomisierens im Randomisierungstest zu verstehen, haben alle Teilnehmer im Interview geäußert, dass das händische Simulieren im Seminar hierfür sehr geholfen hat. Ein Paar hat auch während des Interviews eine händische Simulation durchgeführt, um sich den Prozess in Erinnerung zu rufen. Im Einklang mit den Vorschlägen von G. Cobb (2007), Fitch und Regan (2014), Holcomb, Chance, Rossman, Tietjen, et al. (2010) und Rossman (2008) wird deshalb empfohlen, eine solche Aktivität in eine entsprechende Lerneinheit zu implementieren, um diesen so wichtigen Vorgang konkret erlebbar und dadurch besser verstehbar zu machen.

Es hat sich auch gezeigt, dass es nicht reicht, ein Werkzeug nur zu benutzen, um die Logik des inferentiellen Schließens zu verstehen. Das Werkzeug muss als Instrument (Rabardel 2002) beherrscht werden, wozu eine genaue Analyse des Werkzeugs notwendig ist, um den nötigen Prozess der instrumentellen Genese, d. h. des Entwickelns der Benutzung eines Werkzeugs zur Verwendung als Instrument, bei Lernenden zu steuern. In diesem Fall bedeutet das, dass das Beherrschen des Instruments durch den Lerner notwendig ist, um sich auf die Logik des statistischen Schließens konzentrieren zu können. In der vorliegenden Studie wurde an diversen Stellen Hilfe bei der Bedienung der Software benötigt, hieraus lässt sich ableiten, dass die instrumentelle Genese der Lernenden nicht weit genug fortgeschritten war, um auch acht Wochen nach der Lerneinheit eine Simulation zum Randomisierungstest selbständig durchzuführen. Für eine möglicherweise erneute Durchführung der Lerneinheit sollte für die instrumentelle Genese der Lernenden im Hinblick auf einen durchzuführenden Randomisierungstest das

Werkzeug als Instrument beherrscht werden. Hierzu sollte für die eigenständige Arbeit der Lernenden mehr Zeit in der Lerneinheit eingeplant werden, als dies in der vorliegenden Einheit der Fall war.

Insgesamt zeigt die vorliegende Arbeit detailliert auf, wie die verschiedenen zu einem Randomisierungstest gehörigen Konzepte von Lernenden unter Einsatz der Simulationssoftware TinkerPlots bearbeitet werden und somit Möglichkeiten und Herausforderungen für Lehrende, auf welche Aspekte besonders geachtet werden sollte. Gerade die in 12.2.2.6 zusammengefassten Erkenntnisse können wesentliche Ideen für die Gestaltung einer Lerneinheit zur Inferenzstatistik mit Randomisierungstests liefern.

Für wissenschaftliche Forschende gibt die vorliegende Auswertung der explorativen Studie einen detailreichen Einblick in die verwendeten Methoden der Interaktionsanalyse und der funktionalen Argumentationsanalyse. Beide Methoden haben sich als hilfreich herausgestellt, um tiefe Einblicke in die kognitiven Prozesse von Lernenden zu bekommen. Allerdings muss immer beachtet werden, dass es sich hierbei um interpretative Methoden handelt, die stark subjektiv sind, weshalb die Diskussion der getätigten Interpretationen mit einem oder mehreren unabhängigen Forschern unerlässlich ist. Gerade hierdurch lässt sich jedoch ein hohes Verständnis von Lernerschwierigkeiten erlangen.

12.3.4 Empfehlungen für die Gestaltung einer Lerneinheit zur Logik der Inferenzstatistik mit Randomisierungstests

Aus den Ergebnissen der Fallstudie können verschiedene Empfehlungen gefolgert werden, die für das Unterrichten einer Lerneinheit beachtet werden sollten.

- Eine händische Simulation zu implementieren hat sich als nützlich herausgestellt, um den Prozess des Randomisierens einerseits zu verstehen und andererseits später wieder in Erinnerung zu rufen. Deshalb wird empfohlen, eine solche Aktivität auf jeden Fall in einer Lerneinheit zu integrieren.
- TinkerPlots eignet sich als Werkzeug für eine solche Lerneinheit gut, weil das mentale Modell, das sich ein Nutzer von der Randomisierung bildet, gut in der Software abgebildet werden kann. Außerdem können die einzelnen Schritte (Testgröße identifizieren, Referenzverteilung erstellen, P-Wert bestimmen) aktiv durchgeführt werden im Vergleich zu anderer Software und erzeugen somit ein höheres Verständnis.
- Unabhängig von der Wahl des Werkzeugs sollte die instrumentelle Genese der Lernenden gut angeleitet werden, damit das Werkzeug als Instrument genutzt werden kann und die kognitiven Ressourcen zur Durchführung des Randomisierungstests verwendet werden können und nicht zur Bedienung des Werkzeugs benötigt werden.

- Die neun in Kapitel 6.4 vorgestellten Konzepte sollten in einer Lerneinheit adressiert werden, dabei sollte ein besonderer Fokus auf die Konzepte „mögliche Erklärungen für beobachtete Unterschiede“ und „mögliche Schlussfolgerungen“ gelegt werden.
- Ein Schema zur Durchführung eines Randomisierungstests sollte verwendet werden, um den *cognitive load* zu verringern und das Vorgehen zu strukturieren.

Für ein mögliches Redesign des Bausteins Inferenzstatistik der vorliegenden Arbeit im Sinne des Design-Based Research Ansatzes (P. Cobb et al. 2003) sollten diese Empfehlungen für die nächste Durchführung in besonderer Weise aufgenommen werden.

12.4 Ausblick

Die Ergebnisse der vorliegenden Arbeit zeigen Perspektiven für weitere Forschung auf.

Die durchgeführte Werkzeuganalyse bildet die Voraussetzung für Untersuchungen, wie Lernende die Software TinkerPlots für die stochastische Simulation einsetzen. Hier eröffnet sich insbesondere das Forschungsfeld, wie TinkerPlots die Modellierung stochastischer Zufallsexperimente durch die visuelle Umsetzung mentaler Modelle unterstützt. Da die Software erfolgreich in Deutschland, den USA, Israel und Großbritannien in der Sekundarstufe I eingesetzt wird, lohnt sich hier ein Sprung in dieses Feld und dadurch insbesondere die Beforschung von jungen Lernenden, wie diese mit Hilfe der Zufallsmaschine einen Zugang zu stochastischen Situationen erhalten und wie sie diese modellieren. Hierzu ist es erforderlich, passende Lernumgebungen zu entwickeln, in denen TinkerPlots oder als Weiterentwicklung auch CODAP zur stochastischen Simulation eingesetzt wird. In Reichert (2014) wurde eine solche Lernumgebung beispielhaft für die Sekundarstufe I beschrieben und evaluiert. Ben-Zvi und Aridor (2014) und Lehrer (2017) beschreiben ebenfalls solche Lernumgebungen.

Ein weiterer möglicher Ansatzpunkt für weitere Forschung ist sicherlich die Weiterentwicklung des Bausteins Inferenzstatistik aufgrund der in dieser Arbeit gewonnen Erkenntnisse. Da die Veranstaltungen „Modellieren, Größen, Daten und Zufall I“ und „Computergestützte Lernumgebungen für Modellieren, Größen, Daten und Zufall I“ ab dem Sommersemester 2018 aufgrund der erneuerten Prüfungsordnung der Universität Paderborn zu einer Veranstaltung „Stochastik und ihre Didaktik“ als Vorlesung mit Übung zusammengelegt werden, kann und sollte der Baustein Inferenzstatistik für diese Veranstaltung weiter entwickelt und genutzt werden. Für das laufende Sommersemester 2018 geschieht bereits eine solche Überarbeitung. Hier kann sich weitere Forschung anschließen. Beispielsweise ist angedacht, im Rahmen der Abschlussklausur eine passende Aufgabe zu implementieren, um den Wissensstand aller Teilnehmer zu Randomisierungstests zu erheben, anstatt nur die kognitiven Prozesse einzelner Teilnehmer zu untersuchen.

Auch eine generelle Weiterentwicklung oder Adaption des Bausteins durch andere Lehrende kann auf dieser Arbeit aufbauen. Wertvoll wäre sicherlich weitere Forschung, wie Lernende, möglicherweise auch in der Schule, die Logik der Inferenzstatistik durch eine Einführung in Randomisierungstests verstehen. Hierzu gerade erschienen ist ein Artikel von Justice, Zieffler, Huberty und delMas (2018), der sich genau damit befasst. Darüber hinaus wären weitere qualitative, dann aber vor allem auch quantitative Studien sinnvoll. Eine Weiterentwicklung könnte inhaltlich eine Fortführung des Themas beinhalten. Zum Beispiel könnten Bootstraps thematisiert werden, wie in Engel (2007), Pfannkuch und Budgett (2014) und Wild, Pfannkuch, Regan und Parsonage (2015) vorgeschlagen, um Schlussfolgerungen von der Stichprobe auf die Population machen zu können.

Mögliche weitere Forschung, die sich an die vorliegende Arbeit anschließen kann, ist die Auswertung weiterer Daten. So können beispielsweise die Reflexionen über den Baustein Inferenzstatistik der Teilnehmer der Veranstaltung „Modellieren, Größen, Daten und Zufall I“ aus dem Sommersemester 2017 genutzt werden, um zu analysieren, wie die affektiven Einstellungen der Studierenden hinsichtlich dieses Baustein sind und eine Weiterentwicklung darauf aufbauen.

Dieser Abschnitt hat nur einen Teil möglicher Forschung für die Zukunft aufgezeigt. Allgemein bleibt das statistische Schließen und die Frage, wie dies bei Lernenden entwickelt werden kann, ein wichtiges Forschungsfeld der stochastikdidaktischen Forschung der nächsten Jahre.

Literaturverzeichnis

Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik. (2003). Empfehlungen zu Zielen und zur Gestaltung des Stochastikunterrichts. *Stochastik in der Schule, 23*(3), 21-26.

Artigue, M. (2002). Learning Mathematics in a CAS Environment: The Genesis of a Reflection about Instrumentation and the Dialectics Between technical and coneptual Work. *International Journal of Computers for Mathematical Learning, 7*, 245-274.

Artigue, M., Haspékian, M., Cazes, C., Bottino, R. M., Cerulli, M., Kynigos, C., et al. (2006). *Methodological tools for Comparison of learning theories in technology enhanced learning in mathematics*. (D20.4.1) EU Sixth Framework programme priority 2. <hal-00190114>.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Bakker, A. (2002). *Route-type and landscape-type software for learning statistical data analysis*. Paper presented at the Sixth International Conference of Teaching Statistics (ICOTS6), Cape Town, South Africa.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distributions. In D. Ben-Zvi & J. Garfield (Hrsg.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (S. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bakker, A., & van Eerde, D. (2015). An Introduction to Design-Based Research with an Example From Statistics Education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Hrsg.), *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods* (S. 429-466). Dordrecht: Springer Netherlands.

Barzel, B. (2006). Ich-Du-Wir... Sich mit einem Thema wirklich auseinandersetzen. *Mathematik lehren, 139*, 19-21.

Batanero, C. (2000). Controversies Around the Role of Statistical Tests in Experimental Research. *Mathematical Thinking and Learning, 2*(1-2), 75-97.

Batanero, C., & Borovcnik, M. (2016). *Statistics and Probability in High School*. Rotterdam, Boston, Taipei: Sense Publishers.

Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics - Challenges for teaching and teacher education*. The Netherlands: Springer.

Bauer, D., & Röckmann, D. (1999). *Methoden der Bewertung von Software*. (785). Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Bayer, K. (2007). *Argument und Argumentation. Logische Grundlagen der Argumentationsanalyse*. Göttingen: Vandenhoeck & Ruprecht.

Beck, C., & Maier, H. (1993). Das Interview in der mathematikdidaktischen Forschung. *Journal für Mathematikdidaktik, 14*(2), 147-179.

Beck, C., & Maier, H. (1994). Mathematikdidaktik als Textwissenschaft. *Journal für Mathematik-Didaktik, 15*(1), 35-78. doi:doi.org/10.1007/BF03338800

Ben-Zvi, D., & Aridor, K. (2014). *Students' emergent roles in developing their reasoning about uncertainty and modeling*. Paper presented at the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Ben-Zvi, D., & Ben-Arush, T. (2014). EDA Instrumented Learning with TinkerPlots. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen. Using Tools for Learning Mathematics and Statistcs* (S. 193-208). Wiesbaden: Springer Spektrum.

Ben-Zvi, D., & Garfield, J. (2004). *The Challenge of Developing Statistical Litercy, Reasoning and Thinking*. Dordrecht: Springer Science+Business Media.

Bergmann, J. R. (1981). Ethnomethodologische Konversationsanalyse. In P. Schröder & H. Steger (Hrsg.), *Dialogforschung* (S. 9-52). Düsseldorf: Schwann.

Bezold, A. (2009). *Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule*. Hamburg: Verlag Dr. Kovac.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Hrsg.), *Chance Encounters: Probability in Education* (S. 169-211). The Netherlands: Kluwer Academic Publishers.

Biehler, R., & Engel, J. (2015). Stochastik: Leitidee Daten und Zufall. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thime, & H. G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 221-251). Berlin, Heidelberg: Springer.

Biehler, R., & Frischemeier, D. (2016). Randomisierungstests mit TinkerPlots. *Stochastik in der Schule, 16*(3), 9-15.

Biehler, R., Frischemeier, D., & Podworny, S. (2015). Preservice teachers' reasoning about uncertainty in the context of randomization tests. In A. Zieffler & E. Fry (Hrsg.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning*. Minneapolis, Minnesota: Catalyst Press.

Biehler, R., Frischemeier, D., & Podworny, S. (2017). Elementary preservice teachers' reasoning about modeling a "family factory" with TinkerPlots - A pilot study. *Statistics Education Research Journal, 16*(2), 244-289.

Biehler, R., Kombrink, K., & Schweynoch, S. (2003). MUFFINS: Statistik mit komplexen Datensätzen – Freizeitgestaltung und Mediennutzung von Jugendlichen. *Stochastik in der Schule, 23*(1), 11-25.

Biehler, R., & Maxara, C. (2007). Integration von stochastischer Simulation in den Stochastikunterricht mit Hilfe von Werkzeugsoftware. *Mathematikunterricht, 53*(3), 45-61.

Biehler, R., & Prömmel, A. (2013). Von ersten stochastischen Erfahrungen mit großen Zahlen bis zum 1/√ n–Gesetz – ein didaktisch orientiertes Stufenkonzept. *Stochastik in der Schule, 33*(2), 14-25.

Blum, W., Drüke-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen*. Berlin: Cornelsen.

Blumer, H. (1969). *Symbolic Interactionism. Perspective and Method*. Englewood Cliffs, New Jersey: Prentice Hall.

Bohnsack, R. (2014). *Rekonstruktive Sozialforschung*. Opladen, Toronto: Budrich.

Borchardt, A., & Göthlich, S. E. (2007). Erkenntnisgewinn durch Fallstudien. In S. Albers, D. Klapper, U. Konradt, A. Walter, & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (S. 33-48). Wiesbaden: Gabler.

Borovcnik, M. (2014). Forschungsprozess und probabilistische Modellbildung – Stochastische Denkweisen. In J. Maaß & H.-S. Siller (Hrsg.), *Neue Materialien für einen realitätsbezogenen Mathematikunterricht 2, Realitätsbezüge im Mathematikunterricht,* (S. 11-30). Wiesbaden: Springer.

Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die Klinische Forschung*. Heidelberg: Springer Medizin Verlag.

Boxtel, C. v., Linden, J. v. d., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311-330. http://dx.doi.org/10.1016/S0959-4752(00)00002-5

Brandt, B. (2006). Kinder als Lernende im Mathematikunterricht der Grundschule. In H. Jungwirth & G. Krummheuer (Hrsg.), *Der Blick nach innen. Aspekte der alltäglichen Lebenswelt Mathematikunterricht* (Bd. 1, S. 19-51). Münster: Waxmann.

Brinker, K., & Sager, S. F. (2010). *Linguistische Gesprächsanalyse. Eine Einführung* (5. Aufl.). Berlin: Erich Schmidt.

Bruner, J. (1971). *Studien zur kognitiven Entwicklung*. Stuttgart: Ernst Klett.

Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte*. Berlin, Heidelberg: Springer Spektrum.

Büchter, A., & Henn, H.-W. (2007). *Elementare Stochastik. Eine Einführung in die Mathematik der Daten und des Zufalls*. Berlin, Heidelberg: Springer.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. (2012). *Dynamic visualizations for inference*. Paper presented at the The International Association for Statistical Education Roundtable Conference: Technology in statistics education: Virtualities and Realities, Cebu City, The Philippines.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education, 7*(2). Zugriff unter https://escholarship.org/uc/item/9dg6h7wb

Bueno-Ravel, L., & Gueudet, G. (2007). *Online Resources in mathematics teachers genesis of use.* Paper presented at the Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education, Larnaca, Cyprus.

Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill, & C. Reading (Hrsg.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education: A joint ICMI/IASE study* (S. 57-69). Dordrecht: Springer Science+Business Media.

Busse, A., & Borromeo-Ferri, R. (2003). Methodological reflections on a three-step-design combining observation, stimulated recall and interview. *Zentralblatt für Didaktik der Mathematik, 35*(6), 257-264.

Carlton, M. (2003). Comment on "Confusion over measures of evidence (p's) versus errors (a's) in classical statistical testing". *The American Statistician, 57*(3), 179-181. Zugriff unter http://www.tandfonline.com/doi/abs/10.1198/0003130031856

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98-113.

Chance, B., & Rossman, A. (2006). *Using simulation to teach and learn statistics*. Paper presented at the The Seventh International Conference on Teaching Statistics (ICOTS7), Salvador, Brazil.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chaput, B., Girard, J.-C., & Henry, M. (2011). Frequentist Approach: Modelling and Simulation in Statistics and Probability Teaching. In C. Batanero, G. Burrill, & C. Reading (Hrsg.), *Teaching Statistics in School Mathematics - Challenges for Teaching and Teacher Education* (S. 85-95). Dordecht: Springer.

Cobb, G. (2007). The Introductory Statistics Course: A Ptolemaic Curriculum? *Technology Innovations in Statistics Education, 1*(1), 1-15. https://escholarship.org/uc/item/6hb3k0nz

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9-13.

Cumming, G., Williams, J., & Fidler, F. (2004). Replication, and reserachers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3*, 299-311.

Dahm, M. (2006). *Grundlagen der Mensch-Computer-Interaktion*. München: Pearson Studium.

DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28-58.

Deppermann, A. (2008). *Gespräche analyiseren. Eine Einführung*. Wiesbaden: Verlag für Sozialwissenschaften.

Deppermann, A. (2014). Das Forschungsinterview als soziale Interaktionspraxis. In G. Mey & K. Mruck (Hrsg.), *Qualitative Forschung* (S. 133-150). Wiesbaden: Springer.

Doerr, H. M., & Pratt, D. (2008). The Learning of Mathematics and Mathematical Modeling. In K. M. Heid & G. W. Blume (Hrsg.), *Research on technology and the teaching and learning of mathematics: research syntheses* (S. 259-285). Charlotte, North Carolina: Information Age Publishing.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2. Aufl.). New York: John Wiley & Sons.

Dresing, T., & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse. Anleitung und Regelsysteme für qualitativ Forschende* (Bd. 6.). Marburg. Zugriff unter http://www.audiotranskription.de/praxisbuch

Drijvers, P. (2012). Teachers Transforming Resources into Orchestrations. In G. Gueudet, B. Pepin, & L. Trouche (Hrsg.), *From Text to 'Lived' Resources: Mathematics Curriculum Materials and Teacher Development* (S. 265-281). Dordrecht: Springer Netherlands.

Drijvers, P., Doorman, M., Boon, P., Reed, H., & Gravemeijer, K. (2010). The teacher and the tool: instrumental orchestrations in the technology-rich mathematics classroom. *Educational Studies in Mathematics, 75*(2), 213-234. doi:10.1007/s10649-010-9254-5

Drijvers, P., Doorman, M., Boon, P., & van Gisbergen, S. (2010). Instrumental Orchetration: Theory and Practice. In V. Duran-Guerrier, S. Soury-Lavergne, & F. Arzarello (Hrsg.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education* (S. 1349-1358). Lyon, France: INRP.

Drijvers, P., Doorman, M., Boon, P., van Gisbergen, S., & Gravemeijer, K. (2007). Tool use in a technology-rich learning arrangement for the concept of function. In D. Pitta, P. Philippou, & G. Philippou (Hrsg.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (S. 1389-1398). Larnaca, Cyprus.

Drijvers, P., Kieran, C., Mariotti, M.-A., Ainley, J., Andresen, M., Chan, Y. C., et al. (2010). Integrating Technology into Mathematics Education: Theoretical Perspectives. In C. Hoyles & J.-B. Lagrange (Hrsg.), *Mathematics Education and Technology-Rethinking the Terrain* (S. 89-132). Boston: Springer.

Drijvers, P., & Trouche, L. (2008). From Artifacts to Instruments: A Theoretical Framework behind the Orchestra Metaphor. In K. M. Heid & W. B. Glendon (Hrsg.), *Research on technology and the teaching and learning of mathematics* (Bd. 2, S. 363-392). Charlotte, North Carolina: Information Age Puiblishing.

Edgington, E. S. (1995). *Randomization tests*. New York: Dekker.

Eemeren, F. H. v., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E. C. W., et al. (1996). *Fundamentals of argumentation theory: A handbook of historical background and contemporary developments*. Hillsdale, New Jersey: Lawrence Erlbaum.

Eichler, A., & Vogel, M. (2013). *Leitidee Daten und Zahlen. Von konkreten Beispielen zur Didaktik der Stochastik*. Wiesbaden: Springer.

Engel, J. (2007). On teaching the bootstrap. *Bulletin of the International Statistical Institute 56th Session,* Lisbon.

Engel, J. (2014a). Change point detection tasks to explore students' informal inferential reasoning. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen. Using Tools for Learning Mathematics and Statistics* (S. 113-125). Wiesbaden: Springer Spektrum.

Engel, J. (2014b). *Open data, civil society and monitoring progress: challenges for statistics education*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Engel, J. (2017). Statistical Literacy for active Citizenship: A Call for Data Science Education. *Statistics Education Research Journal, 16*(1), 44-49.

Engel, J., & Grübel, R. (2008). Bootstrap – oder die Kunst, sich selbst aus dem Sumpf zu ziehen. *Mathematische Semesterberichte, 55*(2), 113-130.

Erickson, T. (2006). *Using simulation to learn about inference*. Paper presented at the The Seventh International Conference on Teaching Statistics (ICOTS7), Salvador, Brazil.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statistical Science, 19*(4), 676-685.

Fetzer, M. (2006). Veröffentlichen im Mathematikunterricht - ein Beitrag zu einer Interaktionstheorie grafisch basierten Lernens. In H. Jungwirth & G. Krummheuer (Hrsg.), *Der Blick nach innen: Aspekte der alltäglichen Lebenswelt Mathematikunterricht.* (Bd. 1). Münster: Waxmann.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edingburgh, UK: Oliver & Boyd.

Fisher, R. A. (1935). *The Design of Experiments*. New York: Hafner Publishing Company.

Fitch, M., & Regan, M. (2014). *Accepting the challenge: Constructing a randomisation pathway for inference into our traditional introductory course*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Flick, U. (2013). Design und Prozess qualitativer Forschung. In U. Flick, E. von Kardoff, & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (S. 252-264). Reinbek bei Hamburg: Rowohlt.

Flick, U. (2016). *Qualitative Sozialforschung. Eine Einführung* (7. Aufl.). Reinbek bei Hamburg: Rowohlt.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs. Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education, 32*(2), 124-158.

Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden: Springer Spektrum.

Frischemeier, D., & Biehler, R. (2014). Design and exploratory evaluation of a learning trajectory leading to do randomization tests facilitated by TinkerPlots. *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education*, 799-809.

Frischemeier, D., & Podworny, S. (2017). *Eine innovative Stochastiklehrveranstaltung für Grundschullehramtsstudierende*. Poster präsentiert auf dem Tag der Lehre der Universität Paderborn. Paderborn.

Gal, I. (2002). Adults' Statistical Literacy: Meaning, Components, Responsibilities. *International Statistical Review, 70*(7), 1-51.

Garfield, J., & Ben-Zvi, D. (2008a). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer Science+Business Media.

Garfield, J., & Ben-Zvi, D. (2008b). Helping Students Develop Statistical Reasoning: Implementing a Statistical Reasoning Learning Environment. *Teaching Statistics, 31*(3), 72-77.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM, 44*(7), 883-898.

Garfinkel, H. (1969). *Studies in Ethnomethodology*. New Jersey: Prentice Hall.

Garz, D. (1995). Entwicklungslinien qualitativ-empirischer Sozialforschung. In E. König & P. Zedler (Hrsg.), *Bilanz qualitativer Forschung. Band I: Grundlagen qualitativer Forschung* (S. 11-32). Weinheim: Deutscher Studien-Verlag.

Gigerenzer, G., & Gaissmaier, W. (2006). Denken und Urteilen unter Unsicherheit: Kognitive Heuristiken. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 330-374). Göttingen: Hogrefe.

Gigerenzer, G., Swijtink, Z., Porter, T. M., Daston, L., Beatty, J., & Krüger, L. (1999). *Das Reich des Zufalls*. Heidelberg: Spektrum.

Gnanadesikan, M., Scheaffer, R. L., & Swift, J. (1987). *The Art an Techniques of Simulation*. Canada: Dale Seymour Publications.

Good, P. (2005). *Permutation Tests : A Practical Guide to Resampling Methods for Testing Hypotheses* New York: Springer.

Götz, S. (1993). Verteilungsfreie Testverfahren. *Schriftenreihe zur Didaktik der Mathematik der Ölsterreichischen Mathematischen Gesellschaft, 21*, 47-80.

Gould, R., Davis, G., Patel, R., & Esfandiari, M. (2010). *Enhancing conceptual understanding with data driven labs*. Paper presented at the The Eighth International Conference on Teaching Statistics (ICOTS8), Ljubljana, Slovenia.

Grune, C. (2000). *Lernen in Computernetzen. Analyse didaktischer Konzepte für vernetzte Lernumgebungen*. München: KoPäd Verlag.

Gueudet, G., Pepin, B., & Trouche, L. (2012). *From Text to 'Lived' Resources. Mathematics Curriculum Materials and Teacher Development*. Heidelberg, London, New York: Springer.

Gueudet, G., & Trouche, L. (2009). Towards new documentation systems for mathematics teachers? *Educational Studies in Mathematics, 71*(3), 199-218.

Guin, D., Ruthven, K., & Trouche, L. (2005). *The Didactical Challenge of Symbolic Calculators. Turning a Computational Device into a Mathematical Instrument*. USA: Springer.

Guin, D., & Trouche, L. (1999). The complex process of converting tools into mathematical instruments: the case of calculators. *International Journal of Computers for Mathematical Learning, 3*, 195-227.

Guin, D., & Trouche, L. (2002). Mastering by the teacher of the instrumental genesis in CAS environments: necessity of intrumental orchestrations. *ZDM, 34*(5), 204-211.

Hadas, N., & Hershkowitz, R. (2002). *Activity analyses at the service of task design*. Paper presented at the 26th Conference of the international Group for the Psychology of Mathematics Education, Norwich, Unided Kingdom.

Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: Springer Fachmedien.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research Online, 7*(1), 1-20.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. F. Burrill, & C. Reading (Hrsg.), *Teaching Statistics in School Mathematics - Challenges for Teaching and Teacher Education* (S. 235-246). The Netherlands: Springer Science+Business Media.

Hartwig, R. (2007). *Ergonomie interaktiver Lernmedien: Kriterien und Entwicklungsprozesse für E-Learning-Systeme*. München: Oldenbourg Wissenschaftsverlag.

Hasemann, K., & Mirwald, E. (2012). Daten, Häufigkeit und Wahrscheinlichkeit. In G. Walther, D. van den Heuvel-Panhuizen, D. Granzer, & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule: Mathematik konkret* (S. 141-161). Berlin: Cornelson Scriptor.

Hattermann, M. (2011). *Der Zugmodus in 3D-dynamischen Geometriesystemen (DGS). Analyse von Nutzerverhalten und Typenbildung*. Wiesbaden: Vieweg+Teubner Research.

Heller, V., & Morek, M. (2016). Gesprächsanalyse. Mikroanalytische Beschreibung sprachlicher Interaktion in Bildungs- und Lernzusammenhängen. In J. Boelmann (Hrsg.), *Empirische Erhebungs- und Auswertungsverfahren in der deutschdidaktischen Forschung* (S. 207-231). Baltmanssweiler: Schneider Hohengehren.

Herczeg, M. (2009). *Software-Ergonomie*. München: Oldenbourg Wissenschaftsverlag.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2009). Bootstrap methods and permutations tests. In D. S. Moore & G. P. McCabe (Hrsg.), *Introduction to the Practice of Statistics* (S. 16-11 - 16-60). New York: Freeman.

Höck, G. (2015). *Ko-Konstruktive Problemlösegespräche im Mathematikunterricht*. Münster: Waxmann.

Hofmann, T. (2012). *eFATHOM: Entwicklung und Evaluation einer multimedialen Lernumgebung für einen selbstständigen Einstieg in die Werkzeugsoftware FATHOM*. Wiesbaden: Vieweg+Teubner.

Holcomb, J., Chance, B., Rossman, A., & Cobb, G. (2010). *Assessing student learning about statistical inference*. Paper presented at the The Eigth International Conference on Teaching Statistics, Ljubljana, Slovenia.

Holcomb, J., Chance, B., Rossman, A., Tietjen, E., & Cobb, G. (2010). *Introducing concepts of statistical inference via randomization tests.* Paper presented at the The Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia.

Hoyles, C., & Noss, R. (2003). What can digital technologies take from and bring to research in mathematics education? In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. S. Leung (Hrsg.), *Second International Handbook of Mathematics Education* (S. 323-349). The Netherlands: Springer.

Initiative, C. C. S. S. (2012). *Common core state standards for mathematics*.

Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 331-355). Berlin, Heidelberg: Springer Spektrum.

Jungwirth, H. (2003). Interpretative Forschung in der Mathematikdidaktik - ein Überblick für Irrgäste, Teilzieher und Standvögel. *ZDM, 35*(5), 189-200.

Jungwirth, H. (2005). Interpretative Mathematikdiaktik: methodisches und methodologisches am Beispiel von Normen im Mathematikunterricht. Zugriff unter http://psydok.sulb.uni-saarland.de/volltexte/2005/449/pdf/jungwirth.pdf

Jungwirth, H., & Krummheuer, G. (2006). Banal sozial? Zur Sozialisierung des mathematischen Lehrens und Lernens durch die interpretative Unterrichtsforschung. In H. Jungwirth & G. Krummheuer (Hrsg.), *Der Blick nach innen: Aspekte der alltäglichen Lebenswelt Mathematikunterricht* (S. 7-18). Münster: Waxmann.

Justice, N., Zieffler, A., Huberty, M. D., & delMas, R. (2018). Every rose has its thorn: secondary teachers' reasoning about statistical models. *ZDM*. doi:10.1007/s11858-018-0953-1

Kieran, C., & Drijvers, P. (2006). The Co-Emergence of Machine Techniques, Paper-And-Pencil Techniques, and Theoretical Reflection: A Study of Use in Secondary School Algebra. *International Journal of Computers for Mathematical Learning, 11*, 205-263.

Knapp, H., & Fitzgerald, G. (1989). The antihypertensive effects of fish oil: A controlled study of polyunsaturated fatty acid supllements in essential hypertension. *New England Journal of Medicine, 320*(16), 1037-1043.

Konold, C. (1994). *Understanding probability and statistical inference through resampling*. Paper presented at the Proceedings of the First Scientific Meeting (of the IASE), Perugia, Italy.

Konold, C. (2007). Designing a Data Analysis Tool for Learners. In M. Lovett & P. Shah (Hrsg.), *Thinking with data: The 33rd annual Carnegie Symposium on cognition* (S. 267-291). Hillside, NJ: Lawrence Erlbaum Associates.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217-230.

Konold, C., & Kazak, S. (2008). Reconnecting Data and Chance. *Technology Innovations in Statistics Education, 2*(1), 1-39.

Konold, C., & Miller, C. (2017). TinkerPlots 2.3.3 [Deutsche Version]: Learn Troop.

Krauss, S., & Wassner, C. (2001). Wie man das Testen von Hypothesen einführen sollte. *Stochastik in der Schule, 21*(1), 29-34.

Krummheuer, G. (1997). *Narrativität und Lernen. Mikrosoziologische Studien zur sozialen Konstitution schulischen Lernens*. Weinheim: Deutscher Studien Verlag.

Krummheuer, G. (2003). Argumentationsanalyse in der mathematikdidaktischen Unterrichtsforschung. *ZDM, 35*(6), 247-256.

Krummheuer, G., & Brandt, B. (2001). *Paraphrase und Traduktion. Partizipationstheoretische Elemente einer Interaktionstheorie des Mathematiklernens in der Grundschule*. Weinheim, Basel: Beltz Wissenschaft.

Krummheuer, G., & Fetzer, M. (2005). *Der Alltag im Mathematikunterricht. Beobachten - Verstehen - Gestalten*. München: Elsevier.

Krummheuer, G., & Naujok, N. (1999). *Grundlagen und Beispiele Interpretativer Unterrichtsforschung*. Opladen: Leske+Budrich.

Kuckartz, U., Dresing, T., Rädiker, S., & Stefer, C. (2008). *Qualitative Evaluation. Der Einstieg in die Praxis*. Wiesbaden: VS-Verlag.

Kultusministerkonferenz. (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München: Wolters Kluwer.

Kultusministerkonferenz. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*. Köln: Wolters Kluwer.

Kurtzmann, G., & Sill, H.-D. (2012). Vorschläge zu Zielen und Inhalten stochastischer Bildung in der Primarstufe sowie in der Aus- und Fortbildung von Lehrkräften. In M. Ludwig (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 1005-1008). Münster: WTM Verlag.

Kütting, H., & Sauer, M. J. (2011). *Elementare Stochastik. Mathematische Grundlagen und didaktische Konzepte*. Berlin, Heidelberg: Springer.

Lamnek, S., & Krell, C. (2016). *Qualitative Sozialforschung*. Weinheim, Basel: Beltz.

Lehrer, R. (2017). Modeling signal-noise processes supports student construction of a hierarchical image of sample. *Statistics Education Research Journal, 16*(2), 64-85.

Leiss, D. (2007). *"Hilf mir, es selbst zu tun": Lehrerinterventionen beim mathematischen Modellieren*. Hildesheim: Franzbecker.

Lergenmüller, A., Schmidt, G., Krüger, K., Biehler, R., & Vehling, R. (2012). *Mathematik Neue Wege - Stochastik*. Braunschweig: Bildungshaus Schulbuchverlage.

Liu, Y., & Thompson, P. W. (2009). Mathematics Teachers' Understandings of Proto-Hypothesis Testing. *Pedagogies, 4*(2), 126-138.

Lock, R., Lock, P. F., Lock Morgan, K., Lock, E., & Lock, D. (2013). *Statistics. Unlocking the power of data*. Hoboken, New Jersey: Jon Wiley & Sons.

Lock, R., Lock, P. F., Lock Morgan, K., Lock, E., & Lock, D. (2014). *Intuitive introduction to the important ideas of inference*. Paper presented at the Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Ludbrock, J., & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician, 52*(2), 127-132.

Madden, S. (2008). *High school mathematics teacher's evolving knowledge of comparing distributions.* (Dissertation), Western Michigan University. Zugriff unter http://iase-web.org/documents/dissertations/08.Madden.Dissertation.pdf

Madden, S. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1-2), 109-131. doi:10.1080/10986065.2011.539078

Makar, K., & Confrey, J. (2002). *Comparing Two Distributions: Investigating Secondary Teachers' Statistical Thinking*. Paper presented at the Sixth International Conference on Teaching Statistics (ICOTS6), Cape Town, South Africa.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

Makar, K., & Rubin, A. (2017). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Hrsg.), *International Handbook of Research in Statistics Education* (S. 261-294). New York: Springer.

Marsaglia, G., & Zaman, A. (1991). A New Class of Random Number Generators. *Annals of Applied Probability, 1*(3), 426-480.

Maxara, C. (2006). *Einführung in die stochastische Simulation mit FATHOM*. Kasseler Online-Schriften zur Didaktik der Stochastik (KaDiSto), Band 1. Kassel: Universität Kassel.

Maxara, C. (2009). *Stochastische Simulation von Zufallsexperimenten mit Fathom: eine theoretische Werkzeuganalyse und explorative Fallstudie*. Hildesheim, Berlin: Franzbecker.

Maxara, C., & Biehler, R. (2006). *Students' probabilistic simulation and modeling competence after a computer-intensive elementary course in statistics and probability*. Paper presented at the The Seventh International Conference on Teaching Statistcs (ICOTS7), Salvador, Brazil.

Maxara, C., & Biehler, R. (2007). *Constructing stochastic simulations with a computer tool - students' competencies and difficulties*. Paper presented at the Cerme 5.

Mayring, P. (2010). *Qualitative Inhaltsanalyse. Gundlagen und Techniken*. Weinheim, Basel: Beltz.

Mayring, P. (2016). *Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken* (6., überarbeitete Aufl.). Weinheim: Beltz.

McNamara, A. A. (2015). *Bridging the Gap Between Tools for Learning and for Doing Statistics*. (Dissertation), Los Angeles, USA: University of California.

Meyer, M., & Prediger, S. (2009). Warum? Argumentieren, Begründen, Beweisen. *Praxis der Mathematik in der Schule, 51*(30), 1-7.

Meyfarth, T. (2006). *Ein computergestütztes Kurskonzept für den Stochastik-Leistungskurs mit kontinuierlicher Verwendung der Software Fathom - Didaktisch kommentierte Unterrichtsmaterialien*. Kasseler Online-Schriften zur Didaktik der Stochastik (KaDiSto), Band 2. Kassel: Universität Kassel.

Meyfarth, T. (2009). *Die Konzeption, Durchführung und Analyse eines simulationsintensiven Einstiegs in das Kurshalbjahr Stochastik der gymnasialen Oberstufe - Eine explorative Entwicklungsstudie*. Kasseler Online-Schriften zur Didaktik der Stochastik (KaDiSto), Band 6. Kassel: Universität Kassel.

Miller, M. (1986). *Kollektive Lernprozesse. Studien zur Grundlegung einer soziologischen Lerntheorie*. Frankfurt a. M.: Suhrkamp.

Miller, M. (2006). *Dissens. Zur Theorie diskursiven und systemischen Lernens*. Bielefeld: transcipt.

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *The Teachers College Record, 108*(6), 1017-1054.

Moore, D. S. (2007). *The basic practice of statistics* (4. Aufl.). New York: W. H. Freeman and Company.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston: NCTM.

Neubert, S., Reich, K., & Voß, R. (2001). Lernen als konstruktiver Prozess. In T. Hug (Hrsg.), *Die Wissenschaft und ihr Wissen, Bd. 1* (S. 253-265). Baltmannsweiler: Hohengehren.

Neuhäuser, M. (2010). *Computer-intensive und nichtparametrische statistische Tests*. München: Oldenbourg Verlag.

Neyman, J., & Pearson, E. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical of Physical Character, 231*, 289-337.

Noll, J., & Kirin, D. (2017). TinkerPlots model construction approaches for comparing two groups: Student perspectives. *Statistics Education Research Journal, 16*(2), 213-243.

Noss, R., & Hoyles, C. (1996). *Windows on Mathematical Meanings*. The Netherlands: Kluwer Academics Publisher.

Oesterhaus, J., & Biehler, R. (2013). BeSt@Kontext: Ein schüleraktivierendes Unterrichtskonzept für die Beurteilende Statistik mit computergestützter Simulation in authentischen Kontexten. In G. Greefrathe, F. Käpnick, & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht 2013* (S. 720-723). Münster: WTM-Verlag.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning, 13*(1-2), 27-46.

Pfannkuch, M., & Budgett, S. (2014). *Constructing inferential concepts through bootstrap and randomization-test simulations: A case study*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Pfannkuch, M., Budgett, S., & Arnold, P. (2015). Experiment-to-causation inference: Understanding causality in a probabilistic setting. In A. Zieffler & E. Fry (Hrsg.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (S. 95-128). Minneapolis, Minnesota: Catalyst Press.

Pfannkuch, M., Regan, M., Wild, C., Budgett, S., Forbes, S., Harraway, J., et al. (2011). Inference and the introductory statistics course. *International Journal of Mathematical Education in Science and Technology, 42*(7), 903-913.

Pfannkuch, M., Wild, C., & Regan, M. (2013). Students' difficulties in practicing computer-supported statistcal inference: Some hypothetical generalizations from a study. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen. Using Tools for Learning Mathematics and Statistics* (S. 393-403). Wiesbaden: Springer Spektrum.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. *Supplement to the Journal of the Royal Statistical Society, 4*(1), 119-130.

Podworny, S. (2007). *Hypothesentesten mit P-Werten im Stochastikunterricht der gymnasialen Oberstufe*. Examensarbeit, Kassel: Universität Kassel.

Podworny, S. (2013). Mit TinkerPlots vom einfachen Simulieren zum informellen Hypothesentesten. In G. Greefrath, F. Käpnick, & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht 2013* (S. 324-327). Münster: WTM Verlag.

Podworny, S. (2018). Students' Reflections About a Course for Learning Inferential Reasoning Via Simulations. In C. Batanero & E. J. Chernoff (Hrsg.), *Teaching and Learning Stochastics: Advances in Probability Education Research* (S. 333-349). Springer International Publishing.

Podworny, S., & Biehler, R. (2014). *A learning trajectory on hypothesis testing with TinkerPlots - design and exploratory evaluation*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Podworny, S., Frischemeier, D., & Biehler, R. (2017). Design, realization and evaluation of a statistics course for preservice teachers for primary school in Germany. In A. Molnar (Hrsg.), *Teaching Statistics in a Data Rich World. Proceedings of the Satellite Conference of the International Association of Statistics Education (IASE)*. Rabat, Morocco.

Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 3-4.

Pratt, D., & Ainley, J. (2013). Chance Re-encounters: 'Computers in Probability Education' revisted. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen. Using Tools for Learning Mathematics and Statistcis* (S. 165-177). Wiesbaden: Springer.

Prömmel, A. (2013). *Das GESIM-Konzept. Rekonstruktion von Schülerwissen beim Einstieg in die Stochastik mit Simulationen*. Wiesbaden: Springer Spektrum.

Rabardel, P. (1995). Les hommes et les technologies. Paris: Armand Colin.

Rabardel, P. (2002). *People and technology. A cognitive approach to contemporary instruments. Translated by Heidi Wood*: Université Paris 8.

Rabardel, P., & Samurcay, R. (2001). *From Artifact to Instrument-Mediated Learning*. University of Helsinki: unveröffentlichtes Manuskript.

Ramsey, F. L., & Shafer, D. W. (2013). *The Statistical Sleuth. A Course in Methods of Data Analysis*. Boston, Massachusetts: Cengage Learning.

Reichert, S. (2014). *Design, Durchführung und (beispielhafte) Auswertung einer Unterrichtsreihe zur Einführung in die computergestützte Simulation von Zufallsexperimenten mit TinkerPlots in der Sekundarstufe I*. Bachelorarbeit, Universität Paderborn.

Reimer, K. (2009). Bootstrapping und andere Resampling-Methoden. In S. Albers, D. Klapper, U. Konradt, A. Walter, & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (S. 521-536). Wiesbaden: Gabler Verlag.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529-556.

Ridgway, J. (2016). Implications of the Data Revolutions for Statistics Education. *Internatinal Statistical Review, 84*(3), 528-549. doi:10.1111/insr.12110

Riemer, W. (1991). Das 'Eins durch Wurzel aus n'-Gesetz. Einführung in statistisches Denken auf der Sekundarstufe I. *Stochastik in der Schule, 11*(3), 24-36.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.

Rossman, A., & Chance, B. (1999). Teaching the reasoning of statistical inference: A "top ten" list. *College Mathematics Journal, 30*(4), 297-305. Zugriff unter http://rossmanchance.com/papers/topten.html

Rossman, A., & Chance, B. (2014). Using simulation-based inference for learning introductory statistics. *WIREs Computation Statistics, 6*(4), 211-221. doi:10.1002/wics.1302

Rossman, A., Chance, B., Cobb, G., & Holcomb, R. (2008). *Concepts of statistical inference: Approach, scope, sequence and format for an elementary permutation-based first course*. Zugriff unter http://statweb.calpoly.edu/bchance/csi/CSIcurriculumMay08.doc

Roy, S., Rossman, A., Chance, B., Cobb, G., VanderStoep, J., Tintle, N., et al. (2014). *Using simulation/randomization to introduce p-value in week 1*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). *Exploring informal inference with interactive visualization software*. Paper presented at the The Sixth International Conference on Teaching Statistics (ICOTS6), Cape Town, South Africa.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking in Converstation. *Language, 50*(4), 696-735.

Saint-Mont, U. (2011). *Statistik im Forschungsprozess*. Berlin, Heidelberg: Springer.

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Significance, 1*(1), 15-20.

Schiffer, S. (1998). *Visuelle Programmierung: Grundlagen und Einsatzmöglichkeiten*. Bonn: Addison-Wesley.

Schiller, A., & Engel, J. (2018). *The importance of statistcal literacy for democracy - civic-education through statistics*. Paper presented at the Challenges and Innovations in Statistics Education Mulitplier Conference of ProCivicStat, Szeged, Hungary.

Schoenfeld, A. H. (1998). Making mathematics and making pasta: From cookbook procedures to really cooking. In J. G. Greeno & S. V. Goldman (Hrsg.), *Thinking practices in mathematics and science learning* (S. 299-319). Mahwah, NJ: Lawrence Erlbaum.

Schupp, H., Berg, G., Dabrock, D., & Pohlmann, D. (1992). *PROgramme für den STOchastik-Unterichtt. Ausbaufähiges Software-Paket zum stochastischen Experimentieren und Simulieren*. Bonn: Dümmler.

Schwarz, B. B., Hershkowitz, R., & Prusak, N. (2010). Argumentation and mathematics. In K. Littleton & C. Howe (Hrsg.), *Educational dialogues: Understanding and promoting productive interactions* (S. 115-141). Oxon: Routledge.

Schwarzkopf, R. (2000). *Argumentationsprozesse im Mathematikunterricht. Theoretische Grundlagen und Fallstudien*. Hildesheim: Franzbecker.

Schwarzkopf, R. (2001). Argumentationsanalysen im Unterricht der frühen Jahrgangsstufen — eigenständiges Schließen mit Ausnahmen. *Journal für Mathematik-Didaktik, 22*(3), 253-276. doi:10.1007/bf03338938

Sedlmeier, P. (1999). *Improving Statistical Reasoning. Theoretical Models and Practical Implications*. Mahwah, New Jersey: Lawrence Erlbaumm Associates.

Shaughnessy, M. (2007). Research on Statistics Learning and Reasoning. In F. K. Lester (Hrsg.), *Second Handbook of Research on Mathematics Teaching and Learning* (S. 957-1009). Charlotte, North Carolina: Information Age Publishing.

Sill, H.-D. (2018). Zur Stochastikausbildung im Primarstufenlehramt. In R. Möller & R. Vogel (Hrsg.), *Innovative Konzepte für die Grundschullehrerausbildung im Fach Mathematik* (S. 71-93). Wiesbaden: Springer.

Specker, A. (1997). *Kognitives Software Engineering: ein schema-und scriptbasierter Ansatz*. Zürich: ETH Zürich.

Sproesser, U. (2015). *Einflussfaktoren auf Kompetenz, Sichtweisen und motivationale Variablen im Bereich von Statistical Literacy*. Dissertation, Ludwigsburg: Pädagogische Hochschule Ludwigsburg.

Stangl, W. (2018). Stichwort: "Hathorne-Effekt". Online Lexikon für Psychologie und Pädagogik. Zugriff unter http://lexikon.stangl.eu/1965/hawthorne-effekt/

Steinke, I. (2013). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardoff, & I. Steinke (Hrsg.), *Qualitative Forschung* (S. 319-331). Reinbek bei Hamburg: Rowohlt.

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires post-traning sleep. *Nature Neuroscience, 2*, 1237-1238.

Stratmann, J., Preussler, A., & Kerres, M. (2009). Lernerfolg und Kompetenz: Didaktische Potenziale der Portfolio-Methode im Hochschulstudium. *Zeitschrift für Hochschulentwicklung, 4*(1), 90-103.

Swanson, T., Vanderstoep, J., & Tintle, N. (2014). *Student attitudes toward statistics from a randomization-based curriculum*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-296.

The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32*(1), 5-8. Zugriff unter http://www.jstor.org/stable/3699927

Thompson, P. W., Liu, Y., & Saldanha, L. (2012). Intricacies of Statistical Inference and Teachers' Understandings of Them. In M. C. Lovett & P. Shah (Hrsg.), *Thinking with Data* (S. 207-231). New York: Psychology Press.

Tietze, U.-P., Klika, M., & Wolpers, H. (2002). *Mathematikunterricht in der Sekundarstufe II. Band 3. Didaktik der Stochastik*. Braunschweig: Vieweg.

Tintle, N., Rogers, A., Chance, B., Cobb, G., Rossman, A., Roy, S., et al. (2014). *Quantitative evidence for the use simulation and randomization in the introductory statistics course*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal, 11*(1), 21-40.

Tintle, N., VanderStoep, J., Holmes, V.-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education, 19*(1), 1-25.

Tintle, N., VanderStoep, J., & Swanson, T. (2009). *An active approach to statistical inference, preliminary edition*. Holland, Michigan: Hope College Publishing.

Toulmin, S. E. (1975). *Der Gebrauch von Argumenten*. Kronberg: Scriptor.

Trouche, L. (2004). Managing the complexity of human/machine interactions in computerized learning environments: Guiding students' command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning, 9*(3), 281-307.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute, 58*, 201-204.

Voigt, J. (1984). *Interaktionsmuster und Routinen im Mathematikunterricht: theoret. Grundlagen und mikroethnograph. Falluntersuchungen*. Weinheim: Beltz.

Voigt, J. (1991). Die mikroethnographische Erkundung von Mathematikunterricht - Interpretative Methoden der Interationsanalyse. In H. Maier & J. Voigt (Hrsg.), *Interpretative Unterrichtsforschung: Heinrich Bauersfeld zum 65. Geburtstag*. Köln: Aulis Verlag.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779-804.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1-8.

Wassner, C. (2004). *Förderung Bayesianischen Denkens: Kognitionspsychologische Grundlagen und didaktische Analysen; mit Arbeitsmaterialien und didaktischen Kommentaren zum Thema "Authentisches Bewerten und Urteilen unter Unsicherheit" für den Stochastikunterricht der Sekundarstufe I*. Hildesheim: Franzbecker.

Wassner, C., Biehler, R., Schweynoch, S., & Martignon, L. (2007). *Authentisches Bewerten und Urteilen unter Unsicherheit - Arbeitsmaterialien und didaktische Kommentare für den Themenbereich "Bayessche Regel" für den Stochastikunterricht der Sekundarstufe I*. Kasseler Online-Schriften zur Didaktik der Stochastik (KaDiSto), Band 5. Kassel: Universität Kassel.

Wassong, T., & Biehler, R. (2010). *A Model for Teacher Knowledge as a Basis for Online Courses for Professional Development of Statistics Teacher*. Paper presented at the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia.

Watson, J., & Fitzallen, N. (2016). Statistical software and mathematics education. In L. D. English (Hrsg.), *Handbook of international research in mathematics education* (Bd. 3, S. 563-594). New York: Routledge.

Watson, J., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*(2), 145-168.

West, W. (2014). *Teaching resampling in an introductory statistics course*. Paper presented at the The Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, USA.

Wickmann, D. (1990). *Bayes-Statistik: Einsicht gewinnen und entscheiden bei Unsicherheit*. Mannheim: BI-Wissenschaftsverlag.

Wild, C., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223-265.

Wild, C., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 247-295.

Wild, C., Pfannkuch, M., Regan, M., & Parsonage, R. (2015). Accessible conceptions of statistical inference: Pulling ourselves up by the bootstraps. *Internatinal Statistical Review, 85*(1), 84-107. doi:10.1111/insr.12117

Wittmann, E. C. (1985). Objekte - Operationen - Wirkungen: Das operative Prinzip in der Mathematikdidaktik. *Mathematik lehren, 11*, 7-11.

Wygotski, L. S. (1985). Die instrumentelle Methode in der Psychologie. In J. Lompscher (Hrsg.), *Lew Wygotski. Ausgewählte Schriften* (Bd. 1, S. 309-317). Köln: Pahl-Rugenstein.

Zieffler, A., & Catalysts for Change. (2013). *Statistical Thinking. A simulation approach to modeling uncertainty* (3. edition Aufl.). Minneapolis, Minnesota: Catalyst Press.

Zieffler, A., Garfield, J., DelMas, R., & Reading, C. (2008). A Framework to Support Research on Informal Inferential Reasoning. *Statistics Education Research Journal*, *7*(2), 40-85.

Zieffler, A., Harring, J. R., & Long, J. D. (2011). *Comparing groups. Randomization and bootstrap methods using R*. Hoboken, New Jersy: John Wiley & Sons.