Stochastik

Barth, Friedrich

München, [20]03

17. 4. 3. Die Operationscharakteristik eines Tests

urn:nbn:de:hbz:466:1-83580

Beim üblichen Signifikanzniveau 5% können wir die Nullhypothese »Lady X. rät blind« nicht ablehnen. Ist man jedoch mit einem Signifikanzniveau von 5,5% oder höher zufrieden, so kann man die Nullhypothese »Lady X. rät blind« ablehnen und der Dame Begabung bescheinigen. Die statistische Sicherheit unseres Urteils »Lady X. ist begabt« beträgt dann höchstens 94,5%. Was heißt das? Wenn viele Ladies sich unserer Prüfung unterzögen, attestierten wir ca. 5,5% dieser Damen fälschlicherweise eine gewisse Begabung, weil sie 8 oder mehr Tassen richtig benennen, obwohl sie blind raten.
Was ist aber mit den begabten Damen? Dieser Frage wollen wir im nächsten Abschnitt nachgehen.

17.4.3. Die Operationscharakteristik eines Tests

Beispiel 3: Dem Teetassentest aus Beispiel 2 stellt sich eine Lady, die tatsächlich über eine gewisse Begabung verfügt und mit der Wahrscheinlichkeit $p = 0{,}6$ die Tassen richtig benennt. Mit welcher Wahrscheinlichkeit wird man ihre Begabung verkennen, wenn wir wie in Beispiel 2 als kritischen Bereich die Menge $K = [8; 10]$ nehmen?
Die Wahrscheinlichkeit β', einen solchen Fehler 2. Art zu begehen, ergibt sich zu

$$\beta' = P_{0,6}^{10}(Z \in \bar{K}) = P_{0,6}^{10}(Z \leqq 7) = F_{0,6}^{10}(7) \approx 83{,}3\%.$$

Solchen schwach begabten Damen wird mit unserem Test also oft unrecht getan! Wäre die Begabung der Dame größer, z.B. $p = 0{,}9$, so würden wir sie auch besser erkennen; es ergäbe sich nämlich $\beta' = F_{0,9}^{10}(7) \approx 7{,}0\%$. Weil wir aber über die Begabung der Damen, die sich dem Test unterziehen, nichts wissen, müssen wir uns einen Überblick über alle Wahrscheinlichkeiten für einen Fehler 2. Art verschaffen. Da diese Wahrscheinlichkeiten offensichtlich von p abhängen, betrachten wir die Funktion

$$\beta': p \mapsto P_p^{10}(Z \in \bar{K}), \ D_{\beta'} =]\tfrac{1}{2}; 1].$$

Mit Hilfe einer Wertetabelle können wir den Graphen dieser Funktion zeichnen (Tabelle 353.1 und Figur 353.1).

Man erkennt, daß die Wahrscheinlichkeit β' für einen Fehler 2. Art um so größer wird, je weniger sich die Begabung vom blinden Raten $(p = \frac{1}{2})$ unterscheidet. Da die Definitionsmenge $D_{\beta'}$ links offen ist, gibt es keine größte Irrtumswahrscheinlichkeit 2. Art. Als Ersatz dafür nimmt man das Supremum aller Irrtumswahrscheinlichkeiten 2. Art, also den Wert $1 - \alpha'$. Er ist in unserem Fall etwa 94,5%. Man riskiert also, mit einer Wahrscheinlichkeit bis zu 94,5% begabte – wenn auch sehr schwach begabte – Damen zu Unrecht für unbegabt zu halten. Wir können trotzdem zufrieden sein: Der unangenehme Fall, daß eine Dame nur flunkert und wir ihr dennoch hohe Sensibilität bescheinigen, tritt nur mit 5,5% Wahrscheinlichkeit ein. Daß wir andererseits u.U. einer wirklich begabten Dame ein Unrecht antun, nehmen wir in Kauf in der Gewißheit, daß sich das Genie so oder so eines Tages durchsetzen wird.

p	$\beta' = P_p^{10}(Z \leqq 7)$
0,51	0,94
55	90
60	83
65	74
70	62
75	47
80	32
85	18
90	07
95	01
99	0001
1	0

Tab. 353.1 Wahrscheinlichkeit β' für einen Fehler 2. Art beim kritischen Bereich $K = [8; 10]$

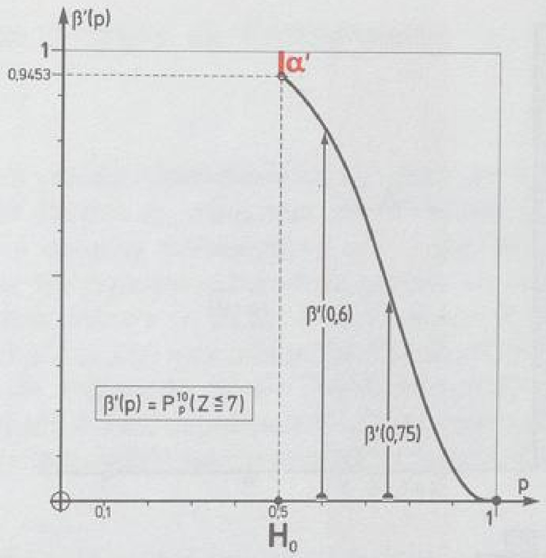

Fig. 353.1 Graph der Funktion $\beta': p \mapsto P_p^{10}(Z \in \overline{K})$

Setzt sich die Gegenhypothese nur aus endlich vielen einfachen Hypothesen zusammen wie bei Theodors Urne in Beispiel 1 von Seite 346, dann besteht der Graph von β' nur aus diskreten Punkten, so wie ihn Figur 347.1 zeigt. In einem solchen Fall gibt es natürlich eine größte Irrtumswahrscheinlichkeit 2. Art.
Es hat sich in der Statistik eingebürgert, die auf der Gegenhypothese H_1 definierte Funktion $p \mapsto \beta'(p)$ auf die Menge *aller* beim Test betrachteten Hypothesen, d. h. auf die zulässige Hypothese $H := H_0 \cup H_1$ fortzusetzen. Diese Funktion heißt dann Operationscharakteristik des Tests, kurz OC des Tests.

Definition 353.1: Es sei auf dem Ergebnisraum Ω der Testgröße Z eine Menge von Wahrscheinlichkeitsverteilungen als zulässige Hypothese H gegeben. Diese Verteilungen lassen sich durch einen Parameter p kennzeichnen. $A \subset \Omega$ sei ein Ereignis. Dann heißt die Funktion

$$\text{OC}: p \mapsto P_p(A), \; D_{\text{OC}} = H$$

die **Operationscharakteristik des Ereignisses A bezüglich H.** Ihr Graph heißt **OC-Kurve.** *

Bemerkung: Der Parameter p muß nicht unbedingt eine Wahrscheinlichkeit sein. So werden z. B. *Poisson*-Verteilungen durch den Parameter »Erwartungswert μ«, Normalverteilungen durch die Parameter μ und σ^2 gekennzeichnet.
Figur 354.1 veranschaulicht am Beispiel des Ereignisses $A := [4; 7]$ und an der Schar $B(16; p)$, $p \in [0; 1]$, als zulässiger Hypothese das Zustandekommen der

* In der Literatur verwendet man vielfach noch die ursprünglich von *Jerzy Neyman* und *E. S. Pearson* zur Kennzeichnung der Güte oder Macht eines Tests eingeführte *power function* = **Gütefunktion g**. Für sie gilt $g(p) := 1 - \text{OC}(p)$.

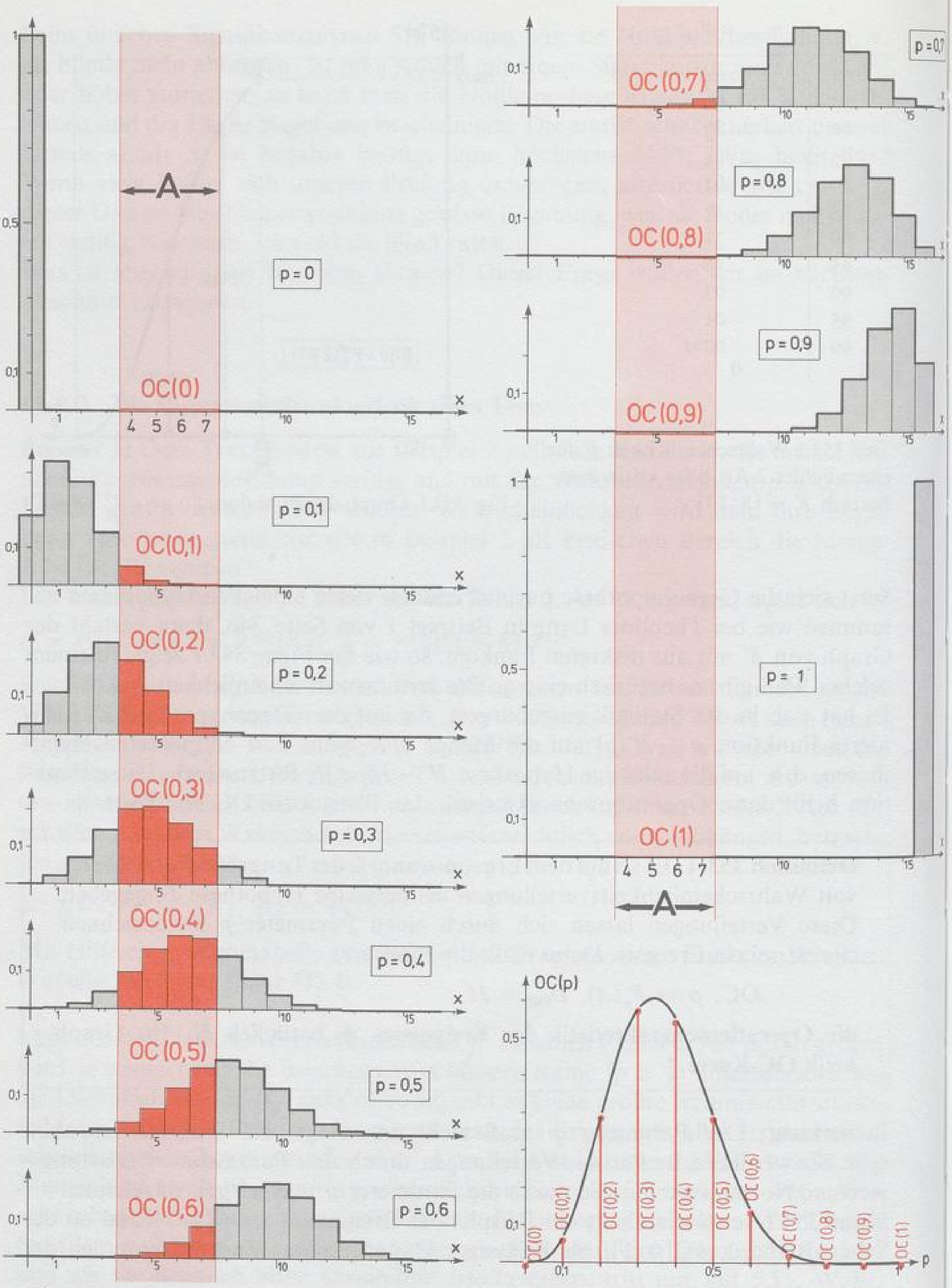

Fig. 354.1 Veranschaulichung der Entstehung der OC des Ereignisses $A := [4; 7]$ bezüglich der zulässigen Hypothese $H := \{B(16; p) \mid p \in [0; 1]\}$. Bedeutet f_p die Dichtefunktion der Binomialverteilung $B(16; p)$, so läßt sich die Operationscharakteristik mittels eines Integrals schreiben, nämlich OC: $p \mapsto \int\limits_{3,5}^{7,5} f_p(t)\,\mathrm{d}t$.

Operationscharakteristik: Zu jedem p gehört als Funktionswert

$$\mathrm{OC}(p) = P_p^{16}(Z \in A) = \sum_{i=4}^{7} \mathrm{B}(16; p; i).$$

Bei einem Signifikanztest spricht man von der Operationscharakteristik der Entscheidungsregel δ mit dem kritischen Bereich K, wenn man $A = \bar{K}$ wählt. Ihre Funktionswerte $\mathrm{OC}(p) = P_p(\bar{K})$ sind dann in Abhängigkeit von p die Wahrscheinlichkeiten, mit denen man die Nullhypothese beibehält, gleich, ob diese Entscheidung die richtige ist oder nicht. Für $p \in H_1$ ist der Funktionswert $P_p(\bar{K})$ jeweils die Irrtumswahrscheinlichkeit 2. Art, daß man nämlich die Nullhypothese nicht ablehnt, obwohl sie nicht zutrifft. Für $p \in H_0$ ist der Funktionswert $P_p(\bar{K})$ jeweils gleich der Sicherheit $1 - \alpha'(p)$, mit der die zutreffende Nullhypothese nicht abgelehnt wird. Dabei ist $\alpha'(p)$ die zu $p \in H_0$ gehörende Irrtumswahrscheinlichkeit 1. Art.

Übrigens kann auch die Nullhypothese H_0 selbst zusammengesetzt sein. Nehmen wir etwa im Teetassentest von Beispiel 2 (**17.4.2.**) als zulässige Hypothese $H := [0; 1]$ und als Nullhypothese $H_0 := [0; \frac{1}{2}]$, dann ergäbe sich als Operationscharakteristik des Ereignisses »$Z \leqq 7$« die Funktion $\mathrm{OC}\colon p \mapsto F_p^{10}(7)$, $D_{\mathrm{OC}} = [0; 1]$, deren Graph Figur 355.1 wiedergibt. Nun gibt es auch unendlich viele Irrtumswahrscheinlichkeiten 1. Art. Zur Charakterisierung des Tests genügt es offenbar, die größte dieser Wahrscheinlichkeiten anzugeben.
Je nach Lage des kritischen Bereichs K haben die Graphen der Operationscharakteristik, kurz OC-Kurven genannt, eine typische Gestalt. Nehmen wir als zulässige Hypothese die Menge aller Binomialverteilungen $\mathrm{B}(n; p)$ mit $p \in [0; 1]$, so gibt es 4 besonders wichtige Typen. Der Nachweis der aufgeführten Eigenschaften wird Aufgabe 372/**48** vorbehalten.

1) $K := [0; k] \Rightarrow \mathrm{OC}\colon p \mapsto 1 - F_p^n(k)$
Ist K linksbündig, so ist die OC-Kurve echt monoton steigend.

2) $K := [k; n] \Rightarrow \mathrm{OC}\colon p \mapsto F_p^n(k-1)$
Ist K rechtsbündig, so ist die OC-Kurve echt monoton fallend.

3) $K := [0; k_1] \cup [k_2; n] \Rightarrow$
$\mathrm{OC}\colon p \mapsto F_p^n(k_2 - 1) - F_p^n(k_1)$
Ist K getrennt, so hat die OC-Kurve einen inneren Hochpunkt.

4) $K := [k_1; k_2] \Rightarrow$
$\mathrm{OC}\colon p \mapsto F_p^n(k_1 - 1) + 1 - F_p^n(k_2)$
Ist K ein inneres Intervall, so hat die OC-Kurve einen inneren Tiefpunkt.

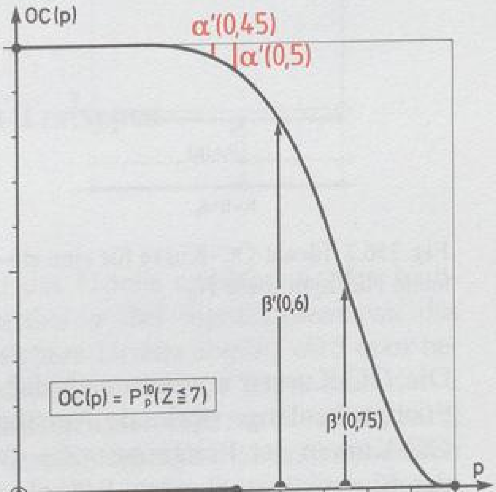

Fig. 355.1 Operationscharakteristik des Ereignisses »$Z \leq 7$« bezüglich $H = [0; 1]$. Vgl. Fig. 353.1

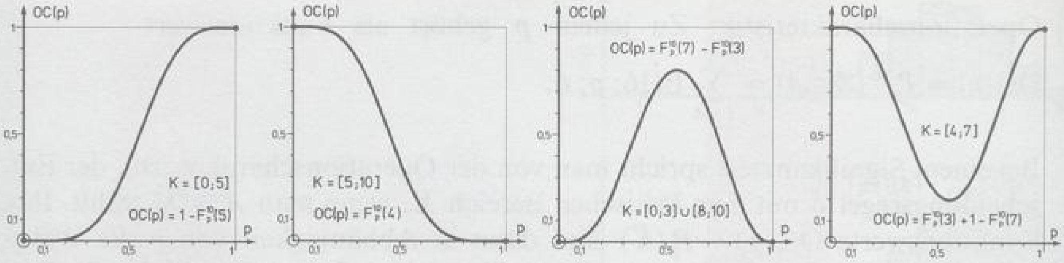

Fig. 356.1 Die 4 wichtigen Typen von OC-Kurven bezüglich $H = \{B(n; p) \mid p \in [0; 1]\}$, veranschaulicht mittels Binomialverteilungen $B(10; p)$

Wie man sich leicht überlegt, sind diese 4 Operationscharakteristiken Polynome n-ten Grades in p. Figur 356.1 veranschaulicht sie für $n = 10$.

Die OC-Kurve gibt uns einen Hinweis auf die Güte des Tests. Je steiler sie nämlich in ihren Flanken ist, desto schneller werden die Irrtumswahrscheinlichkeiten 2. Art klein. Im Idealfall wären für jedes $p \in H_0$ die Irrtumswahrscheinlichkeit $\alpha'(p) = 0$ und für jedes $p \in H_1$ die Irrtumswahrscheinlichkeit $\beta'(p) = 0$. Dann würde man nur richtige Urteile abgeben! Die zugehörige OC-Kurve hätte über H_0 konstant den Wert 1 und über H_1 konstant den Wert 0. Figur 356.2 zeigt die ideale OC-Kurve für eine einfache Nullhypothese, Figur 356.3 für eine zusammengesetzte Nullhypothese.

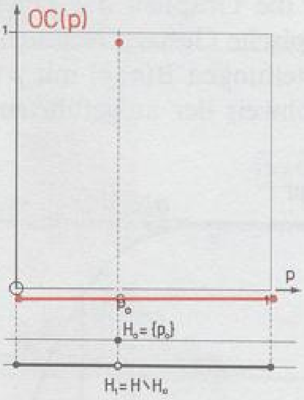

Fig. 356.2 Ideale OC-Kurve für eine einfache Nullhypothese H_0

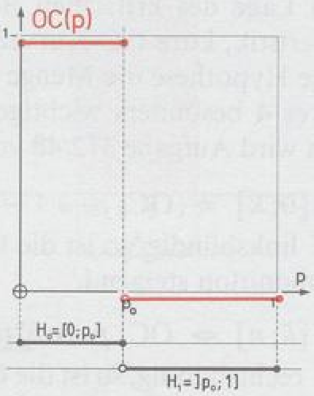

Fig. 356.3 Ideale OC-Kurve für eine zusammengesetzte Nullhypothese H_0

Die OC-Kurven erweisen sich daher als praktisches Hilfsmittel, bei gegebener Stichprobenlänge optimale Annahmebereiche zu finden. Figur 357.1 zeigt die OC-Kurven der Ereignisse »$Z = 0$«, »$Z \leqq 1$«, …, »$Z \leqq 5$« bezüglich der Schar der Binomialverteilungen $B(5; p)$, $p \in [0; 1]$, als zulässiger Hypothese H. Man entnimmt ihr z. B., daß man für die Entscheidung zwischen den Hypothesen $H_0 = \{0{,}15\}$ und $H_1 = \{0{,}4\}$ am besten das Ereignis »$Z \leqq 2$« heranzieht, wenn die Wahrscheinlichkeit für einen Fehler 1. Art unter 5% liegen soll. Ohne diese Bedingung würde man sich für »$Z \leqq 1$« entscheiden, weil dann $\alpha' + \beta'$ minimal

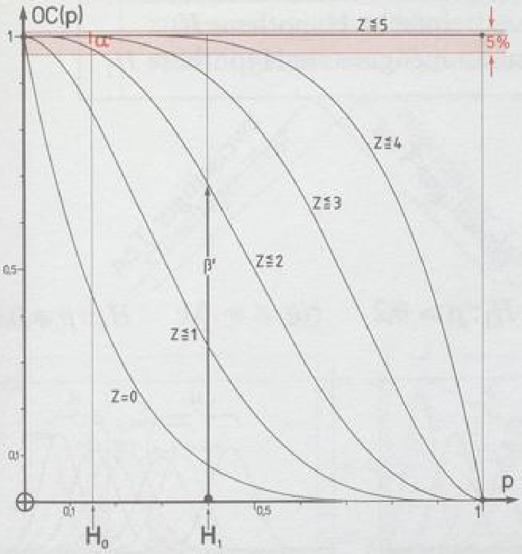

Fig. 357.1 Alternativtest für $H_0 = \{0{,}15\}$, $H_1 = \{0{,}4\}$ und $A = [0; k]$ mit $k \in \{0, 1, 2, 3, 4, 5\}$. Auswahl des optimalen Tests für die Schranke $\alpha = 5\%$

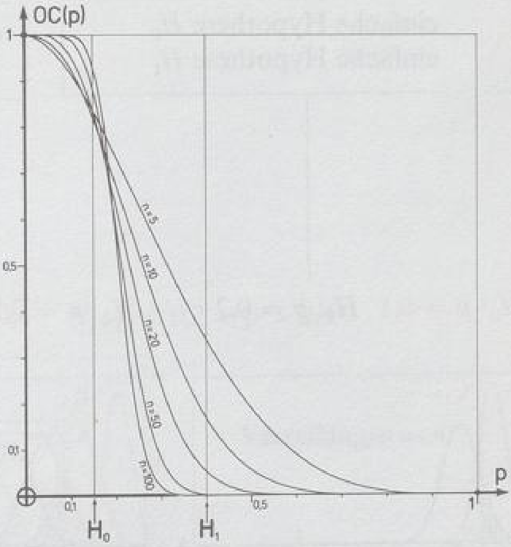

Fig. 357.2 Illustration des Einflusses der Stichprobenlänge n auf die Trennschärfe $H_0 = \{0{,}15\}$; $H_1 = \{0{,}4\}$; $A = [0; 0{,}2n]$; $n \in \{5; 10; 20; 50; 100\}$.

wird. Ein Ereignis ist desto besser für eine Entscheidungsregel geeignet, je stärker die OC-Kurve von dem einen der beiden in Frage kommenden p-Werte bis zum anderen abfällt. Andererseits läßt sich der Einfluß der Stichprobenlänge n auf die **Trennschärfe** des Tests an Hand der zugehörigen OC-Kurven beobachten (Figur 357.2). Wie erwartet fallen die OC-Kurven für größere n steiler von 1 auf 0 ab und trennen daher die Hypothesen besser. Für $n \to \infty$ hätte man einen idealen Test mit senkrecht abfallender OC-Kurve. Die Trennung ist perfekt, die Fehler haben die Wahrscheinlichkeit 0.

17.5. Überblick über die behandelten Testtypen

Siehe Seite 358f.

17.6. Verfälschte Tests

Bei einem Signifikanztest hat die Sicherheit des Urteils »Ablehnung der Nullhypothese« mindestens den Wert $1 - \alpha$, wobei α das Signifikanzniveau des Tests ist. Da man natürlich gern möglichst sichere Urteile abgibt, wird man bestrebt sein, das Signifikanzniveau α möglichst klein zu halten. Wählt man nun α und damit auch den kritischen Bereich K sehr klein, dann muß man leider in Kauf nehmen, daß nur noch in seltenen Fällen die Nullhypothese abgelehnt werden kann; d. h., der Test wird sehr häufig kein brauchbares Ergebnis liefern. Dieser Sachverhalt könnte einen Tester nun in die Versuchung bringen, erst einmal den Ausfall der Stichprobe abzuwarten und dann den kritischen Bereich K möglichst eng um das Stichprobenergebnis herumzulegen und damit das Signifikanzniveau recht klein zu machen. Der Versuchsausgang erschiene dann in einem besonders