Stochastik

Barth, Friedrich

München, [20]03

17. 6. Verfälschte Tests

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

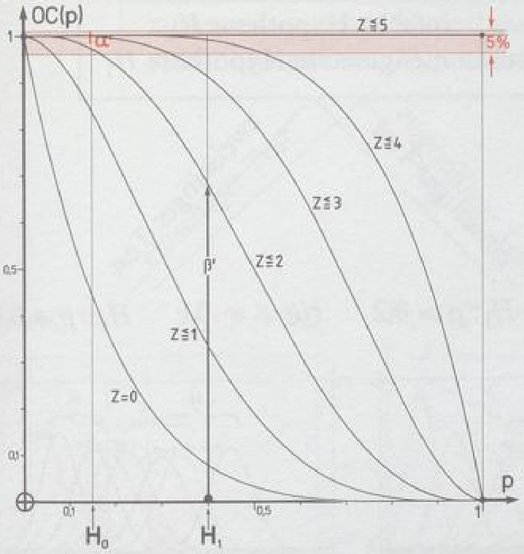


Fig. 357.1 Alternativtest für $H_0 = \{0{,}15\}$, $H_1 = \{0{,}4\}$ und $A = [0; k]$ mit $k \in \{0, 1, 2, 3, 4, 5\}$. Auswahl des optimalen Tests für die Schranke $\alpha = 5\%$

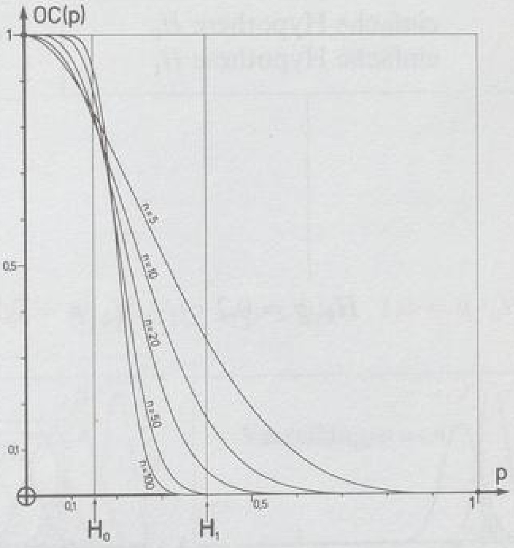


Fig. 357.2 Illustration des Einflusses der Stichprobenlänge n auf die Trennschärfe $H_0 = \{0{,}15\}$; $H_1 = \{0{,}4\}$; $A = [0; 0{,}2n]$; $n \in \{5; 10; 20; 50; 100\}$.

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

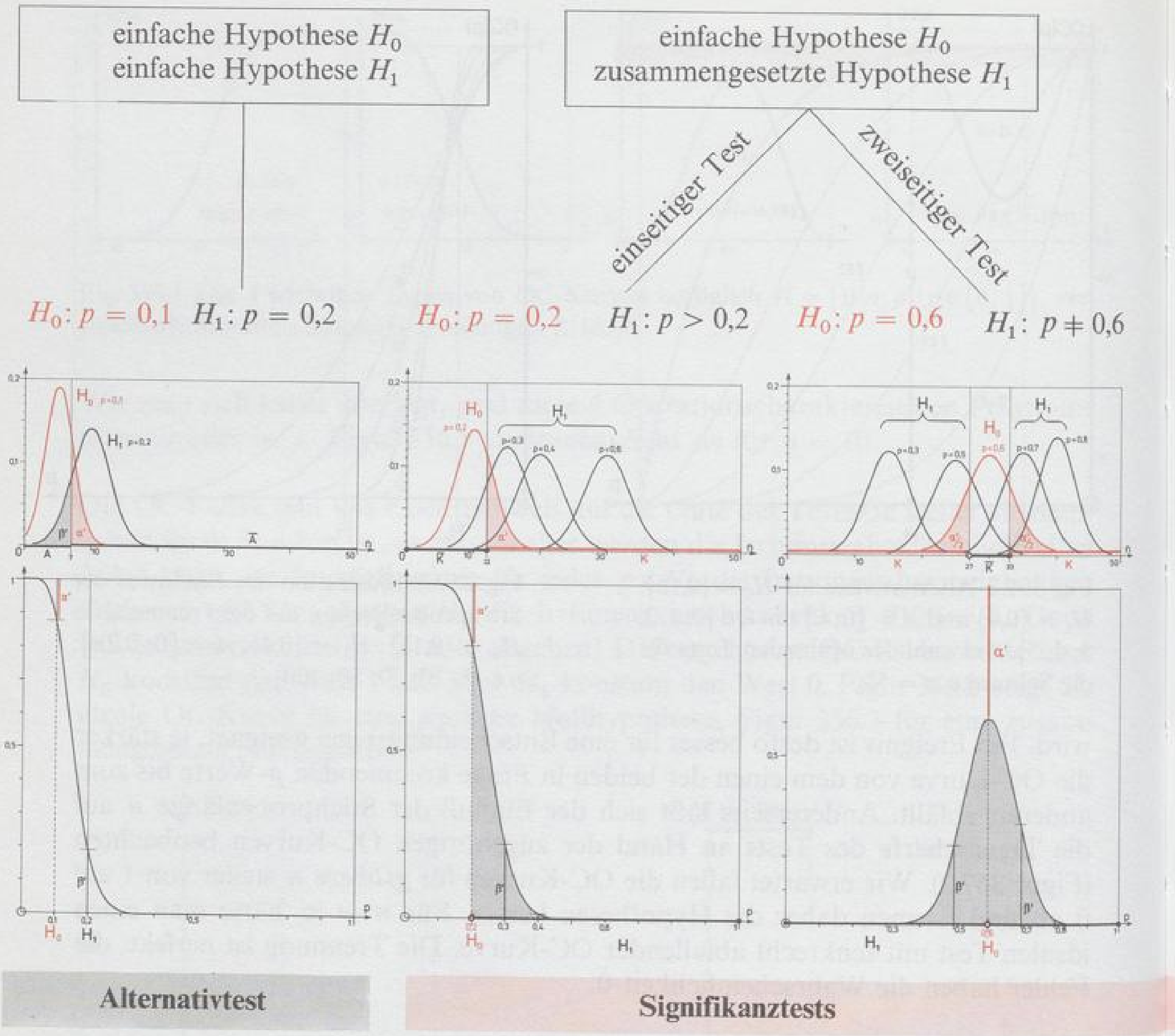
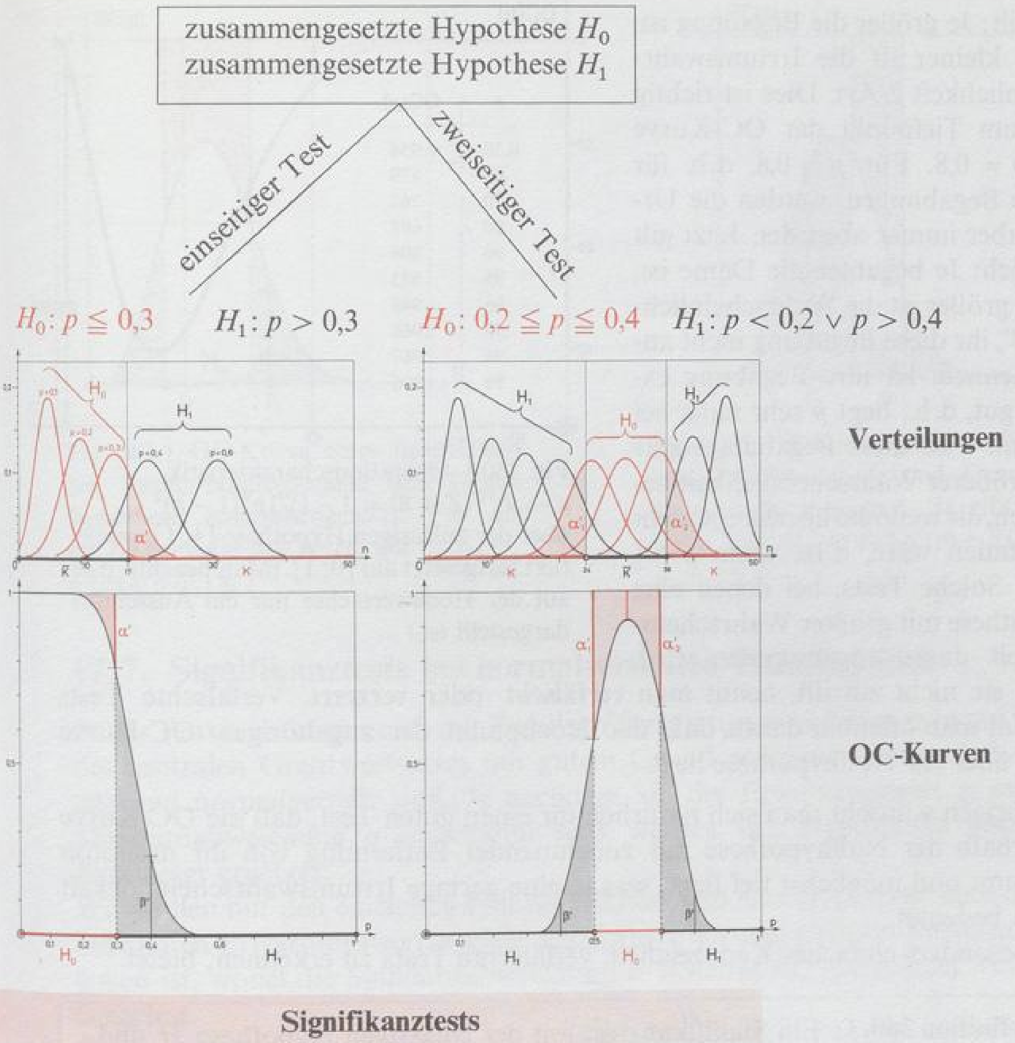
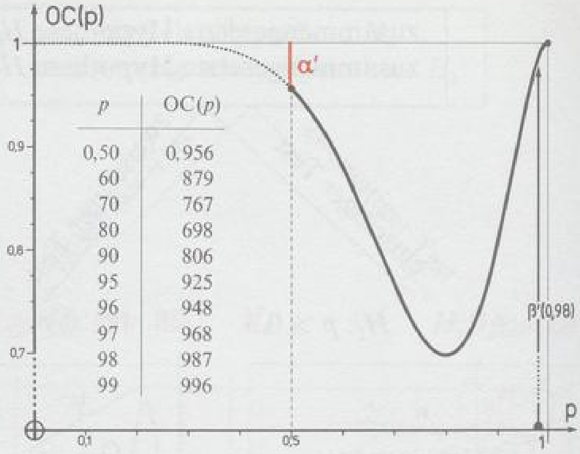


Fig. 358.1 Überblick über die behandelten Testtypen

günstigen Licht! Allerdings ist die Auswahl des Tests nach dem Ausgang der zufallsbedingten Stichprobe sehr gefährlich. Wählt man nämlich den Test nach Bedarf, also letzten Endes zufallsbestimmt, so verlieren die errechneten Wahrscheinlichkeiten jeden Sinn, und man kann sie dann auch nicht mehr als Beleg für irgendwelche Behauptungen anführen. Bei der Vielzahl verschiedener Tests, die man in einem Handbuch der mathematischen Statistik finden kann, ist natürlich die Versuchung groß, sich einen solchen auszusuchen, der irgendeine erwünschte Aussage am besten »bestätigt«. Vor einem derartigen Mißbrauch der Statistik muß daher ganz besonders gewarnt werden. Wir betrachten ein abschreckendes



Beispiel: Im Teetassentest (Seite 350) möchte Sir Y., ein leidenschaftlicher Verehrer von Lady X., ihr mit möglichst großer Sicherheit Begabung bescheinigen. Weil er weiß, daß Lady X. 8 Tassen richtig benannt hat, nimmt er einen kleinstmöglichen kritischen Bereich, bei dem er ihr noch Begabung attestieren kann, also $K = \{8\}$. Für den Fehler 1. Art ergibt sich die Wahrscheinlichkeit $\alpha' = P_{0,5}^{10}(Z = 8) \approx 4{,}4\%$. Sir Y. wird erläutern, daß nach seinem sehr scharfen Test nur in 4,4% der Fälle, in denen in Wirklichkeit $p = \frac{1}{2}$ vorliegt, auf »$p > \frac{1}{2}$« erkannt wird. Damit hat er zweifellos recht. Aber sehen wir uns die OC-Kurve des Ereignisses $\bar{K}$ an (Figur 360.1). Sie zeigt, daß für kleine Begabungen (d.h. p wenig größer als $\frac{1}{2}$) der Test durchaus brauchbar ist. Denn für solche Begabungen gilt: Je größer die Begabung ist, desto kleiner ist die Irrtumswahrscheinlichkeit 2. Art. Dies ist richtig bis zum Tiefpunkt der OC-Kurve bei $p = 0{,}8$. Für $p > 0{,}8$, d.h. für große Begabungen, werden die Urteile aber immer absurder. Jetzt gilt nämlich: Je begabter die Dame ist, desto größer ist die Wahrscheinlichkeit β', ihr diese Begabung nicht anzuerkennen. Ist ihre Begabung extrem gut, d.h., liegt p sehr nahe bei 1, dann wird diese Begabung sogar mit größerer Wahrscheinlichkeit bestritten, als wenn sie überhaupt nicht vorhanden wäre, d.h., wenn $p = \frac{1}{2}$ wäre. Solche Tests, bei denen eine Hypothese mit größter Wahrscheinlichkeit dann angenommen wird, wenn sie nicht zutrifft, nennt man **verfälscht** oder **verzerrt**. Verfälschte Tests erkennt man offenbar daran, daß der Hochpunkt der zugehörigen OC-Kurve nicht über der Nullhypothese liegt.



p	OC(p)
0,50	0,956
60	879
70	767
80	698
90	806
95	925
96	948
97	968
98	987
99	996

Fig. 360.1 Operationscharakteristik $p \mapsto P_p^{10}(Z \neq 8) = 1 - \binom{10}{8} p^8 (1-p)^2$ über der zulässigen Hypothese $[\frac{1}{2}; 1]$, punktiert fortgesetzt auf $[0; 1]$. (Man beachte, daß auf der Hochwertachse nur ein Ausschnitt dargestellt ist.)

Im übrigen wünscht man sich natürlich für einen guten Test, daß die OC-Kurve außerhalb der Nullhypothese mit zunehmender Entfernung von ihr monoton abnimmt und möglichst tief liegt, was ja eine geringe Irrtumswahrscheinlichkeit 2. Art bedeutet.

Ein besonders einfaches Kennzeichen, verfälschte Tests zu erkennen, bietet

Definition 360.1: Ein Signifikanztest mit der zulässigen Hypothese H und der Nullhypothese H_0 heißt **unverfälscht**, wenn für jedes $p \in H_0$ und jedes $p_1 \in H \setminus H_0$

$$\alpha'(p) + \beta'(p_1) \leqq 1$$

gilt. Andernfalls heißt der Test **verfälscht**.

In dieser Definition kommt zum Ausdruck, daß bei einem unverfälschten Test das Maximum der Operationscharakteristik über der Nullhypothese H_0 angenommen werden muß. Die Figuren 361.1 und 361.2 veranschaulichen Definition 360.1.

Was bedeutet eigentlich die Bedingung $\alpha'(p) + \beta'(p_1) > 1$, die einen verfälschten Test kennzeichnet? Sie besagt, daß

$$\alpha'(p) > 1 - \beta'(p_1),$$

d.h.: Die Wahrscheinlichkeit, die Nullhypothese abzulehnen, falls sie zutrifft, ist größer als die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie nicht zutrifft.

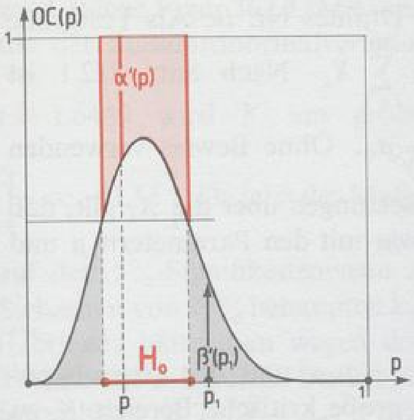


Fig. 361.1 OC-Kurve eines unverfälschten Tests. Nullhypothese und Gegenhypothese zusammengesetzt. Für alle $p \in H_0$ und alle $p_1 \in H \backslash H_0$ gilt: $\alpha'(p) + \beta'(p_1) \leqq 1$.

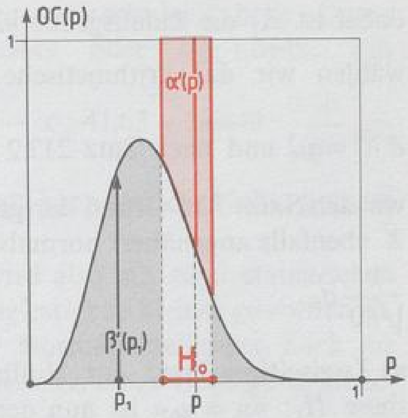


Fig. 361.2 OC-Kurve eines verfälschten Tests. Nullhypothese und Gegenhypothese zusammengesetzt. Es gibt $p \in H_0$ und $p_1 \in H \backslash H_0$, so daß $\alpha'(p) + \beta'(p_1) > 1$.

17.7. Signifikanztests bei normalverteilten Zufallsgrößen

In der Praxis hat man es oft mit Zufallsgrößen zu tun, von denen man auf Grund des zentralen Grenzwertsatzes mit gutem Grund annehmen kann, daß sie annähernd normalverteilt sind. Je nachdem, ob der Erwartungswert μ oder die Standardabweichung σ unbekannt sind, werden sich Hypothesen über diese Parameter ergeben.
Wir wollen nur den einfachen Fall besprechen, daß eine Hypothese über den unbekannten Erwartungswert μ einer angenähert normalverteilten Zufallsgröße zu testen ist, wobei die Standardabweichung σ bekannt ist. Ein solcher Test heißt ***Gaußtest.***

Beispiel: Die Untersuchung von Drähten einer bestimmten Legierung ergab für die Zugfestigkeit den Mittelwert $41{,}62\,\mathrm{N/mm^2}$ und die Standardabweichung $0{,}60\,\mathrm{N/mm^2}$. Wir dürfen annehmen, daß die Zufallsgröße »Zugfestigkeit eines Drahtes« angenähert normalverteilt ist mit $\mu_0 = 41{,}62\,\mathrm{N/mm^2}$ und $\sigma_0 = 0{,}60\,\mathrm{N/mm^2}$. Eine Versuchsserie an 80 Drähten mit einer etwas veränderten Legierung ergab eine mittlere Zugefstigkeit von $41{,}50\,\mathrm{N/mm^2}$ bei gleicher Standardabweichung. Kann man auf dem 5%-Niveau die Hypothese »Die mittlere Zugfestigkeit hat sich nicht verändert« ablehnen?
Ob man einseitig oder zweiseitig testen wird, hängt davon ab, welche Alternativen man in Betracht ziehen will. Man kann sich auf den Standpunkt stellen, daß die Zugfestigkeit sowohl größer als auch kleiner geworden ist, und dann zweiseitig testen, oder man nimmt auf Grund des Stichprobenergebnisses an, daß die Zugfestigkeit höchstens kleiner geworden sein kann, und dann einseitig testen.

Lösung: Die neue Legierung besitze die mittlere Zugfestigkeit μ, die unbekannt ist. Die Zugfestigkeit eines Drahtes ist dann angenähert normalverteilt mit μ und σ_0^2. Im Zufallsexperiment wurde die Stichprobe $(X_1 | X_2 | \ldots | X_{80})$ bestimmt;