

Wahrscheinlichkeit $1 - \alpha$ einschließt. Wir nehmen an, dass n so groß ist, dass die Normalapproximation zulässig ist. Dann ist $\bar{X}$ näherungsweise nach $\mathcal{N}(\mu, \sigma^2/n)$ verteilt, und aus Gl. (6.10) folgt

$$\mathbb{P}\left(\bar{X} \in \left[\mu \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]\right) \simeq 1 - \alpha. \quad (6.11)$$

Dabei bedeutet $\simeq$ ungefähre Gleichheit mit einem kleinen, nicht weiter quantifizierten Fehler. Sind die X_i normalverteilt, so gilt (6.11) sogar *exakt*, auch für kleine n ; das folgt aus dem Additionssatz der Normalverteilung. In jedem Fall wird das Intervall, das den Mittelwert mit vorgegebener Wahrscheinlichkeit $1 - \alpha$ einschließt, immer kleiner, je größer n ist. Wir können auch sagen, wie groß n sein muss, damit das Intervall eine vorgegebene Breite nicht überschreitet. Denn aus (6.11) können wir folgern: Ist $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, so ist die Lösung von

$$z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \delta, \quad \text{also} \quad n = \left(\frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\delta}\right)^2, \quad (6.12)$$

die Zahl der Versuche, die mindestens ausgeführt werden muss, damit die Abweichung des Mittelwerts vom Erwartungswert mit Wahrscheinlichkeit $1 - \alpha$ nicht größer ist als δ . Überlegungen dieser Art spielen in der *Versuchsplanung* eine außerordentlich wichtige Rolle, nicht zuletzt bei der Planung der Stichprobengröße in medizinischen Studien.

Beispiel. Wie oft muss man (mit einem idealen Würfel) mindestens würfeln, damit der Mittelwert der Augenzahl mit 95% Wahrscheinlichkeit zwischen 3.4 und 3.6 liegt? Wir kennen Erwartungswert und Varianz der Augenzahl beim einmaligen Würfeln als $\mu = 3.5$, $\sigma^2 = 2.917$. Es ist $\delta = 0.1$ und $z_{0.975} = 1.96$. Einsetzen in (6.12) ergibt $n = \frac{1.96^2 \cdot 2.917}{0.1^2} \simeq 1121$.

Wir halten nun fest, dass die Aussage “ $\bar{X} \in [\mu \pm \delta]$ ” gleichbedeutend ist mit “ $\mu \in [\bar{X} \pm \delta]$ ” — denn beides besagt, dass $\bar{X}$ und μ nicht weiter als δ voneinander entfernt sind. Wir können (6.11) daher auch lesen als

$$\mathbb{P}\left(\mu \in \left[\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]\right) \simeq 1 - \alpha. \quad (6.13)$$

Das zufällige Intervall $[\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$ heißt **Konfidenzintervall** (oder **Vertrauensintervall**) für den Erwartungswert μ zum Konfidenzniveau $1 - \alpha$. Ganz allgemein definiert man:

Definition 6.2 (Konfidenzintervall). Seien C_1 und C_2 (mit $C_1 < C_2$) Zufallsvariablen auf einem Wahrscheinlichkeitsraum (diese können etwa über eine mathematische Stichprobe $(X_1, \dots, X_n)$ definiert sein). Es sei u ein reeller Parameter, und $0 < \alpha < 1$ eine reelle Zahl. Das zufällige Intervall $[C_1, C_2]$ heißt **Konfidenzintervall für den unbekannten Parameter u zum Konfidenzniveau $1 - \alpha$** , wenn die Beziehung

$$\mathbb{P}(C_1 \leq u \leq C_2) = 1 - \alpha$$

erfüllt ist. In diesem Fall nennt man für Realisierungen c_1 und c_2 von C_1 und C_2 das Intervall $[c_1, c_2]$ das **beobachtete** Konfidenzintervall für u zum Konfidenzniveau $1 - \alpha$.

Bemerkung. Ein Konfidenzintervall $[C_1, C_2]$ ist ein *zufälliges* Intervall. Werden C_1 und C_2 durch die entsprechenden Realisierungen c_1 und c_2 ersetzt, die etwa über eine beobachtete Stichprobe $(x_1, \dots, x_n)$ bestimmt sind, so erhält man das *beobachtete* Konfidenzintervall $[c_1, c_2]$. Wir werden im folgenden solche Intervalle $[c_1, c_2]$ einfach Konfidenzintervalle nennen, weil klar ist, dass es sich um beobachtete Konfidenzintervalle handelt. Die Aussage " $u \in [c_1, c_2]$ " ist dann in $(1 - \alpha) \cdot 100\%$ aller Fälle richtig. Vorsicht: Dies heißt nicht, dass mit Wahrscheinlichkeit $(1 - \alpha)$ der wahre Wert im Intervall $[c_1, c_2]$ liegt; solch eine Aussage macht gar keinen Sinn, weil ja u gegeben ist und c_1, c_2 zufällige Realisierungen sind.

Im Allgemeinen müssen Konfidenzintervalle nicht symmetrisch sein, wir werden es in diesem Kurs aber nur mit symmetrischen Intervallen zu tun haben. Insbesondere ist die Symmetrie immer dann eine natürliche Eigenschaft, wenn es um den Erwartungswert einer Verteilung geht, die symmetrisch um eben diesen Erwartungswert ist.

Die folgende Liste gibt über einige wichtige Anwendungen Auskunft.

1. **Konfidenzintervall für den Erwartungswert der Normalverteilung:** Gegeben sei eine Stichprobe vom Umfang n mit Mittelwert $\bar{x}$. Es wird vorausgesetzt, dass die zugrundeliegende Zufallsgröße normalverteilt ist, oder dass n so groß ist, dass die Normalapproximation anwendbar ist.

- (a) Wenn die **Varianz** σ^2 **von** X **bekannt** ist, gilt (6.11), und das (symmetrische) $1 - \alpha$ -Konfidenzintervall für den Erwartungswert μ von X lautet:

$$\left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

- (b) Wenn σ^2 unbekannt ist, muss mit der **empirischen Varianz** s^2 gearbeitet werden. In Analogie zu Z_n aus (4.4) betrachten wir zu diesem Zweck die ZV

$$T := \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

wobei S^2 die empirische Varianz *als Zufallsvariable* bezeichnet, vgl. Abschnitt 6.1.4. Man kann zeigen: Die Zufallsvariable T ist **Student-t-verteilt** mit $k = n - 1$ Freiheitsgraden (kurz: t_{n-1} -verteilt),

$$\mathbb{P}(T \leq t) = c \int_{-\infty}^t \left(1 + \frac{\tau^2}{n-1} \right)^{-n/2} d\tau,$$

wobei c eine explizit bekannte, von n abhängige Normierungskonstante ist. (Wie bei der Normalverteilung kennt man keinen geschlossenen

Ausdruck für die Verteilungsfunktion und muss die zugehörigen Werte in Tabellen nachschlagen.) Wie die Standardnormalverteilung hat Die Dichte der t_{n-1} -Verteilung hat — wie die der Standardnormalverteilung — die Gestalt einer Glockenkurve und ist symmetrisch um 0. Das symmetrische $1 - \alpha$ -Konfidenzintervall für den Erwartungswert μ ist dann

$$\left[\bar{x} \pm t_{n-1, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right], \quad (6.14)$$

wobei $t_{n-1,q}$ das q -Quantil der t_{n-1} -Verteilung ist, also derjenige Wert, der

$$\mathbb{P}(T \leq t_{n-1,q}) = q$$

erfüllt. Quantilwerte können der (Quantil-)Tabelle der Student-t-Verteilung entnommen werden. Die t-Verteilung hat eine größere Varianz als die Standardnormalverteilung (die Glockenkurve ist etwas ‘breiter’), daher ist $t_{n-1,q}$ für endliche n immer größer als z_q (so ist z.B. $t_{2,0.975} = 4.3 > 1.96 = z_{0.975}$); entsprechend größer fällt das Intervall aus. Es gilt jedoch $\lim_{n \rightarrow \infty} t_{n-1} = \mathcal{N}(0, 1)$ und somit auch $\lim_{n \rightarrow \infty} t_{n-1,q} = z_q$. Tatsächlich ist für große Werte ($n \geq 30$) die Approximation durch die Normalverteilung eine gute Näherung. Dies kann man auch schon mit der expliziten Darstellung der Verteilungsfunktion erkennen (vgl. die Herleitung der Poisson-Verteilung und den Beweis des ZGWS).

2. Konfidenzintervall für den Parameter p eines Bernoulli-Experiments.

Bei n Wiederholungen eines Bernoulli-Experiments wurde n_A mal das Ereignis A beobachtet, woraus sich $\bar{x} = \hat{p} = n_A/n$ als Schätzer für p , die Wahrscheinlichkeit von A , ergibt. Für großes n erhält man als Schätzer für die Varianz des (einfachen) Bernoulli-Experiments

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left[\left(\sum_{i=1}^n x_i^2 \right) - n\bar{x}^2 \right] = \frac{1}{n-1} (n\bar{x} - n\bar{x}^2) = \frac{n}{n-1} \bar{x}(1 - \bar{x}) \\ &= \frac{n}{n-1} \hat{p}(1 - \hat{p}) \simeq \hat{p}(1 - \hat{p}), \end{aligned}$$

wobei wir im zweiten Schritt benutzt haben, dass $x_i^2 = x_i$ da $x_i \in \{0, 1\}$. Für $n\hat{p}(1 - \hat{p}) \geq 9$ ist die Normalapproximation zulässig, so dass das Konfidenzintervall für den Erwartungswert $\mu = p$ des Bernoulliexperiments durch (6.14) gegeben ist. Einsetzen von $\bar{x}$ und s^2 in (6.14) liefert somit

$$\left[\hat{p} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} \right] \quad (6.15)$$

als ist das symmetrische $1 - \alpha$ -Konfidenzintervall für p . (Falls (6.15) Intervallgrenzen < 0 bzw. > 1 liefert, müssen diese natürlich durch 0 bzw. 1

ersetzt werden.) Beachten Sie, dass in diesem Beispiel die Varianz unbekannt ist und deshalb eigentlich mit der t -Verteilung gerechnet werden muss; weil n groß ist, verwenden wir die Approximation von $t_{n-1, 1-\frac{\alpha}{2}}$ durch $z_{1-\frac{\alpha}{2}}$.

Beispiel. Die Geburtsstatistik der Schweiz weist zwischen 1950 und 1970 $n = 1944700$ Geburten aus, davon $n_J = 997600$ Jungengeburten. Es sollen das 99% und das 99.9% Konfidenzintervall für die Wahrscheinlichkeit p einer Jungengeburt ermittelt werden. Man erhält $\hat{p} = n_J/n = 0.512984$ und $s^2 = 0.249831$. Weiterhin ist, da n riesig ist, die Normalapproximation anwendbar, die relevanten Quantilwerte sind $z_{0.995} = 2.575$ und $z_{0.9995} = 3.300$. Mit diesen Zahlen ergeben sich die 99% und 99.9% Konfidenzintervalle als $[0.51206, 0.51391]$ bzw. $[0.51190, 0.51416]$. Keines schließt den Wert 0.5 ein. Wir können daher mit Recht behaupten, dass Jungen- und Mädchengeburten nicht gleich häufig sind: Wir wissen dies zwar nicht mit Sicherheit, haben aber ein Verfahren benutzt, das mit 99.9%iger Sicherheit wahre Behauptungen liefert.

Wir haben in diesem Abschnitt Konfidenzintervalle für Erwartungswerte und verwandte Größen bestimmt, unter der Normalverteilungsannahme. Gelegentlich benötigt man auch Konfidenzintervalle für die Varianz oder für andere Parameter einer (unbekannten) Verteilung. Diese Probleme kann man mit den hier besprochenen Methoden behandeln, siehe die angeführte Literatur.

6.3 Das Testen von Hypothesen

Experimentelle Daten dienen letzten Endes immer dem Zweck, bestimmte Vermutungen zu bestätigen oder zu widerlegen. Bei zufälligen Größen kann man solche Entscheidungen nie mit letzter Sicherheit treffen; man benötigt hier das Konzept des **statistischen Tests**.

Eine **Hypothese** ist in diesem Zusammenhang ganz allgemein eine Annahme über eine Zufallsvariable. Ein **Test** einer Hypothese ist ein Prüfverfahren, das man anwendet, um festzustellen, ob die Hypothese abgelehnt werden soll oder nicht. Dabei stehen jeweils zwei Hypothesen zur Debatte: Die zu prüfende oder **Nullhypothese** H_0 , und die **Alternativhypothese** H_A , für die man sich bei Ablehnung von H_0 entscheidet. Das Konzept des statistischen Tests soll nun am Beispiel des sogenannten **Gauß-Tests** illustriert werden, s. Abb. 6.4.

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt, μ aber unbekannt ist. Mit Hilfe einer Stichprobe vom Umfang n möchte man $H_0 : \mu = \mu_0$ gegen $H_A : \mu \neq \mu_0$ überprüfen.

Als **Prüfgröße** dient der Stichprobenmittelwert $\bar{X}$. Wir geben (vor der Durchführung des Tests!) eine kleine positive Zahl α vor, das **Signifikanzniveau** des Tests oder die **Irrtumswahrscheinlichkeit**. Wir überlegen dann, welcher Verteilung der Stichprobenmittelwert folgen würde, wenn die H_0 richtig wäre: Unter H_0 wäre $\bar{X} \sim \mathcal{N}(\mu_0, \sigma^2/n)$. Wir lehnen die H_0 dann ab, wenn der beobachtete Stichprobenmittelwert so weit von μ_0 abweicht, dass es "zu unwahrscheinlich" ist, dass es

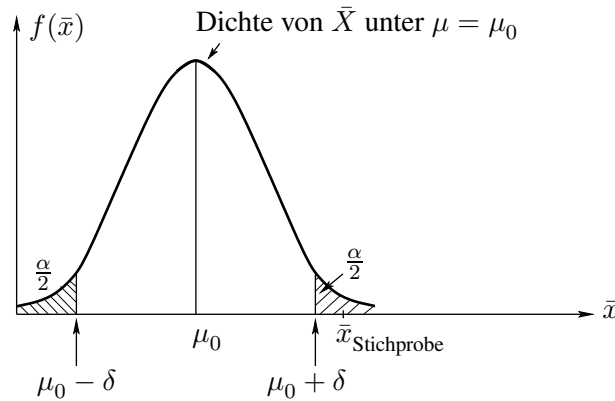


Abbildung 6.4: Prinzip des einfachen, zweiseitigen Gauß-Tests. μ_0 ist der Erwartungswert unter H_0 . Die Nullhypothese wird abgelehnt, wenn der Stichprobenmittelwert, hier mit $\bar{x}_{\text{Stichprobe}}$ bezeichnet, in den Ablehnungsbereich fällt.

sich um eine zufällige Schwankung handelt. Genauer: H_0 wird abgelehnt, wenn $\bar{x}$ in den **Ablehnungsbereich** $\{\bar{x} : |\bar{x} - \mu_0| > \delta\}$ fällt. Dabei bestimmen wir δ so, dass

$$\mathbb{P}(|\bar{X} - \mu_0| > \delta) = \alpha, \quad \text{wenn } \mu = \mu_0. \quad (6.16)$$

Der Ablehnungsbereich besteht also gerade aus den — unter H_0 — $\alpha \cdot 100\%$ extremsten Werten. Nach Gl. (6.12) muss δ dann gerade als

$$\delta = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (6.17)$$

gewählt werden. Wir verwerfen H_0 also, falls

$$\frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n} > z_{1-\frac{\alpha}{2}}.$$

Andernfalls behalten wir die Nullhypothese bei. Wir können dann allerdings nicht schließen, dass H_0 richtig ist. Wir können lediglich folgern, dass wir aufgrund der Stichprobe keine signifikante Abweichung zum Signifikanzniveau $1 - \alpha$ feststellen können.

Noch ein Wort zur Wahl der Nullhypothese. Die Faustregel ist, dass die Nullhypothese “konservativ” ist: Sie verneint in der Regel Unterschiede, Veränderungen usw., deren Feststellung das bisher Bekannte in Frage stellen könnte. Will man umgekehrt einen vermuteten Zusammenhang erhärten, so wählt man als H_0 die Verneinung des Zusammenhangs; wird diese dann aufgrund der Beobachtungen abgelehnt, so darf der Zusammenhang als statistisch signifikant angesehen werden.

Beispiel. Für das *Geburtenbeispiel* wird $H_0: p = 1/2$ gegen $H_A: p \neq 1/2$ zum Signifikanzniveau $\alpha = 0.001$ getestet. Unter der Nullhypothese ist — für jede einzelne Geburt — $\mu_0 = p = 1/2$ und $\sigma^2 = p(1 - p) = 1/4$. Da $\frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n} = 36.20 > z_{0.9995} = 3.3$, wird H_0 abgelehnt.

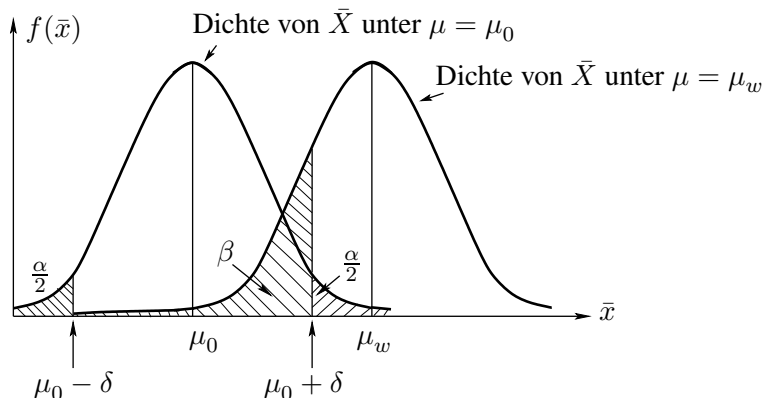


Abbildung 6.5: Fehler 1. Art und Fehler 2. Art im einfachen, zweiseitigen Gauß-Test. μ_0 ist der Erwartungswert unter H_0 , μ_w ist der wahre (aber unbekannte) Erwartungswert.

Wir sind hier — wie beim Konfidenzintervall — wieder mit der Tatsache konfrontiert, dass man statistische Aussagen nie mit letzter Sicherheit treffen kann. Denn selbst der Wert $\bar{x} = 0.512984$ im Geburtenbeispiel, der — relativ zur Standardabweichung von $\bar{X}$ unter H_0 — eine sehr große Abweichung vom Erwartungswert darstellt, ist unter H_0 möglich; allerdings ist dieser oder ein noch extremerer Wert äußerst unwahrscheinlich. Wir müssen uns deshalb genauer mit den Fehlern auseinander setzen, die beim Testen auftreten können. Es sind dies:

Fehler 1. Art: Die Nullhypothese wird abgelehnt, obwohl sie richtig ist; der Test ist gerade so konstruiert, dass dies mit der vorgegebenen Irrtumswahrscheinlichkeit α geschieht (bei einseitigen Fragestellungen (s.u.) mit Wahrscheinlichkeit $\leq \alpha$).

Fehler 2. Art: Die Nullhypothese wird beibehalten, obwohl sie falsch ist. Die Wahrscheinlichkeit β , mit der dies geschieht, ist in Abb. 6.5 illustriert. Sie hängt von der wahren (aber unbekannten) Verteilung ab und ist deshalb sehr viel schwerer zu quantifizieren als der Fehler 1. Art. Allgemein lässt sich aber sagen, dass die Wahl eines kleineren α zu einem größeren β führt und umgekehrt.

Dies ist in der folgenden Tabelle noch einmal zusammengefasst, vergleiche auch Abbildung 6.5.

	Das Testergebnis lautet	
	H_0 ablehnen	H_0 beibehalten
H_0 ist richtig	Fehler 1. Art α	kein Fehler $1 - \alpha$
H_0 ist falsch	kein Fehler $1 - \beta$	Fehler 2. Art β

Einseitige und zweiseitige Tests. Der bisher betrachtete Gauß-Test war ein sogenannter *zweiseitiger* Test. Hier haben wir nur danach gefragt, ob zwei Werte gleich

oder verschieden sind, wobei kein Anfangsverdacht geäußert wurde, welcher der größere ist, wir hatten also eine *zweiseitige Fragestellung*. Entsprechend lautete das Hypothesenpaar $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$, und große Abweichungen der Prüfgröße nach oben *oder* nach unten (bezüglich des Idealwerts μ_0) führten gleichermaßen zur Ablehnung der Nullhypothese (wie im GEBURTENBEISPIEL). Ist man dagegen nur an Abweichungen in einer Richtung interessiert, so ist die Fragestellung *einseitig*. Typische Beispiele sind Fragen wie: Werden Stickoxidgrenzwerte überschritten? Ist der Cholesterinspiegel zu hoch? Sind nur Abweichungen nach oben (unten) relevant, so lautet das Hypothesenpaar $H_0 : \mu \leq \mu_0$ versus $H_A : \mu > \mu_0$ ($H_0 : \mu \geq \mu_0$ versus $H_A : \mu < \mu_0$). Der Ablehnungsbereich besteht dann aus den (unter H_0) $\alpha \cdot 100\%$ größten (kleinsten) Werten; vgl. Abb. 6.6. Genauer wählt man für das Hypothesenpaar $H_0 : \mu \leq \mu_0$ versus $H_A : \mu > \mu_0$ unser δ so, dass, in Analogie zu (6.16):

$$\mathbb{P}(\bar{X} - \mu_0 > \delta) = \alpha, \quad \text{wenn } \mu = \mu_0 \quad (6.18)$$

und lehnt H_0 ab, wenn $\bar{X} - \mu_0 > \delta$; für das Hypothesenpaar $H_0 : \mu \geq \mu_0$ versus $H_A : \mu < \mu_0$ wählt man δ so, dass

$$\mathbb{P}(\bar{X} - \mu_0 < -\delta) = \alpha, \quad \text{wenn } \mu = \mu_0 \quad (6.19)$$

und lehnt H_0 ab, wenn $\bar{X} - \mu_0 < -\delta$. Dies garantiert, dass der Fehler erster Art höchstens gleich dem vorgegebenen α ist:

$$\mathbb{P}(H_0 \text{ wird abgelehnt wenn wahr}) \leq \alpha.$$

Die Ungleichung rührt daher, dass wir die H_0 hier als Ungleichung formuliert haben. Gleichheit gilt gerade dann, wenn $\mu = \mu_0$; andernfalls ist der Fehler erster Art kleiner als α . Ganz allgemein arbeitet man bei einseitigen Fragestellungen mit der Verteilung der Prüfgröße im ‘ungünstigsten’ Fall, d.h. in dem Fall, der den größten Fehler 1. Art liefert; dieser entspricht dem Gleichheitszeichen.

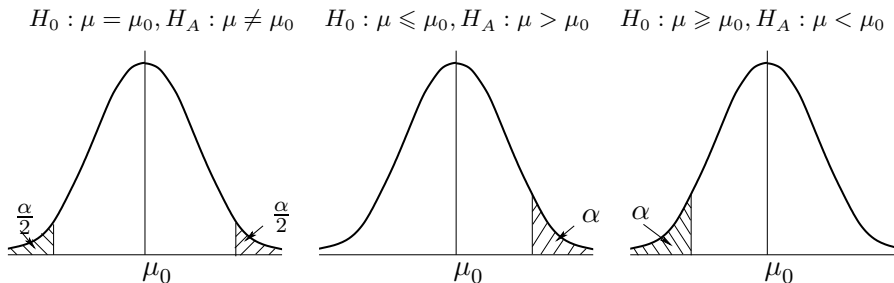


Abbildung 6.6: Zweiseitiger Test (links) und einseitige Tests (Mitte und rechts).

Allgemeines Prinzip: Das bisher beschriebene Beispiel illustriert ein ganz allgemeines Prinzip, nach dem für sehr allgemeine Situationen (nicht nur die hier besprochenen) Tests hergeleitet werden können. Man geht in vier Schritten vor:

1. Hypothesenpaar formulieren (betrifft oft Erwartungswerte, oder andere Parameter von Verteilungen) und α festlegen
2. Prüfgröße festlegen (eine geeignete, aus der Stichprobe ermittelte Größe; oft der Stichprobenmittelwert, aber keineswegs immer)
3. Verteilung der Prüfgröße unter H_0 bestimmen (oft eine Normalverteilung, aber keineswegs immer)
4. H_0 ablehnen, wenn der gemessene Wert der Prüfgröße in den Ablehnungsbereich fällt, das sind die (unter H_0 höchstens) $\alpha \cdot 100\%$ extremsten Werte der Prüfgröße. Manchmal ist die explizite Konstruktion des Ablehnungsbereichs gar nicht notwendig, wie Sie im folgenden Beispiel sehen werden.

Beispiel. Louis Pasteur impfte im Jahr 1880 sechs Hühner mit einem Serum gegen Cholera. Als er diese und sechs ungeimpfte Hühner mit Cholera-Erregern infizierte, waren am nächsten Tag sechs Hühner tot. Diese 6 toten Hühner waren gerade die 6 ungeimpften.

Wir wollen nun statistisch nachweisen, dass das Serum eine Wirkung hatte und gehen dabei gemäß obiger ‘Gebrauchsanleitung’ vor:

1. Die Fragestellung ist einseitig. Wir testen die Nullhypothese ‘Tod durch Cholera ist bei ungeimpften Hühnern nicht häufiger als bei geimpften’ gegen die Alternativhypothese ‘Tod durch Cholera ist bei ungeimpften Hühnern häufiger’ ($\alpha = 0.05$).
2. Als Prüfgröße nehmen wir X , die Anzahl der ungeimpften Hühner unter den 6 toten.
3. Falls das Serum unwirksam gewesen wäre, wären die 6 toten Hühner eine zufällige Auswahl aus den 6 geimpften und den 6 ungeimpften gewesen, also $X \sim \text{hyper}(6, 6, 6)$. Dies ist die Verteilung der Prüfgröße unter H_0 (‘mit dem Gleichheitszeichen’).
4. Wir konstruieren zunächst den Ablehnungsbereich. Wegen der einseitigen Fragestellung interessieren uns die unter H_0 höchstens 5 % größten Werte. Wir erhalten

$$\mathbb{P}(X = 6) = \frac{\binom{6}{6} \binom{6}{0}}{\binom{12}{6}} = \frac{1 \cdot 1}{924} = 0.0011$$

$$\mathbb{P}(X = 5) = \frac{\binom{6}{5} \binom{6}{1}}{\binom{12}{6}} = \frac{6 \cdot 6}{924} = 0.038$$

$$\mathbb{P}(X = 4) = \frac{\binom{6}{4} \binom{6}{2}}{\binom{12}{6}} = \frac{15 \cdot 15}{924} = 0.244.$$

Da $\mathbb{P}(X = 5) + \mathbb{P}(X = 6) = 0.039 < 0.05 = \alpha < 0.283 = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6)$, besteht der Ablehungsbereich aus den Werten 5 und 6. Man wird also dem Serum eine Wirkung zuschreiben, wenn mindestens 5 der toten Hühner ungeimpft waren. Im vorliegenden Fall wird die H_0 also abgelehnt, die Wirkung des Serums ist statistisch signifikant.

In diesem Beispiel kann man sich die Konstruktion des Ablehungsbereichs allerdings auch sparen. Denn es reicht, folgendermaßen zu argumentieren: Wir haben $X = 6$ beobachtet. Die Wahrscheinlichkeit unter H_0 für diesen oder einen noch größeren Wert ist $\mathbb{P}(X = 6) = 0.0011$ (in diesem Fall ist ja gar kein größerer Wert möglich, aber Vorsicht: Im Allgemeinen sind hier *alle* Werte zu beachten, die mindestens so extrem sind wie der beobachtete; hätte man also $X = 5$ beobachtet, wäre die relevante Wahrscheinlichkeit $\mathbb{P}(X = 5) + \mathbb{P}(X = 6) = 0.039$). Da $0.0011 < \alpha$, gehört der beobachtete Wert offenbar zum Ablehnungsbereich.

6.3.1 Der Zoo der Erwartungswerttests

Nachdem wir das Prinzip des statistischen Tests kennengelernt haben, soll jetzt eine Übersicht über einige Tests folgen, die sich auf Erwartungswerte von Verteilungen beziehen (sogenannte Erwartungswerttests). Wir besprechen hier solche Tests, in denen die betrachteten Zufallsvariablen normalverteilt sind oder die Stichprobe so groß ist, dass die Normalapproximation anwendbar ist. Je nach Situation sind verschiedene Tests zu verwenden. Um den jeweils “richtigen” Test zu finden, geht man anhand folgender Kriterien vor:

1. Hat man es mit einer oder mit zwei Zufallsvariablen zu tun? Im ersten Fall vergleicht man seine Beobachtungen mit einem vorgegebenen “Idealwert” μ_0 einer Zufallsvariablen X (wie im GEBURTENBEISPIEL). Der Idealwert stammt oft aus der Theorie, oder aus langjähriger Erfahrung (‘Lehrbuchwert’). Das Hypothesenpaar lautet dann z.B.

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu \neq \mu_0,$$

und man macht einen **Ein-Stichproben-Test** (manchmal auch **einfacher Test** genannt). Im zweiten Fall vergleicht man zwei Zufallsgrößen X und Y miteinander (oft sind das Daten aus “Experiment” und “Kontrolle”, z.B. das Wachstum von Pflanzen mit und ohne Wachstumshormon); hier lautet das typische Hypothesenpaar $H_0 : \mu_X = \mu_Y$ versus $H_A : \mu_X \neq \mu_Y$, und man macht **Zwei-Stichproben-Tests** (auch **doppelte Tests** genannt).

2. Ist die Varianz der Zufallsgröße (unter der Nullhypothese) bekannt, oder muss sie aus der Stichprobe geschätzt werden? Ist die Varianz bekannt, so verwendet man **Gauß-Tests**, ansonsten **t-Tests**.
3. Ist die Fragestellung **einseitig** oder **zweiseitig**? Diese Unterscheidung haben wir weiter oben schon besprochen.

Bei dem eingangs besprochenen Beispiel (und insbesondere beim Geburtenbeispiel) handelt es sich also um einen einfachen, zweiseitigen Gauß-Test.

Es folgt nun eine Übersicht über wichtige Erwartungswerttests.

I. Ein-Stichproben-Tests

Die Frage, ob der unbekannte Erwartungswert μ einer normalverteilten Zufallsvariablen X vom vermuteten Wert μ_0 abweicht, soll zum Signifikanzniveau α beantwortet werden. Gegeben sei eine Stichprobe vom Umfang n mit Mittelwert $\bar{x}$ und empirischer Varianz s^2 .

1. **Einfacher Gauß-Test:** Die Varianz σ^2 (unter H_0) wird als bekannt vorausgesetzt. Der Test beruht darauf, dass die ZV (die *standardisierte Prüfgröße*)

$$Z := \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

standardnormalverteilt ist, falls $\mu = \mu_0$ gilt.

- (a) Zweiseitiger Test: $H_0 : \mu = \mu_0$ gegen $H_A : \mu \neq \mu_0$. Da unter H_0 gilt

$$\mathbb{P}(|Z| > z_{1-\frac{\alpha}{2}}) = \alpha$$

(vergleiche (6.16) und (6.17)), verwirft man H_0 , falls

$$\frac{|\bar{x} - \mu_0|}{\sigma} \sqrt{n} > z_{1-\frac{\alpha}{2}},$$

wobei $z_{1-\frac{\alpha}{2}}$ wie immer das $1 - \frac{\alpha}{2}$ -Quantil der Standardnormalverteilung bezeichnet. (Dieser Test wurde schon weiter oben ausführlich besprochen.)

- (b) Einseitiger Test: $H_0 : \mu \leq \mu_0$ gegen $H_A : \mu > \mu_0$. Da unter H_0 (ausgewertet am Gleichheitszeichen) gilt

$$\mathbb{P}(Z > z_{1-\alpha}) = \alpha$$

(vergleiche (6.18) sowie die Definition des Quantils (6.9)), verwirft man H_0 , falls

$$\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha}.$$

- (c) Einseitiger Test: $H_0 : \mu \geq \mu_0$ gegen $H_A : \mu < \mu_0$. Da nun unter H_0 (ausgewertet am Gleichheitszeichen) gilt

$$\mathbb{P}(Z < -z_{1-\alpha}) = \alpha$$

(vergleiche (6.19)), verwirft man H_0 , falls

$$\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} < -z_{1-\alpha}.$$

2. **Einfacher t-Test:** Die Varianz σ^2 ist *unbekannt*; daher muss die *empirische Varianz* s^2 der Stichprobe herangezogen werden. Der Test beruht darauf, dass die ZV (die standardisierte Prüfgröße)

$$T := \frac{\bar{X} - \mu_0}{S} \sqrt{n}$$

unter $\mu = \mu_0$ Student-t-verteilt ist mit $k = n - 1$ Freiheitsgraden (dies ist ganz analog zur Überlegung zu Konfidenzintervallen bei unbekannter Varianz).

- (a) Zweiseitiger Test: $H_0 : \mu = \mu_0$ gegen $H_A : \mu \neq \mu_0$. Da unter H_0 ('ausgewertet mit dem Gleichheitszeichen')

$$\mathbb{P}(|T| > t_{n-1, 1-\frac{\alpha}{2}}) = \alpha,$$

ist H_0 abzulehnen, falls

$$\frac{|\bar{x} - \mu_0|}{s} \sqrt{n} > t_{n-1, 1-\frac{\alpha}{2}},$$

wobei $t_{n-1, 1-\frac{\alpha}{2}}$ das $1 - \frac{\alpha}{2}$ -Quantil der Student-t-Verteilung mit $k = n - 1$ Freiheitsgraden bezeichnet.

- (b) Einseitige Tests: $H_0 : \mu \leq \mu_0$ gegen $H_A : \mu > \mu_0$. Da unter H_0 (jeweils für $\mu = \mu_0$ ausgewertet) gilt, dass

$$\mathbb{P}(T \geq t_{n-1, 1-\alpha}) = \alpha,$$

ist H_0 abzulehnen, falls

$$\frac{\bar{x} - \mu_0}{s} \sqrt{n} \geq t_{n-1, 1-\alpha}.$$

Beispiel. Frage: Ist der Cholesterinwert von Vegetariern signifikant ($\alpha = 0.01$) geringer als der Normalwert (180 mg/100ml)? Man prüft $H_0 : \mu \geq 180$ gegen $H_A : \mu < 180$. Bei einer Gruppe von 9 Vegetariern werden folgende Werte gemessen:

154, 119, 177, 150, 138, 185, 167, 158, 174.

Man erhält $\bar{x} = 158$, $s = 20.7$, und wegen

$$\frac{158 - 180}{20.7} \sqrt{9} = -3.19 < -2.90 = -t_{8, 0.99}$$

wird H_0 abgelehnt. Der Cholesterinwert von Vegetariern ist also signifikant niedriger als der Normalwert.