Testverteilungen

Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ und

$$\overline{X} = \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

sowie

$$S^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

t-Verteilung

Die Verteilung von

$$T = \sqrt{n} \frac{\overline{X} - \mu}{S}$$

heißt t-**Verteilung mit** $n-1$ **Freiheitsgraden**.
Notation: $t(n-1)$. p-Quantil: $t(n-1)_p$.

χ^2-Verteilung

Sind $U_1, \ldots, U_k$ i.i.d. $\sim N(0,1)$, dann heißt die Verteilung von

$$Q = \sum_{i=1}^{k} U_i^2$$

χ^2-**Verteilung mit** k **Freiheitsgraden**.
Momente: Es gilt: $E(Q) = k$ und $\mathrm{Var}(Q) = 2k$.
Gilt mit einer Konstanten $c > 0$:

$$T/c \sim \chi^2(k),$$

dann heißt T **gestreckt** χ^2**-verteilt mit** k **Freiheitsgraden**.
Man schreibt auch: $T \sim c \cdot \chi^2(k)$.

Verteilung der Varianzschätzer

Annahme: Normalverteilungsmodell, d.h.

$$X_1, \dots, X_n \overset{d}{\sim} N(\mu, \sigma^2)$$

Welchen Varianzschätzer $\widehat{\sigma}_n^2$ wann verwenden?

Fall 1: μ bekannt: Verwende $\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$. Dann gilt (per def.)

$$\frac{n}{\sigma^2} \widehat{\sigma}_n^2 \sim \chi^2(n)$$

Fall 2: μ unbekannt. Verwende $\widehat{\sigma}_n^2 := S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$. Dann:

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1).$$

F-Verteilung

F-Verteilung

Seien $Q_1 \sim \chi^2(n_1)$ und $Q_2 \sim \chi^2(n_2)$ unabhängig. Dann heißt die Verteilung des Quotienten

$$F = \frac{Q_1/n_1}{Q_2/n_2}$$

F-**Verteilung mit** n_1 **und** n_2 **Freiheitsgraden**.
Notation: $F(n_1, n_2)$.
p-Quantil: $F(n_1, n_2)_p$.
Momente: $E(F) = \frac{n_2}{n_2-2}$, $\mathrm{Var}(F) = \frac{2n_2^2(n_2+n_1-2)}{n_1(n_2-2)^2(n_2-4)}$.

F-Verteilung: Vergleich von Varianzschätzungen

$X_{11}, \ldots, X_{1,n_1}$ und $X_{21}, \ldots, X_{2,n_2}$ seien zwei unabhängige normalverteilte Stichproben mit

$$X_{1i} \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2), \qquad i = 1, \ldots, n_1,$$

und

$$X_{2i} \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2), \qquad i = 1, \ldots, n_2,$$

Erwartungstreue und unabhängig Schätzungen der Varianzen σ_1^2 und σ_2^2 sind

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2, \qquad i = 1, 2.$$

Zahlenbeispiel: $s_1^2 = 3.5$ und $s_2^2 = 5.5$. Frage: Besteht tatsächlich ein Unterschied?

F-Verteilung

Man kann prinzipiell $S_2^2 - S_1^2$ mit 0 vergleichen oder S_1^2/S_2^2 mit 1. In der Statistik betrachtet man den Quotienten, da dieser einer (gestreckten) F-Verteilung folgt:

$$Q_i = \frac{n_i - 1}{\sigma_i^2} S_i^2 \sim \chi^2(n_i - 1), \qquad i = 1, 2.$$

Q_1 und Q_2 sind unabhängig und χ^2-verteilt. Daher gilt:

$$F = \frac{Q_1/(n_1 - 1)}{Q_2/(n_2 - 1)} \sim F(n_1 - 1, n_2 - 1)$$

Ausrechnen:

$$F = \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2}.$$

Im Fall $\sigma_1^2 = \sigma_2^2$ folgt: $F \sim F(n_1 - 1, n_2 - 1)$.

Konfidenzintervalle

Kritik an Punktschätzungen:
Ein Datensatz liefere:

$$\overline{x} = 11.34534, \qquad s/\sqrt{n} = 5.45$$

Hinweis: $s/\sqrt{n}$ ist eine Schätzung der Standardabweichung von $\overline{X}$ und heißt **Standardfehler**.
Die Angabe vieler Nachkommastellen suggeriert eine Genauigkeit, die statistisch nicht unbedingt gerechtfertigt ist!
Besser:
Gebe ein datenbasiertes Intervall $[L, U]$ an, welches mit einer definierten (Mindest-) Wahrscheinlichkeit den Parameter überdeckt.

Anschauung: Sollte die Schätzung mit einem Microliner oder einem mehr oder weniger dicken Edding markiert werden?

Konfidenzintervalle

Konfidenzintervall

Ein Intervall $[L, U]$ mit datenabhängigen Intervallgrenzen

$$L = L(X_1, \ldots, X_n)$$
$$U = U(X_1, \ldots, X_n)$$

heißt **Konfidenzintervall (Vertrauensbereich) zum Konfidenzniveau** $1 - \alpha$, wenn f‚r alle $\vartheta \in \Theta$ gilt:

$$P([L, U] \ni \vartheta) \geq 1 - \alpha.$$

Im Unterschied hierzu: **Prognoseintervall** für eine ZV X:

$$P(a < X \leq b) \geq 1 - \alpha$$

(Nehme Quantile $a = F_X^{-1}(\alpha/2)$ und $b = F_X^{-1}(1 - \alpha/2)$.)

Konfidenzintervall für μ

Modell:

$$X_1, \dots, X_n \overset{d}{\sim} N(\mu, \sigma^2)$$

Ausgangspunkt: Prognoseintervall für $T = \sqrt{n}(\overline{X} - \mu)/S \sim t(n-1)$:
Mit Wahrscheinlichkeit $1 - \alpha$ gilt:

$$-t(n-1)_{1-\alpha/2} \leq \sqrt{n}\frac{\overline{X} - \mu}{S} \leq t(n-1)_{1-\alpha/2}$$

(Beachte: $t(n-1)_{1-\alpha/2}$ ist das $(1 - \frac{\alpha}{2})$-Quantil der $t(n-1)$-Verteilung!)
Umformen, so dass nur μ in der Mitte stehen bleibt:

$$\overline{X} - t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}}.$$

$(1-\alpha)$-KI für μ ist gegeben durch:

$$[L, U] = \left[\overline{X} - t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Konfidenzintervall für μ

Verbreitet in der Praxis: 'Error Bounds' $\overline{X}_n \pm S_n/\sqrt{n}$ (zu optimistisch!).

Statistiker verwendet $\overline{X}_n \pm t(n-1)_{1-\alpha/2} \frac{S_n}{\sqrt{n}}$ (klare Interpretation).

Mit einer Fehlerwahrscheinlichkeit von α ist die Aussage

$$\overline{X} - t(n-1)_{1-\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t(n-1)_{1-\alpha/2} \frac{S}{\sqrt{n}}$$

über μ korrekt.

Konfidenzintervall für μ

Zweiseitiges KI, σ unbekannt:

$$\left[\overline{X} - t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t(n-1)_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Einseitige KIs:

1. Einseitiges unteres KI: $\left(-\infty, \overline{X} + t(n-1)_{1-\alpha} \cdot S/\sqrt{n}\right]$.
   Mit Wkeit $1-\alpha$ ist die Aussage „$\mu \leq \overline{X} + t(n-1)_{1-\alpha} \cdot S/\sqrt{n}$“ richtig (obere Schranke).
2. Einseitiges oberes KI: $\left[\overline{X} - t(n-1)_{1-\alpha} \cdot S/\sqrt{n}, \infty\right)$ liefert analog eine untere Schranke.

Falls σ bekannt ist: Ersetze in den Formeln:

1. S durch σ.
2. $t(n-1)_{1-\alpha/2}$ durch $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$.
3. $t(n-1)_{1-\alpha}$ durch $z_{1-\alpha}$.

$z_{1-\alpha}$: $(1-\alpha)$-Quantil der $N(0,1)$-Verteilung.

Konfidenzintervall für μ

Computersimulation: Simulation von 10 Stichproben vom Umfang n (=10) aus einer $N(2,1)$-Verteilung.

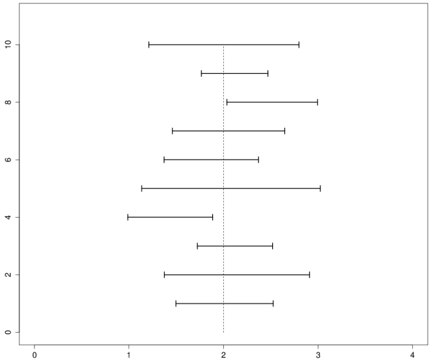

Abbildung: Computersimulation: Dargestellt sind 10 Konfidenzintervalle für μ, die aus 10 unabhängigen Stichproben berechnet wurden. Der im Experiment eingestellte Wert $\mu = 2$ ist gestrichelt eingezeichnet.

Konfidenzintervall für σ^2

Ausgangspunkt: Schätzer $\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Mit Wahrscheinlichkeit $1-\alpha$ gilt:

$$\chi^2(n-1)_{\alpha/2} \leq \frac{(n-1)\widehat{\sigma}^2}{\sigma^2} \leq \chi^2(n-1)_{1-\alpha/2}$$

Umformen liefert zweiseitiges Konfidenzintervall für σ^2:

$$\left[\frac{n-1}{\chi^2(n-1)_{1-\alpha/2}}\,\widehat{\sigma}^2, \frac{n-1}{\chi^2(n-1)_{\alpha/2}}\,\widehat{\sigma}^2\right]$$

Analog:

- einseitiges oberes Konfidenzintervall: $[0, (n-1)\widehat{\sigma}^2/\chi^2(n-1)_\alpha]$
- einseitiges unteres Konfidenzintervall $[(n-1)\widehat{\sigma}^2/\chi^2(n-1)_{1-\alpha}, \infty)$

Konfidenzintervall für p

Modell: $Y \sim \mathrm{Bin}(n, p)$.
Approximatives Konfidenzintervall (aus ZGWS):

$$
\begin{aligned}
L &= \widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\
U &= \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}
\end{aligned}
$$

Konfidenzintervall für p

ZGWS für Binomialverteilung mit geschätztem $\sigma = \sqrt{p(1-p)}$ im Nenner: $\sqrt{n}\frac{\widehat{p}-p}{\sqrt{\widehat{p}(1-\widehat{p})}} \sim_{approx} N(0,1)$.
Mit Wahrscheinlichkeit $1-\alpha$ gilt näherungsweise (für großes n):

$$-z_{1-\alpha/2} \leq \sqrt{n}\frac{\widehat{p}-p}{\sqrt{\widehat{p}(1-\widehat{p})}} \leq z_{1-\alpha/2}$$

Dies ist äquivalent zu (Umformen, so das p in der Mitte stehen bleibt):

$$\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p}}{)}n}$$

Somit überdeckt $[L, U]$ die unbekannte Erfolgswahrscheinlichkeit p mit Wahrscheinlichkeit $1-\alpha$.

Konfidenzintervall für p

Besser (bei kleinen Stichprobenumfängen):
Konfidenzintervalle $[p_L, p_U]$ nach Pearson-Clopper:

$$p_L = \frac{y \cdot f_{\alpha/2}}{n - y + 1 + y \cdot f_{\alpha/2}}, \quad p_U = \frac{(y+1) f_{1-\alpha/2}}{n - y + (y+1) f_{1-\alpha/2}}$$

mit den folgenden Quantilen der F-Verteilung:

$$f_{\alpha/2} = F(2y, 2(n-y+1))_{\alpha/2},$$
$$f_{1-\alpha/2} = F(2(y+1), 2(n-y))_{1-\alpha/2}.$$

Konfidenzintervall für p:

Wie genau sind Wahlumfragen?

Forschungsgruppe Wahlen: $n = 2500$ (Politbarometer).

Allensbach: $n = 1000$

Sonntagsfrage Januar 2013:

Partei	Allensbach	Forschungsgruppe Wahlen	Bundestagswahl 2009
CDU/CSU	39.0	41.0	33.8
SPD	28.0	29.0	23.0
GRÜNE	14.0	13	10.7
FDP	5	4	14.6
DIE LINKE	7	6	11.9
PIRATEN	3	3	2.0
Sonstige	4	4	4.0

Konfidenzintervall für p:

Beispiel: CDU/CSU als große Partei.

Schätzungen 39.0% (Allensbach) bzw. 41.0% (FG Wahlen). Wir berechnen die KI zur Konfidenz 95%.

Auswertung Allensbach-Umfrage:

Mit $z_{0.975} \approx 1.96$ und $n = 1000$ ergibt sich das realisierte KI

$$\left[0.39 - 1.96\sqrt{\frac{0.39(1-0.39)}{1000}}, 0.39 + 1.96\sqrt{\frac{0.39(1-0.39)}{1000}}\right] = [0.3598; 0.4202].$$

Auswertung FG-Wahlen-Umfrage mit $n = 2500$:

$$\left[0.41 - 1.96\sqrt{\frac{0.41(1-0.41)}{2500}}, 0.41 + 1.96\sqrt{\frac{0.41(1-0.41)}{2500}}\right] = [0.3907; 0.4293].$$

Konfidenzintervall für p:

Kleine Parteien: Wir nehmen die Daten der FG Wahlen (größeres n): Auswertung PIRATEN, Schätzung 3%.

Es ergibt sich das realisierte KI

$$\left[0.03 - 1.96\sqrt{\frac{0.03(1-0.03)}{2500}}, 0.03 + 1.96\sqrt{\frac{0.03(1-0.03)}{2500}}\right] = [0.0233; 0.0367]$$

Auswertung FDP: Schätzung 4% :

$$\left[0.04 - 1.96\sqrt{\frac{0.04(1-0.04)}{2500}}, 0.04 + 1.96\sqrt{\frac{0.04(1-0.04)}{2500}}\right] = [0.0323; 0.0477]$$

Aufgabe 37

Textaufgabe:

Eine Fluggesellschaft möchte wissen, wie hoch der Anteil p der Passagiere ist, die ihren Flug nicht antreten. Hierzu soll ein Konfidenzintervall für p bestimmt werden.

Die Überprüfung von 1000 zufällig ausgewählten Passagieren ergibt, dass 74 von ihnen den Flug nicht angetreten haben.

Bestimmen Sie anhand dieses Ergebnisses ein approximatives zweiseitiges Konfidenzintervall für p zum Konfidenzniveau 90%.

Aufgabe 37

Beispiel: Beobachte die Anzahl Y der von einer künstlichen Intelligenz richtig erkannten Testbeispiele unter $n = 30$ Beispielen.

Modell: $Y \sim \text{Bin}(n = 30, p)$

p: wahre Wahrscheinlichkeit, dass der Detektor korrekt erkennt. p ist unbekannt.

Entscheidungproblem:

$p = p_0 = 1/2$: nur so gut wie eine Entscheidung per Münzwurf.

$p = p_1 = 0.9$: Wunschrate korrekter Detektionen.

Wir wollen entscheiden zwischen $p = p_0$ und $p = p_1$.

$\rightarrow$ Zwei Verteilungen (Zähldichten) für Y.

$$\text{bin}(30, 1/2) \text{ oder } \text{bin}(30, 0.9).$$

Beispiel: Erhebe n Messungen $X_1, \ldots, X_n$ der Ozonkonzentration X (in $\mu g/m^3$). Aus langjährigen Voruntersuchungen sei die Standardabweichung $\sigma = 5$ bekannt.

Modell: $X_1, \ldots, X_n \sim N(\mu, 5^2)$

μ: wahre Ozonkonzentration ($\mu = E(X)$), μ unbekannt

Entscheidungsproblem:

$\mu = \mu_0 = 240$: Alarmschwellwert lt. Ozon-Gesetz

$\mu = \mu_1 = 200$: Zielwert der Gemeinde

$\rightarrow$ Zwei Verteilungen (Dichten) für die Daten X:

$$\varphi_{(240,25)}(x) \text{ oder } \varphi_{(200,25)}(x).$$

Testproblem, Nullhypothese, Alternative

Sind f_0 und f_1 zwei mögliche Verteilungen für eine Zufallsvariable X, dann wird das **Testproblem**, zwischen $X \sim f_0$ und $X \sim f_1$ zu entscheiden, in der Form

$$H_0 : f = f_0 \qquad \text{gegen} \qquad H_1 : f = f_1$$

notiert, wobei f die wahre Verteilung von X bezeichnet. H_0 heißt **Nullhypothese** und H_1 **Alternative** (**Alternativhypothese**).

Statistische Testtheorie

Datenmaterial $X_1, \dots, X_n$
Statistik $T = T(X_1, \dots, X_n)$

Statistischer Test

Ein **(statistischer) Test** ist eine Entscheidungsregel, die basierend auf T entweder zugunsten von H_0 (Notation: „H_0“) oder zugunsten von H_1 („H_1“) entscheidet.

Fehler 1. und 2. Art

Entscheidung für H_1, obwohl H_0 richtig ist, heißt **Fehler 1. Art**. H_0 wird dann fälschlicherweise verworfen. Eine Entscheidung für H_0, obwohl H_1 richtig ist, heißt **Fehler 2. Art**. H_0 wird fälschlicherweise akzeptiert.

Insgesamt sind vier Konstellationen möglich, die in der folgenden Tabelle zusammengefasst sind:

	H_0	H_1
„H_0“	$\surd$	Fehler 2. Art
„H_1“	Fehler 1. Art	$\surd$

Signifikanzniveau, Test zum Niveau α

Bezeichnet „H_1“ eine Annahme der Alternative und „H_0“ eine Annahme der Nullhypothese durch eine Entscheidungsregel, dann ist durch diese Regel ein **statistischer Test zum Signifikanzniveau (Niveau)** α gegeben, wenn

$$P_{H_0}(„H_1“) \leq \alpha \, .$$

Genauer ist die linke Seite ist das tatsächliche Signifikanzniveau des Tests und die rechte Seite das vorgegebene **nominale** Signifikanzniveau.

Hinweis: Die Wahrscheinlichkeit eines Fehlers 2. Art wird nicht unbedingt kontrolliert. Dies erfordert eine Planung der Stichprobengröße.

Schärfe (Power)

Die Wahrscheinlichkeit eines Fehlers 2. Art wird üblicherweise mit β bezeichnet. Die Gegenwahrscheinlichkeit,

$$1 - \beta = P_{H_1}(„H_1“)(= E_{H_1}(1 - \phi)),$$

dass der Test die Alternative H_1 tatsächlich aufdeckt, heißt **Schärfe (Power)** des Testverfahrens.

Entscheidungskonstellationen und die Wahrscheinlichkeiten:

	H_0	H_1
„H_0“	$\surd$ $1 - \alpha$	Fehler 2. Art β
„H_1“	Fehler 1. Art α	$\surd$ $1 - \beta$: Schärfe (Power)

Frage: Wie sollen die Hypothesen H_0 und H_1 zugeordnet werden?

Vorgehen 1: Risikoüberlegung

$\rightarrow$ Ein Signifikanztest kontrolliert stets den Fehler 1. Art, aber nicht unbedingt den Fehler 2. Art.

- Entscheide für das vorliegende Problem, welcher Fehler schlimmer ist und auf jeden Fall kontrolliert werden soll.
- Stelle Hypothesen so auf, dass der Fehler 1. Art der schlimmere ist.

Vorgehen 2: Nachweisformulierung

• Sehr oft stellt eine der Hypothesen das etablierte Wissen (Stand der Technik) da und die andere Hypothese den vermuteten neuen, besonderen Effekt.

• Der Effekt kann z.B. ein Überschreiten eines Grenzwerts, ein Unterschreiten einer Zielvorgabe des Managements, die Wirksamkeit eines neuen Wirkstoffs für ein Medikament oder die Überlegenheit einer künstlichen Intelligenz für Problem X sein.

→ Wer den Effekt nachweisen will, muss die Gegenseite überzeugen.

→ Die Anhänger des etablierten Wissens werden sich nur dann von dem Effekt überzeugen lassen und ihre Meinung ändern, wenn die Wahrscheinlichkeit einer Fehlentscheidung zu Gunsten des Effekts (sehr) klein ist!

Vorgehen 2: Nachweisformulierung

Ansatz: Test kontrolliert $P_{H_0}(„H_1“) \leq \alpha$.

Lege H_0 und H_1 so fest, dass gilt:

$$P(\text{„Entscheidung für Effekt, obwohl kein Effekt existiert.“}) = P_{H_0}(„H_1“)$$

Die Hypothese, die den Effekt beschreibt wird die Alternativhypothese!
Die Hypothese, die das etablierte Wissen (kein Effekt) beschreibt, wird die Nullhypothese!

In dieser Formulierung wird die Fehlerwahrscheinlichkeit durch α kontrolliert, fälschlicherweise von dem Vorliegen des Effekts auszugehen.

Verallgemeinerung: s. Buch

$$H_0 : \vartheta \in \Theta_0 \qquad \text{versus} \qquad H_1 : \vartheta \in \Theta_1,$$

wobei $\Theta_0 \cup \Theta_1 = \Theta$ eine disjunkte Zerlegung des Parameterraums Θ ist.

Beispiel: Modell $N(\mu, 25)$, $\mu \in \mathbb{R}$. Gilt $\mu = 200$ oder $\mu = 240$?

$\Theta = \{200, 240\}$.
$\Theta_0 = \{200\} \leftrightarrow H_0 : \mu = 200$
$\Theta_1 = \{240\} \leftrightarrow H_1 : \mu = 240$.

Hypothesen

Meist möchte man aber nicht nur zwei Verteilungen gegeneinander testen, sondern z.B. $\mu \leq 240$ (Grenzwert eingehalten) testen gegen $\mu > 240$ (Alarmwert überschritten).

Hypothesen (über den Erwartungswert μ)

Einseitiges Testproblem:

$$H_0 : \mu \leq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0,$$

bzw.

$$H_0 : \mu \geq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu < \mu_0.$$

Zweiseitiges Testproblem:

$$H_0 : \mu = \mu_0 \qquad \text{gegen} \qquad H_1 : \mu \neq \mu_0.$$

WICHTIG: Der Grenzfall $'='$ wird immer H_0 zugeschlagen.

Der Gauß-Test

Gegeben: $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$ mit *bekannter* Varianz $\sigma^2 \in (0, \infty)$

Teststatistik: $T = \sqrt{n}\,\dfrac{\overline{X}_n - \mu_0}{\sigma}$ ($\mu_0 \in \mathbb{R}$ vorgegebener Sollwert)

Verteilung der Teststatistik: $T \sim \mathrm{N}(0, 1)$ für $\mu = \mu_0$

(In der Teststatistik wird $\overline{X}_n$ mit μ_0 verglichen, dem am schwersten von H_1 zu unterscheidendem Fall.)

Der Gauß-Test

Einseitiger Gauß-Test (1)

Der einseitige Gaußtest verwirft die Nullhypothese $H_0 : \mu \leq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu > \mu_0$, wenn $T > z_{1-\alpha}$.

Einseitiger Gauß-Test (2)

Der einseitige Gaußtest verwirft $H_0 : \mu \geq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu < \mu_0$, wenn $T < -z_{1-\alpha} = z_\alpha$.

Zweiseitiger Gauß-Test

Der zweiseitige Gauß-Test verwirft die Nullhypothese $H_0 : \mu = \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu \neq \mu_0$, wenn $|T| > z_{1-\alpha/2}$.

(Hierbei bezeichnet z_p das p-Quantil zu $N(0,1)$ für $p \in (0,1)$.)