Schließende Statistik

Vorlesung an der Universität des Saarlandes

Dr. Martin Becker

Wintersemester 2015/16

Organisatorisches I

- Vorlesung: Freitag, 12-14 Uhr, Gebäude B4 1, Audimax
- Übungen: siehe Homepage, Beginn: ab Montag (26.10.)
- Prüfung: 2-stündige Klausur nach Semesterende (1. Prüfungszeitraum)

Wichtig:

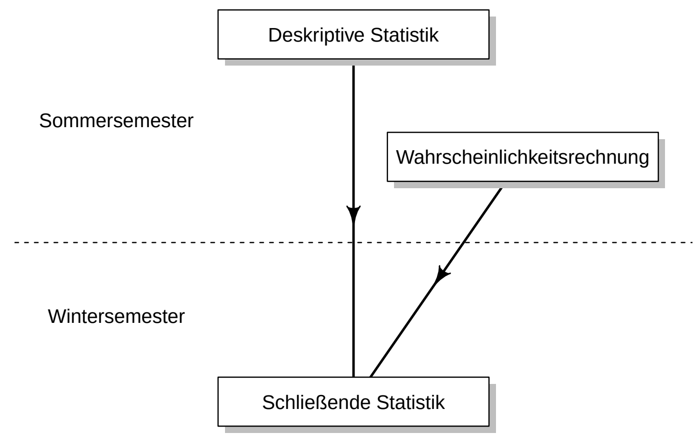
Anmeldung (ViPa) vom 10.-25. November (bis 15 Uhr) möglich
Abmeldung bis 28. Januar 2016 (12 Uhr) möglich

- Hilfsmittel für Klausur
  - „Moderat“ programmierbarer Taschenrechner, auch mit Grafikfähigkeit
  - 2 *beliebig gestaltete* DIN A 4–Blätter (bzw. 4, falls nur einseitig)
  - Benötigte Tabellen werden gestellt, aber **keine weitere Formelsammlung!**
- Durchgefallen — was dann?
  - „Wiederholungskurs“ im kommenden (Sommer-)Semester
  - „Nachprüfung“ (voraussichtlich) erst September/Oktober 2016 (2. Prüfungszeitraum)
  - „Reguläre“ Vorlesung/Übungen wieder im Wintersemester 2016/17

Organisatorisches II

- Kontakt: Dr. Martin Becker
  Geb. C3 1, 2. OG, Zi. 2.17
  e-Mail: `martin.becker@mx.uni-saarland.de`
- Sprechstunde nach Vereinbarung (Terminabstimmung per e-Mail)
- Informationen und Materialien auf Homepage:
  `http://www.lehrstab-statistik.de`
- Material zu dieser Veranstaltung: Vorlesungsfolien *i.d.R. vor Vorlesung* zum Download (inklusive Drucker-freundlicher 2-auf-1 bzw. 4-auf-1 Versionen)
- Wie in „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“:
  - Neben theoretischer Einführung der Konzepte auch einige Beispiele auf Vorlesungsfolien
  - Einige wichtige Grundlagen werden gesondert als „Definition“, „Satz“ oder „Bemerkung“ hervorgehoben
  - **Aber:** Auch vieles, was nicht formal als „Definition“, „Satz“ oder „Bemerkung“ gekennzeichnet ist, ist wichtig!
- Übungsblätter i.d.R. nach Vorlesung zum Download
- Ergebnisse *(keine Musterlösungen!)* zu einigen Aufgaben im Netz
- Ausführlichere Lösungen zu den Übungsaufgaben (nur) in Übungsgruppen!

Organisation der Statistik-Veranstaltungen



Benötigte Konzepte

aus den mathematischen Grundlagen

- Rechnen mit Potenzen

$$a^m \cdot b^m = (a \cdot b)^m \qquad a^m \cdot a^n = a^{m+n} \qquad \frac{a^m}{a^n} = a^{m-n} \qquad (a^m)^n = a^{m \cdot n}$$

- Rechnen mit Logarithmen

$$\ln(a \cdot b) = \ln a + \ln b \qquad \ln\left(\frac{a}{b}\right) = \ln a - \ln b \qquad \ln(a^r) = r \cdot \ln a$$

- Rechenregeln auch mit Summen-/Produktzeichen, z.B.

$$\ln\left(\prod_{i=1}^{n} x_i^{r_i}\right) = \sum_{i=1}^{n} r_i \ln(x_i)$$

- Maximieren differenzierbarer Funktionen
  - Funktionen (ggf. partiell) ableiten
  - Nullsetzen von Funktionen (bzw. deren Ableitungen)
- „Unfallfreies" Rechnen mit 4 Grundrechenarten und Brüchen...

Benötigte Konzepte

aus Veranstaltung „Deskriptive Statistik und Wahrscheinlichkeitsrechnung"

- Diskrete und stetige Zufallsvariablen X, Verteilungsfunktionen, Wahrscheinlichkeitsverteilungen, ggf. Dichtefunktionen
- Momente (Erwartungswert $\mathrm{E}(X)$, Varianz $\mathrm{Var}(X)$, höhere Momente $\mathrm{E}(X^k)$)
- „Einbettung" der deskriptiven Statistik in die Wahrscheinlichkeitsrechnung
  - Ist Ω die (endliche) Menge von Merkmalsträgern einer deskriptiven statistischen Untersuchung, $\mathcal{F} = \mathcal{P}(\Omega)$ und P die Laplace-Wahrscheinlichkeit

    $$P : \mathcal{P}(\Omega) \to \mathbb{R}; B \mapsto \frac{\#B}{\#\Omega} ,$$

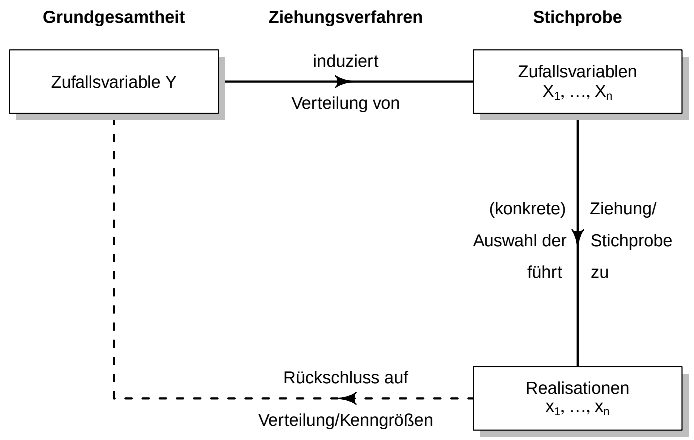
    so kann jedes numerische Merkmal X als Zufallsvariable $X : \Omega \to \mathbb{R}$ verstanden werden.
  - Der Träger von X entspricht dann dem Merkmalsraum $A = \{a_1, \dots, a_m\}$, die Punktwahrscheinlichkeiten den relativen Häufigkeiten, d.h. es gilt $p(a_j) = r(a_j)$ bzw. — äquivalent — $P_X(\{a_j\}) = r(a_j)$ für $j \in \{1, \dots, m\}$.
- Verteilung von $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ für unabhängig identisch verteilte X_i
  - falls X_i normalverteilt
  - falls $n \to \infty$ (Zentraler Grenzwertsatz!)

Grundidee der schließenden Statistik

- Ziel der schließenden Statistik/induktiven Statistik:

  Ziehen von Rückschlüssen auf die Verteilung einer (größeren) Grundgesamtheit auf Grundlage der Beobachtung einer (kleineren) Stichprobe.
- Rückschlüsse auf die Verteilung können sich auch beschränken auf spezielle Eigenschaften/Kennzahlen der Verteilung, z.B. den Erwartungswert.
- „Fundament": **Drei Grundannahmen**
  1. Der interessierende Umweltausschnitt kann durch eine (ein- oder mehrdimensionale) Zufallsvariable Y beschrieben werden.
  2. Man kann eine *Menge* W von Wahrscheinlichkeitsverteilungen angeben, zu der die *unbekannte* wahre Verteilung von Y gehört.
  3. Man beobachtet Realisationen $x_1, \dots, x_n$ von (Stichproben-)Zufallsvariablen $X_1, \dots, X_n$, deren gemeinsame Verteilung *in vollständig bekannter Weise* von der Verteilung von Y abhängt.
- Ziel ist es also, aus der Beobachtung der n Werte $x_1, \dots, x_n$ mit Hilfe des bekannten Zusammenhangs zwischen den Verteilungen von $X_1, \dots, X_n$ und Y Aussagen über die Verteilung von Y zu treffen.

„Veranschaulichung" der schließenden Statistik



Bemerkungen zu den 3 Grundannahmen

- Die 1. Grundannahme umfasst insbesondere die Situation, in der die Zufallsvariable Y einem (ein- oder mehrdimensionalen) Merkmal auf einer *endlichen* Menge von Merkmalsträgern entspricht, vgl. die Einbettung der deskriptiven Statistik in die Wahrscheinlichkeitsrechnung auf Folie 6.
  In diesem Fall interessiert man sich häufig für Kennzahlen von Y, z.B. den Erwartungswert von Y (als Mittelwert des Merkmals auf der Grundgesamtheit).
- Die Menge W von Verteilungen aus der 2. Grundannahme ist häufig eine *parametrische* Verteilungsfamilie, zum Beispiel die Menge aller Exponentialverteilungen oder die Menge aller Normalverteilungen mit Varianz $\sigma^2 = 2^2$.
  In diesem Fall ist die Menge der für die Verteilung von Y denkbaren Parameter interessant (später mehr!). Wir betrachten dann nur solche Verteilungsfamilien, in denen verschiedene Parameter auch zu verschiedenen Verteilungen führen („Parameter sind *identifizierbar*.“).
- Wir beschränken uns auf *sehr* einfache Zusammenhänge zwischen der Verteilung der interessierenden Zufallsvariablen Y und der Verteilung der Zufallsvariablen $X_1, \ldots, X_n$.

Beispiel I

Stichprobe aus endlicher Grundgesamtheit Ω

- Grundgesamtheit: $N = 4$ Kinder (**A**nna, **B**eatrice, **C**hristian, **D**aniel) gleichen Alters, die in derselben Straße wohnen: $\Omega = \{A, B, C, D\}$
- Interessierender Umweltausschnitt: monatliches Taschengeld Y (in €) bzw. später spezieller: Mittelwert des monatlichen Taschengelds der 4 Kinder (entspricht $\mathrm{E}(Y)$ bei Einbettung wie beschrieben)
- (Verteilungsannahme:) Verteilung von Y unbekannt, aber sicher in der Menge der diskreten Verteilungen mit maximal $N = 4$ (nichtnegativen) Trägerpunkten und Punktwahrscheinlichkeiten, die Vielfaches von $1/N = 1/4$ sind.
  Im Beispiel nun: Zufallsvariable Y nehme Werte

ω	A	B	C	D
$Y(\omega)$	15	20	25	20

an, habe also folgende zugehörige Verteilung:

y_i	15	20	25	Σ
$p_Y(y_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Beispiel II

Stichprobe aus endlicher Grundgesamtheit Ω

- **Beachte:** Verteilung von Y nur im Beispiel bekannt, in der Praxis: Verteilung von Y natürlich unbekannt!
- Einfachste Möglichkeit, um Verteilung von Y bzw. deren Erwartungswert zu ermitteln: alle 4 Kinder nach Taschengeld befragen!
- Typische Situation in schließender Statistik: nicht alle Kinder können befragt werden, sondern nur eine kleinere Anzahl $n < N = 4$, beispielsweise $n = 2$.
  Erwartungswert von Y (mittleres Taschengeld aller 4 Kinder) kann dann nur noch **geschätzt** werden!
- Ziel: Rückschluss aus der Erhebung von $n = 2$ Taschengeldhöhen auf die größere Grundgesamtheit von $N = 4$ Kindern durch
  - Schätzung des mittleren Taschengeldes aller 4 Kinder
  - Beurteilung der Qualität der Schätzung (mit welchem „Fehler“ ist zu rechnen)
- (Qualität der) Schätzung hängt ganz entscheidend vom Ziehungs-/Auswahlverfahren ab!

Beispiel III

Stichprobe aus endlicher Grundgesamtheit Ω

- Erhebung von 2 Taschengeldhöhen führt zu Stichprobenzufallsvariablen X_1 und X_2.
- X_1 bzw. X_2 entsprechen in diesem Fall dem Taschengeld des 1. bzw. 2. befragten Kindes
- **Sehr wichtig** für Verständnis:
  X_1 und X_2 sind Zufallsvariablen, da ihr Wert (Realisation) davon abhängt, **welche Kinder** man zufällig ausgewählt hat!
- Erst **nach Auswahl** der Kinder (also nach „Ziehung der Stichprobe“) steht der Wert (die Realisation) x_1 von X_1 bzw. x_2 von X_2 fest!

Variante A

- Naheliegendes Auswahlverfahren: nacheinander **rein zufällige** Auswahl von 2 der 4 Kinder, d.h. **zufälliges Ziehen ohne Zurücklegen mit Berücksichtigung der Reihenfolge**
- Alle $(4)_2 = 12$ Paare (A, B); (A, C); (A, D); (B, A); (B, C); (B, D); (C, A); (C, B); (C, D); (D, A); (D, B); (D, C) treten dann mit der gleichen Wahrscheinlichkeit (1/12) auf und führen zu den folgenden „Stichprobenrealisationen“ (x_1, x_2) der Stichprobenvariablen (X_1, X_2):

Beispiel IV

Stichprobe aus endlicher Grundgesamtheit Ω

- Realisationen (x_1, x_2) zur Auswahl von 1. Kind (Zeilen)/2. Kind (Spalten):

	A	B	C	D
A	unmöglich	(15,20)	(15,25)	(15,20)
B	(20,15)	unmöglich	(20,25)	(20,20)
C	(25,15)	(25,20)	unmöglich	(25,20)
D	(20,15)	(20,20)	(20,25)	unmöglich

- Resultierende gemeinsame Verteilung von (X_1, X_2):

$x_1 \backslash x_2$	15	20	25	Σ
15	0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$
20	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
25	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- Es fällt auf (**Variante A**):
  - X_1 und X_2 haben die gleiche Verteilung wie Y.
  - X_1 und X_2 sind **nicht** stochastisch unabhängig.

Beispiel V

Stichprobe aus endlicher Grundgesamtheit Ω

- Naheliegend: Schätzung des Erwartungswertes $E(Y)$, also des mittleren Taschengeldes aller 4 Kinder, durch den (arithmetischen) Mittelwert der erhaltenen Werte für die 2 befragten Kinder.
- **Wichtig: Nach** Auswahl der Kinder ist dieser Mittelwert eine Zahl, es ist aber sehr nützlich, den Mittelwert schon **vor** Auswahl der Kinder (dann) als Zufallsvariable (der Zufall kommt über die zufällige Auswahl der Kinder ins Spiel) zu betrachten!
- Interessant ist also die Verteilung der **Zufallsvariable** $\overline{X} := \frac{1}{2}(X_1 + X_2)$, also des Mittelwerts der Stichprobenzufallsvariablen X_1 und X_2.
  Die (hiervon zu unterscheidende!) **Realisation** $\overline{x} = \frac{1}{2}(x_1 + x_2)$ ergibt sich erst (als Zahlenwert) nach Auswahl der Kinder (wenn die Realisation (x_1, x_2) von (X_1, X_2) vorliegt)!
- Verteilung von $\overline{X}$ hier (**Variante A**):

$\overline{x}_i$	17.5	20	22.5	Σ
$p_{\overline{X}}(\overline{x}_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Beispiel VI

Stichprobe aus endlicher Grundgesamtheit Ω

Variante B

- Weiteres mögliches Auswahlverfahren: 2-fache **rein zufällige** und **voneinander unabhängige** Auswahl eines der 4 Kinder, wobei erlaubt ist, dasselbe Kind mehrfach auszuwählen, d.h. **zufälliges Ziehen mit Zurücklegen und Berücksichtigung der Reihenfolge**
- Alle $4^2 = 16$ Paare (A, A); (A, B); (A, C); (A, D); (B, A); (B, B); (B, C); (B, D); (C, A); (C, B); (C, C); (C, D); (D, A); (D, B); (D, C); (D, D) treten dann mit der gleichen Wahrscheinlichkeit (1/16) auf und führen zu den folgenden „Stichprobenrealisationen" (x_1, x_2) der Stichprobenvariablen (X_1, X_2) (zur Auswahl von 1. Kind (Zeilen)/2. Kind (Spalten)):

	A	B	C	D
A	(15,15)	(15,20)	(15,25)	(15,20)
B	(20,15)	(20,20)	(20,25)	(20,20)
C	(25,15)	(25,20)	(25,25)	(25,20)
D	(20,15)	(20,20)	(20,25)	(20,20)

Beispiel VII

Stichprobe aus endlicher Grundgesamtheit Ω

- Resultierende gemeinsame Verteilung von (X_1, X_2):

$x_1 \backslash x_2$	15	20	25	Σ
15	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
20	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
25	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

- Es fällt auf (**Variante B**):
  - X_1 und X_2 haben die gleiche Verteilung wie Y.
  - X_1 und X_2 **sind** stochastisch unabhängig.
- Verteilung von $\overline{X}$ hier (**Variante B**):

$\overline{x}_i$	15	17.5	20	22.5	25	Σ
$p_{\overline{X}}(\overline{x}_i)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	1

Zufallsstichprobe

- Beide Varianten zur Auswahl der Stichprobe führen dazu, dass alle Stichprobenzufallsvariablen X_i $(i = 1, 2)$ **identisch** verteilt sind wie Y.
- Variante **B** führt außerdem dazu, dass die Stichprobenzufallsvariablen X_i $(i = 1, 2)$ **stochastisch unabhängig** sind.

Definition 2.1 ((Einfache) Zufallsstichprobe)

Seien $n \in \mathbb{N}$ und $X_1, \dots, X_n$ Zufallsvariablen einer Stichprobe vom Umfang n zu Y. Dann heißt $(X_1, \dots, X_n)$

- **Zufallsstichprobe** vom Umfang n zu Y, falls die Verteilungen von Y und X_i für alle $i \in \{1, \dots, n\}$ übereinstimmen, alle X_i also identisch verteilt sind wie Y,
- **einfache (Zufalls-)Stichprobe** vom Umfang n zu Y, falls die Verteilungen von Y und X_i für alle $i \in \{1, \dots, n\}$ übereinstimmen und $X_1, \dots, X_n$ außerdem stochastisch unabhängig sind.

- (X_1, X_2) ist in Variante A des Beispiels also eine Zufallsstichprobe vom Umfang 2 zu Y, in Variante B sogar eine einfache (Zufalls-)Stichprobe vom Umfang 2 zu Y.

- $X_1, \dots, X_n$ ist also nach Definition 2.1 auf Folie 17 genau dann eine **Zufallsstichprobe**, falls für die Verteilungsfunktionen zu $Y, X_1, \dots, X_n$
$$F_Y = F_{X_1} = \dots = F_{X_n}$$
gilt.
- Ist $X_1, \dots, X_n$ eine **einfache Stichprobe** vom Umfang n zu Y, so gilt für die *gemeinsame* Verteilungsfunktion von $(X_1, \dots, X_n)$ sogar
$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_Y(x_1) \cdot \ldots \cdot F_Y(x_n) = \prod_{i=1}^{n} F_Y(x_i) \; .$$
Ist Y diskrete Zufallsvariable gilt also insbesondere für die beteiligten Wahrscheinlichkeitsfunktionen
$$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = p_Y(x_1) \cdot \ldots \cdot p_Y(x_n) = \prod_{i=1}^{n} p_Y(x_i) \; ,$$
ist Y stetige Zufallsvariable, so existieren Dichtefunktionen von Y bzw. $(X_1, \dots, X_n)$ mit
$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_Y(x_1) \cdot \ldots \cdot f_Y(x_n) = \prod_{i=1}^{n} f_Y(x_i) \; .$$

Stichprobenrealisation/Stichprobenraum

Definition 2.2 (Stichprobenrealisation/Stichprobenraum)

Seien $n \in \mathbb{N}$ und $X_1, \dots, X_n$ Zufallsvariablen einer Stichprobe vom Umfang n zu Y. Seien $x_1, \dots, x_n$ die beobachteten Realisationen zu den Zufallsvariablen $X_1, \dots, X_n$. Dann heißt

- $(x_1, \dots, x_n)$ **Stichprobenrealisation** und
- die Menge $\mathcal{X}$ aller möglichen Stichprobenrealisationen **Stichprobenraum**.

- Es gilt offensichtlich immer $\mathcal{X} \subseteq \mathbb{R}^n$.
- „Alle möglichen Stichprobenrealisationen" meint alle Stichprobenrealisationen, die für *irgendeine* der möglichen Verteilungen W von Y aus der Verteilungsannahme möglich sind.
- Wenn man davon ausgeht, dass ein Kind „schlimmstenfalls" 0 € Taschengeld erhält, wäre im Beispiel also $\mathcal{X} = \mathbb{R}_+^2$ (Erinnerung: $\mathbb{R}_+ := \{x \in \mathbb{R} \,|\, x \geq 0\}$).
- Meist wird die Information der Stichprobenzufallsvariablen bzw. der Stichprobenrealisation weiter mit sog. „Stichprobenfunktionen" aggregiert, die oft (große) Ähnlichkeit mit Funktionen haben, die in der deskriptiven Statistik zur Aggregierung von Urlisten eingesetzt werden.

Stichprobenfunktion/Statistik

Definition 2.3 (Stichprobenfunktion/Statistik)

Seien $n \in \mathbb{N}$ und $X_1, \dots, X_n$ Zufallsvariablen einer Stichprobe vom Umfang n zu Y mit Stichprobenraum $\mathcal{X}$. Dann heißt eine Abbildung
$$T : \mathcal{X} \to \mathbb{R}; (x_1, \dots, x_n) \mapsto T(x_1, \dots, x_n)$$
Stichprobenfunktion oder **Statistik**.

- Stichprobenfunktionen sind also Abbildungen, deren Wert mit Hilfe der Stichprobenrealisation bestimmt werden kann.
- Stichprobenfunktionen müssen (geeignet, z.B. $\mathcal{B}^n$-$\mathcal{B}$-) messbare Abbildungen sein; diese Anforderung ist aber für alle hier interessierenden Funktionen erfüllt, Messbarkeitsüberlegungen bleiben also im weiteren Verlauf außen vor.
- Ebenfalls als Stichprobenfunktion bezeichnet wird die (als Hintereinanderausführung zu verstehende) Abbildung $T(X_1, \dots, X_n)$, wegen der Messbarkeitseigenschaft ist dies immer eine **Zufallsvariable**.
Die Untersuchung der zugehörigen Verteilung ist für viele Anwendungen von **ganz wesentlicher** Bedeutung.

- Wenn man sowohl die Zufallsvariable $T(X_1, \ldots, X_n)$ als auch den aus einer vorliegenden Stichprobenrealisation $(x_1, \ldots, x_n)$ resultierenden Wert $T(x_1, \ldots, x_n)$ betrachtet, so bezeichnet man $T(x_1, \ldots, x_n)$ oft auch als **Realisation** der Stichprobenfunktion.
- Im Taschengeld-Beispiel war die betrachtete Stichprobenfunktion das arithmetische Mittel, also konkreter

$$T : \mathbb{R}^2 \to \mathbb{R};\, T(x_1, x_2) = \overline{x} := \frac{1}{2}(x_1 + x_2)$$

bzw. — als Zufallsvariable betrachtet —

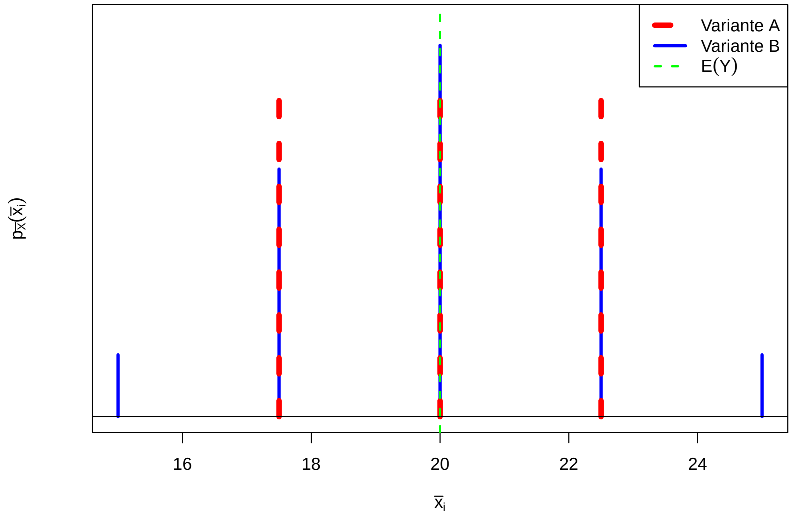
$$T(X_1, X_2) = \overline{X} := \frac{1}{2}(X_1 + X_2) \ .$$

- Je nach Anwendung erhalten Stichprobenfunktionen auch speziellere Bezeichnungen, z. B.
  - **Schätzfunktion** oder **Schätzer**, wenn die Stichprobenfunktion zur Schätzung eines Verteilungsparameters oder einer Verteilungskennzahl verwendet wird (wie im Beispiel!),
  - **Teststatistik**, wenn auf Grundlage der Stichprobenfunktion Entscheidungen über die Ablehnung oder Annahme von Hypothesen über die Verteilung von Y getroffen werden.

Beispiel VIII

Stichprobe aus endlicher Grundgesamtheit Ω

Vergleich der Verteilungen von $\overline{X}$ in beiden Varianten:



Beispiel IX

Stichprobe aus endlicher Grundgesamtheit Ω

- Verteilung von Y

y_i	15	20	25	Σ
$p_Y(y_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

hat Erwartungswert $E(Y) = 20$ und Varianz $\mathrm{Var}(Y) = 12.5$.

- Verteilung von $\overline{X}$ (Variante **A**):

$\overline{x}_i$	17.5	20	22.5	Σ
$p_{\overline{X}}(\overline{x}_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

hat Erwartungswert $E(\overline{X}) = 20$ und Varianz $\mathrm{Var}(\overline{X}) = 4.1\overline{6}$.

- Verteilung von $\overline{X}$ (Variante **B**):

$\overline{x}_i$	15	17.5	20	22.5	25	Σ
$p_{\overline{X}}(\overline{x}_i)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	1

hat Erwartungswert $E(\overline{X}) = 20$ und Varianz $\mathrm{Var}(\overline{X}) = 6.25$.

Beispiel X

Stichprobe aus endlicher Grundgesamtheit Ω

- In beiden Varianten schätzt man das mittlere Taschengeld $E(Y) = 20$ also „im Mittel“ richtig, denn es gilt für beide Varianten $E(\overline{X}) = 20 = E(Y)$.
- Die Varianz von $\overline{X}$ ist in Variante A kleiner als in Variante B; zusammen mit der Erkenntnis, dass beide Varianten „im Mittel“ richtig liegen, schätzt also Variante A „genauer“.
- In beiden Varianten hängt es vom Zufall (genauer von der konkreten Auswahl der beiden Kinder — bzw. in Variante B möglicherweise zweimal desselben Kindes — ab), ob man *nach Durchführung der Stichprobenziehung* den tatsächlichen Mittelwert als Schätzwert erhält oder nicht.
- Obwohl $\overline{X}$ in Variante A die kleinere Varianz hat, erhält man in Variante B den tatsächlichen Mittelwert $E(Y) = 20$ mit einer größeren Wahrscheinlichkeit (3/8 in Variante B gegenüber 1/3 in Variante A).

Parameterpunktschätzer

- Im Folgenden: Systematische Betrachtung der Schätzung von Verteilungsparametern, wenn die Menge W der (möglichen) Verteilungen von Y eine **parametrische** Verteilungsfamilie gemäß folgender Definition ist: *(Z.T. Wdh. aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“)*

Definition 3.1 (Parametrische Verteilungsfamilie, Parameterraum)

1. Eine Menge von Verteilungen W heißt **parametrische Verteilungsfamilie**, wenn jede Verteilung in W durch einen endlich-dimensionalen Parameter $\theta = (\theta_1, \ldots, \theta_K) \in \Theta \subseteq \mathbb{R}^K$ charakterisiert wird.
   Um die Abhängigkeit von θ auszudrücken, notiert man die Verteilungen, Verteilungsfunktionen sowie die Wahrscheinlichkeits- bzw. Dichtefunktionen häufig als
   $$P(\,\cdot\,|\theta_1, \ldots, \theta_K),\; F(\,\cdot\,|\theta_1, \ldots, \theta_K) \text{ sowie } p(\,\cdot\,|\theta_1, \ldots, \theta_K) \text{ bzw. } f(\,\cdot\,|\theta_1, \ldots, \theta_K)\,.$$
2. Ist W die Menge von Verteilungen aus der 2. Grundannahme („Verteilungsannahme“), so bezeichnet man W auch als **parametrische Verteilungsannahme**. Die Menge Θ heißt dann auch **Parameterraum**.

Bemerkungen

- Wir betrachten nur „identifizierbare“ parametrische Verteilungsfamilien, das heißt, unterschiedliche Parameter aus dem Parameterraum Θ müssen auch zu unterschiedlichen Verteilungen aus W führen.
- Die Bezeichnung θ dient lediglich zur vereinheitlichten Notation. In der Praxis behalten die Parameter meist ihre ursprüngliche Bezeichnung.
- In der Regel gehören alle Verteilungen in W zum gleichen Typ, zum Beispiel als
  - Bernouilliverteilung $B(1, p)$: Parameter $p \equiv \theta$, Parameterraum $\Theta = [0, 1]$
  - Poissonverteilung $\text{Pois}(\lambda)$: Parameter $\lambda \equiv \theta$, Parameterraum $\Theta = \mathbb{R}_{++}$
  - Exponentialverteilung $\text{Exp}(\lambda)$: Parameter $\lambda \equiv \theta$, Parameterraum $\Theta = \mathbb{R}_{++}$
  - Normalverteilung $N(\mu, \sigma^2)$: Parameter**vektor** $(\mu, \sigma^2) \equiv (\theta_1, \theta_2)$, Parameterraum $\mathbb{R} \times \mathbb{R}_{++}$

  (mit $\mathbb{R}_{++} := \{x \in \mathbb{R} \,|\, x > 0\}$).
- Suche nach **allgemein anwendbaren** Methoden zur Konstruktion von Schätzfunktionen für unbekannte Parameter θ aus parametrischen Verteilungsannahmen.
- Schätzfunktionen für einen Parameter(vektor) θ *sowie deren Realisationen (!)* werden üblicherweise mit $\widehat{\theta}$, gelegentlich auch mit $\widetilde{\theta}$ bezeichnet.
- Meist wird vom Vorliegen einer einfachen Stichprobe ausgegangen.

Methode der Momente (Momentenmethode)

- Im Taschengeldbeispiel: Schätzung des Erwartungswerts $\text{E}(Y)$ *naheliegenderweise* durch das arithmetische Mittel $\overline{X} = \frac{1}{2}(X_1 + X_2)$.
- Dies entspricht der Schätzung des 1. (theoretischen) Moments von Y durch das 1. empirische Moment der Stichprobenrealisation (aufgefasst als Urliste im Sinne der deskriptiven Statistik).
- Gleichsetzen von theoretischen und empirischen Momenten bzw. Ersetzen theoretischer durch empirische Momente führt zur gebräuchlichen **(Schätz-)Methode der Momente** für die Parameter von parametrischen Verteilungsfamilien.
- Grundlegende Idee: Schätze Parameter der Verteilung so, dass zugehörige theoretische Momente $\text{E}(Y)$, $\text{E}(Y^2)$, ... mit den entsprechenden empirischen Momenten $\overline{X}$, $\overline{X^2}$, ... der Stichprobenzufallsvariablen $X_1, \ldots, X_n$ (bzw. deren Realisationen) übereinstimmen.
- Es werden dabei (beginnend mit dem ersten Moment) gerade so viele Momente einbezogen, dass das entstehende Gleichungssystem für die Parameter eine eindeutige Lösung hat.
  Bei eindimensionalen Parameterräumen genügt *i.d.R.* das erste Moment.

Momente von Zufallsvariablen

- Bereits aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“ bekannt ist die folgende Definition für die (theoretischen) Momente von Zufallsvariablen:

Definition 3.2 (k-te Momente)

Es seien Y eine (eindimensionale) Zufallsvariable, $k \in \mathbb{N}$.
Man bezeichnet den Erwartungswert $\text{E}(Y^k)$ (falls er existiert) als das **(theoretische) Moment k-ter Ordnung** von Y, oder auch das **k-te (theoretische) Moment** von Y und schreibt auch kürzer

$$\text{E}\,Y^k := \text{E}(Y^k).$$

- Erinnerung (unter Auslassung der Existenzbetrachtung!):
  Das k-te Moment von Y berechnet man für diskrete bzw. stetige Zufallsvariablen Y durch
  $$\text{E}(Y^k) = \sum_{y_i} y_i^k \cdot p_Y(y_i) \qquad \text{bzw.} \qquad \text{E}(Y^k) = \int_{-\infty}^{\infty} y^k \cdot f_Y(y)\,dy\,,$$
  wobei y_i (im diskreten Fall) alle Trägerpunkte von Y durchläuft.

Empirische Momente von Stichproben

- Analog zu empirischen Momenten von Urlisten in der deskriptiven Statistik definiert man empirische Momente von Stichproben in der schließenden Statistik wie folgt:

Definition 3.3 (empirische Momente)

Ist $(X_1, \ldots, X_n)$ eine (einfache) Zufallsstichprobe zu einer Zufallsvariablen Y, so heißt

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

das **empirische k-te Moment**, oder auch das **Stichprobenmoment der Ordnung** k. Zu einer Realisation $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ bezeichnet

$$\overline{x^k} := \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

entsprechend die zugehörige **Realisation** des k-ten empirischen Moments.

Durchführung der Momentenmethode

- Zur Durchführung der Momentenmethode benötigte Anzahl von Momenten meist gleich der Anzahl der zu schätzenden Verteilungsparameter.
- Mehrere Möglichkeiten zur Durchführung der Momentenmethode.
- 1. Möglichkeit:
  - Ausdrücken/Berechnen der theoretischen Momente in Abhängigkeit der Verteilungsparameter
  - Gleichsetzen der theoretischen Momente mit den entsprechenden empirischen Momenten und Auflösen der entstehenden Gleichungen nach den Verteilungsparametern.
- 2. Möglichkeit:
  - Ausdrücken der Verteilungsparameter in Abhängigkeit theoretischer Momente der Verteilung (eventuell erst nach vorheriger Berechnung wie in Möglichkeit 1)
  - Ersetzen der theoretischen Momente in den resultierenden „Formeln“ für die Verteilungsparameter durch die entsprechenden empirischen Momente.
- Beide Möglichkeiten führen (natürlich) zum selben Ergebnis!
- Nützlich ist hier gelegentlich der Varianzzerlegungssatz

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 \ .$$

Beispiele (Momentenmethode) I

1. Schätzung des Parameters p einer Alternativ-/Bernoulliverteilung:
   - Verteilungsannahme: $W = \{B(1,p) \mid p \in \Theta = [0,1]\}$
   - Theoretisches 1. Moment: $\mathrm{E}(Y) = p$ (bekannt aus W'rechnung)
   - Gleichsetzen (hier besonders einfach!) von $\mathrm{E}(Y)$ mit 1. empirischen Moment $\overline{X}$ liefert sofort Momentenmethodenschätzer (Methode 1) $\widehat{p} = \overline{X}$.

   Der Schätzer $\widehat{p}$ für die Erfolgswahrscheinlichkeit p nach der Methode der Momente entspricht also gerade dem Anteil der Erfolge in der Stichprobe.
2. Schätzung des Parameters λ einer Exponentialverteilung:
   - Verteilungsannahme: $W = \{\mathrm{Exp}(\lambda) \mid \lambda \in \Theta = \mathbb{R}_{++}\}$
   - Theoretisches 1. Moment: $\mathrm{E}(Y) = \frac{1}{\lambda}$ (bekannt aus W'rechnung)
   - Gleichsetzen von $E(Y)$ mit 1. empirischen Moment $\overline{X}$ liefert (Methode 1)

$$\overline{X} \overset{!}{=} \mathrm{E}(Y) = \frac{1}{\lambda} \quad \Rightarrow \quad \widehat{\lambda} = \frac{1}{\overline{X}} \ .$$

(Vorsicht bei Berechnung der Realisation: $\frac{1}{\overline{x}} \neq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$)

Beispiele (Momentenmethode) II

3. Schätzung der Parameter (μ, σ^2) einer Normalverteilung:
   - Verteilungsannahme: $W = \{N(\mu, \sigma^2) \mid (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_{++}\}$
     Hier bekannt: $E(Y) = \mu$ und $\mathrm{Var}(Y) = \sigma^2$.
     $\rightsquigarrow$ Methode 2 bietet sich an (mit Varianzzerlegungssatz):
   - Verteilungsparameter $\mu = \mathrm{E}(Y)$
     Verteilungsparameter $\sigma^2 = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$
   - Einsetzen der empirischen Momente anstelle der theoretischen Momente liefert $\widehat{\mu} = \overline{X}$ sowie $\widehat{\sigma^2} = \overline{X^2} - \overline{X}^2$ als Schätzer nach der Momentenmethode.
   - Am Beispiel der Realisation

$$8.75, 10.37, 8.33, 13.19, 10.66, 8.36, 10.97, 11.48, 11.15, 9.39$$

einer Stichprobe vom Umfang 10 erhält man mit

$$\overline{x} = 10.265 \qquad \text{und} \qquad \overline{x^2} = 107.562$$

die realisierten Schätzwerte

$$\widehat{\mu} = 10.265 \qquad \text{und} \qquad \widehat{\sigma^2} = 107.562 - 10.265^2 = 2.192 \ .$$

Maximum-Likelihood-Methode (ML-Methode)

- Weitere geläufige Schätzmethode: **Maximum-Likelihood-Methode**
- **Vor** Erläuterung der Methode: einleitendes Beispiel

Beispiel: ML-Methode durch Intuition (?)

Ein „fairer" Würfel sei auf einer unbekannten Anzahl $r \in \{0, 1, 2, 3, 4, 5, 6\}$ von Seiten rot lackiert, auf den übrigen Seiten andersfarbig.
Der Würfel wird 100-mal geworfen und es wird festgestellt, wie oft eine rote Seite (oben) zu sehen war.

- Angenommen, es war 34-mal eine rote Seite zu sehen; wie würden Sie die Anzahl der rot lackierten Seiten auf dem Würfel schätzen?
- Angenommen, es war 99-mal eine rote Seite zu sehen; wie würden Sie nun die Anzahl der rot lackierten Seiten auf dem Würfel schätzen?

Welche Überlegungen haben Sie insbesondere zu dem zweiten Schätzwert geführt?

Erläuterung Beispiel I

- Bei der Bearbeitung des obigen Beispiels wendet man (zumindest im 2. Fall) vermutlich intuitiv die Maximum-Likelihood-Methode an!
- Prinzipielle Idee der Maximum-Likelihood-Methode:
  Wähle denjenigen der möglichen Parameter als Schätzung aus, bei dem die beobachtete Stichprobenrealisation am plausibelsten ist!
- Im Beispiel interessiert die (unbekannte) Anzahl der roten Seiten.
- Kenntnis der Anzahl der roten Seiten ist (Würfel ist „fair"!) gleichbedeutend mit der Kenntnis der Wahrscheinlichkeit, dass eine rote Seite oben liegt; offensichtlich ist diese Wahrscheinlichkeit nämlich $\frac{r}{6}$, wenn $r \in \{0, \ldots, 6\}$ die Anzahl der roten Seiten bezeichnet.
- Interessierender Umweltausschnitt kann also durch die Zufallsvariable Y beschrieben werden, die den Wert 1 annimmt, falls bei einem Würfelwurf eine rote Seite oben liegt, 0 sonst.
- Y ist dann offensichtlich $B(1, p)$-verteilt mit unbekanntem Parameter $p \in \{0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1\}$, die 2. Grundannahme ist also erfüllt mit

$$W = \left\{ B(1, p) \mid p \in \left\{ 0, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1 \right\} \right\} .$$

Erläuterung Beispiel II

- 100-maliges Werfen des Würfels und jeweiliges Notieren einer 1, falls eine rote Seite oben liegt, einer 0 sonst, führt offensichtlich zu einer Realisation $x_1, \ldots, x_n$ einer einfachen Stichprobe $X_1, \ldots, X_n$ vom Umfang $n = 100$ zu Y, denn $X_1, \ldots, X_n$ sind als Resultat wiederholter Würfelwürfe offensichtlich unabhängig identisch verteilt wie Y.
- Wiederum (vgl. Taschengeldbeispiel) ist es aber nützlich, sich schon *vorher* Gedanken über die Verteilung der Anzahl der (insgesamt geworfenen) Würfe mit obenliegender roten Seite zu machen!
- Aus Veranstaltung „Deskriptive Statistik und Wahrscheinlichkeitsrechnung" bekannt: Für die Zufallsvariable Z, die die Anzahl der roten Seiten bei 100-maligem Werfen beschreibt, also für

$$Z = \sum_{i=1}^{100} X_i = X_1 + \ldots + X_{100} ,$$

  gilt $Z \sim B(100, p)$, falls $Y \sim B(1, p)$.
- Ziel: Aus Stichprobe $X_1, \ldots, X_{100}$ bzw. der Realisation $x_1, \ldots, x_{100}$ (über die Stichprobenfunktion Z bzw. deren Realisation $z = x_1 + \ldots + x_{100}$) auf unbekannten Parameter p und damit die Anzahl der roten Seiten r schließen.

Erläuterung Beispiel III

- Im Beispiel: Umsetzung der ML-Methode besonders einfach, da Menge W der möglichen Verteilungen (aus Verteilungsannahme) **endlich**.
- „Plausibilität" einer Stichprobenrealisation kann hier direkt anhand der Eintrittswahrscheinlichkeit der Realisation gemessen und für alle möglichen Parameter p bestimmt werden.
- Wahrscheinlichkeit (abhängig von p), dass Z Wert z annimmt:

$$P\{Z = z|p\} = \binom{100}{z} \cdot p^z \cdot (1 - p)^{100-z}$$

- Für die erste Realisation $z = 34$ von Z:

r	0	1	2	3	4	5	6
p	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
$P\{Z = 34\|p\}$	0	$1.2 \cdot 10^{-5}$	$8.31 \cdot 10^{-2}$	$4.58 \cdot 10^{-4}$	$1.94 \cdot 10^{-11}$	$5.17 \cdot 10^{-28}$	0

- Für die zweite Realisation $z = 99$ von Z:

r	0	1	2	3	4	5	6
p	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1
$P\{Z = 99\|p\}$	0	$7.65 \cdot 10^{-76}$	$3.88 \cdot 10^{-46}$	$7.89 \cdot 10^{-29}$	$1.23 \cdot 10^{-16}$	$2.41 \cdot 10^{-7}$	0

Bemerkungen zum Beispiel

- Die angegebenen Wahrscheinlichkeiten für Z fassen jeweils mehrere mögliche Stichprobenrealisationen zusammen (da für den Wert von Z irrelevant ist, *welche* der Stichprobenzufallsvariablen X_i den Wert 0 bzw. 1 angenommen haben), für die ML-Schätzung ist aber eigentlich die Wahrscheinlichkeit einer einzelnen Stichprobenrealisation maßgeblich. Die Wahrscheinlichkeit einer einzelnen Stichprobenrealisation erhält man, indem der Faktor $\binom{100}{z}$ entfernt wird; dieser ist jedoch in jeder der beiden Tabellen konstant und beeinflusst daher die Bestimmung des Maximums nicht.
- Eher untypisch am Beispiel (aber umso geeigneter zur Erklärung der Methode!) ist die Tatsache, dass W eine endliche Menge von Verteilungen ist. In der Praxis wird man in der Regel unendlich viele Möglichkeiten für die Wahl des Parameters haben, z.B. bei Alternativverteilungen $p \in [0, 1]$. Dies ändert zwar *nichts* am Prinzip der Schätzung, wohl aber an den zur Bestimmung der „maximalen Plausibilität" nötigen (mathematischen) Techniken.
- Dass die „Plausibilität" hier genauer einer Wahrscheinlichkeit entspricht, hängt an der diskreten Verteilung von Y. Ist Y eine stetige Zufallsvariable, übernehmen Dichtefunktionswerte die Messung der „Plausibilität".

Maximum-Likelihood-Methode (im Detail)

Schritte zur ML-Schätzung

Die Durchführung einer ML-Schätzung besteht aus folgenden Schritten:

1. Aufstellung der sog. **Likelihood-Funktion** $L(\theta)$, die *in Abhängigkeit des (unbekannten) Parametervektors* θ die Plausibilität der beobachteten Stichprobenrealisation misst.
2. Suche des (eines) Parameters bzw. Parametervektors $\widehat{\theta}$, der den (zu der beobachteten Stichprobenrealisation) maximal möglichen Wert der Likelihoodfunktion liefert.
   Es ist also *jeder* Parameter(vektor) $\widehat{\theta}$ ein ML-Schätzer, für den gilt:
$$L(\widehat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

- Je nach Anwendungssituation unterscheidet sich die Vorgehensweise in beiden Schritten erheblich.
- Wir setzen bei der Durchführung von ML-Schätzungen **stets** voraus, dass eine **einfache (Zufalls-)Stichprobe** vorliegt!

1. Schritt: Aufstellen der Likelihoodfunktion

- „Plausibilität" oder „Likelihood" der Stichprobenrealisation wird gemessen
  - mit Hilfe der **Wahrscheinlichkeit**, die Stichprobenrealisation $(x_1, \ldots, x_n)$ zu erhalten, d.h. dem Wahrscheinlichkeitsfunktionswert
  $$L(\theta) := p_{X_1,\ldots,X_n}(x_1, \ldots, x_n \,|\, \theta) \ ,$$
  falls Y diskrete Zufallsvariable ist,
  - mit Hilfe der **gemeinsamen Dichtefunktion** ausgewertet an der Stichprobenrealisation $(x_1, \ldots, x_n)$,
  $$L(\theta) := f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \,|\, \theta) \ ,$$
  falls Y stetige Zufallsvariable ist.
- Bei Vorliegen einer einfachen Stichprobe lässt sich die Likelihoodfunktion für diskrete Zufallsvariablen Y **immer** darstellen als
$$\begin{aligned} L(\theta) &= p_{X_1,\ldots,X_n}(x_1, \ldots, x_n \,|\, \theta) \\ &\overset{X_i \text{ unabhängig}}{=} \prod_{i=1}^{n} p_{X_i}(x_i|\theta) \\ &\overset{X_i \text{ verteilt wie } Y}{=} \prod_{i=1}^{n} p_Y(x_i|\theta) \ . \end{aligned}$$

- Analog erhält man bei Vorliegen einer einfachen Stichprobe für stetige Zufallsvariablen Y **immer** die Darstellung
$$\begin{aligned} L(\theta) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \,|\, \theta) \\ &\overset{X_i \text{ unabhängig}}{=} \prod_{i=1}^{n} f_{X_i}(x_i|\theta) \\ &\overset{X_i \text{ verteilt wie } Y}{=} \prod_{i=1}^{n} f_Y(x_i|\theta) \ . \end{aligned}$$
  für die Likelihoodfunktion.
- Ist der Parameterraum Θ endlich, kann im Prinzip $L(\theta)$ für alle $\theta \in \Theta$ berechnet werden und eines der θ als ML-Schätzwert $\widehat{\theta}$ gewählt werden, für das $L(\theta)$ maximal war.
  Für diese (einfache) Situation wird Schritt 2 nicht weiter konkretisiert.
- Ist der Parameterraum Θ ein Kontinuum (z.B. ein Intervall in $\mathbb{R}^K$), müssen für den 2. Schritt i.d.R. Maximierungsverfahren aus der Analysis angewendet werden.

2. Schritt: Maximieren der Likelihoodfunktion

(falls Θ ein Intervall in $\mathbb{R}^K$ ist)

- Wichtige Eigenschaft des Maximierungsproblems aus Schritt 2:

  Wichtig ist nicht der **Wert** des Maximums $L(\widehat{\theta})$ der Likelihoodfunktion, sondern die **Stelle** $\widehat{\theta}$, an der dieser Wert angenommen wird!
- Aus Gründen (zum Teil ganz erheblich) vereinfachter Berechnung:
  - Bilden der **logarithmierten** Likelihoodfunktion (Log-Likelihoodfunktion) $\ln L(\theta)$.
  - Maximieren der Log-Likelihoodfunktion $\ln L(\theta)$ **statt** Maximierung der Likelihoodfunktion.
- Diese Änderung des Verfahrens ändert nichts an den Ergebnissen, denn
  - $\ln : \mathbb{R}_{++} \to \mathbb{R}$ ist eine streng monoton wachsende Abbildung,
  - es genügt, die Likelihoodfunktion in den Bereichen zu untersuchen, in denen sie *positive* Werte annimmt, da nur dort das Maximum angenommen werden kann. Dort ist auch die log-Likelihoodfunktion definiert.

- Maximierung von $\ln L(\theta)$ kann *oft* (aber nicht immer!) auf die aus der Mathematik bekannte Art und Weise erfolgen:
  1. Bilden der ersten Ableitung $\frac{\partial \ln L}{\partial \theta}$ der log-Likelihoodfunktion.
     (Bei mehrdimensionalen Parametervektoren: Bilden der partiellen Ableitungen
     $$\frac{\partial \ln L}{\partial \theta_1}, \ldots, \frac{\partial \ln L}{\partial \theta_K}$$
     der log-Likelihoodfunktion.)
  2. Nullsetzen der ersten Ableitung, um „Kandidaten“ für Maximumsstellen von $\ln L(\theta)$ zu finden:
     $$\frac{\partial \ln L}{\partial \theta} \overset{!}{=} 0 \qquad \rightsquigarrow \qquad \widehat{\theta}$$
     (Bei mehrdimensionalen Parametervektoren: Lösen des Gleichungssystems
     $$\frac{\partial \ln L}{\partial \theta_1} \overset{!}{=} 0, \qquad \ldots \qquad , \frac{\partial \ln L}{\partial \theta_K} \overset{!}{=} 0$$
     um „Kandidaten“ $\widehat{\theta}$ für Maximumsstellen von $\ln L(\theta)$ zu finden.)
  3. Überprüfung anhand des Vorzeichens der 2. Ableitung $\frac{\partial^2 \ln L}{(\partial \theta)^2}$ (bzw. der Definitheit der Hessematrix), ob tatsächlich eine Maximumsstelle vorliegt:
     $$\frac{\partial^2 \ln L}{(\partial \theta)^2}(\widehat{\theta}) \overset{?}{<} 0$$

- Auf die Überprüfung der 2. Ableitung bzw. der Hessematrix verzichten wir häufig, um nicht durch mathematische Schwierigkeiten von den statistischen abzulenken.
- Durch den Übergang von der Likelihoodfunktion zur log-Likelihoodfunktion erhält man gegenüber den Darstellungen aus Folie 39 und 40 im diskreten Fall nun
  $$\ln L(\theta) = \ln\left(\prod_{i=1}^{n} p_Y(x_i|\theta)\right) = \sum_{i=1}^{n} \ln\left(p_Y(x_i|\theta)\right)$$
  und im stetigen Fall
  $$\ln L(\theta) = \ln\left(\prod_{i=1}^{n} f_Y(x_i|\theta)\right) = \sum_{i=1}^{n} \ln\left(f_Y(x_i|\theta)\right) .$$
- Die wesentliche Vereinfachung beim Übergang zur log-Likelihoodfunktion ergibt sich meist dadurch, dass die Summen in den obigen Darstellungen deutlich leichter abzuleiten sind als die Produkte in den Darstellungen der Likelihoodfunktion auf Folie 39 und Folie 40.
- Falls „Standardverfahren“ keine Maximumsstelle liefert $\rightsquigarrow$ „Gehirn einschalten“

Beispiel: ML-Schätzung für Exponentialverteilung

Erinnerung: $f_Y(y|\lambda) = \lambda e^{-\lambda y}$ für $y > 0$, $\lambda > 0$

1. Aufstellen der Likelihoodfunktion (im Fall $x_i > 0$ für alle i):
   $$L(\lambda) = \prod_{i=1}^{n} f_Y(x_i|\lambda) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i}\right)$$
2. Aufstellen der log-Likelihoodfunktion (im Fall $x_i > 0$ für alle i):
   $$\ln L(\lambda) = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n} (\ln \lambda + (-\lambda x_i)) = n \cdot \ln \lambda - \lambda \cdot \sum_{i=1}^{n} x_i$$
3. Ableiten und Nullsetzen der log-Likelihoodfunktion:
   $$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \overset{!}{=} 0$$
   liefert
   $$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$
   als ML-Schätzer (2. Ableitung $\frac{\partial^2 \ln L}{(\partial \lambda)^2}\left(\frac{1}{\overline{x}}\right) = -\frac{n}{\lambda^2} < 0$).

Bemerkungen

- Häufiger wird die Abhängigkeit der Likelihoodfunktion von der Stichprobenrealisation auch durch Schreibweisen der Art $L(\theta; x_1, \ldots, x_n)$ oder $L(x_1, \ldots, x_n|\theta)$ ausgedrückt.
- Vorsicht geboten, falls Bereich positiver Dichte bzw. der Träger der Verteilung von Y von Parametern abhängt!
Im Beispiel: Bereich positiver Dichte $\mathbb{R}_{++}$ *unabhängig* vom Verteilungsparameter λ, Maximierungsproblem unter Vernachlässigung des Falls *„mindestens ein x_i kleiner oder gleich 0"* betrachtet, da dieser Fall **für keinen der möglichen Parameter** mit positiver Wahrscheinlichkeit eintritt.
Dieses „Vernachlässigen" ist nicht immer unschädlich!
- Bei diskreten Zufallsvariablen mit „wenig" verschiedenen Ausprägungen oft Angabe der absoluten Häufigkeiten für die einzelnen Ausprägungen in der Stichprobe statt Angabe der Stichprobenrealisation $x_1, \ldots, x_n$ selbst.
Beispiel: Bei Stichprobe vom Umfang 25 zu alternativverteilter Zufallsvariablen Y häufiger Angabe von „18 Erfolge in der Stichprobe der Länge 25" als Angabe der Stichprobenrealisation

$$0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1 \,.$$

Beispiel: ML-Schätzung für Alternativverteilungen I

- Verteilungsannahme: $Y \sim B(1, p)$ für $p \in \Theta = [0, 1]$ mit

$$p_Y(y|p) = \left\{ \begin{array}{cl} p & \text{falls } y = 1 \\ 1 - p & \text{falls } y = 0 \end{array} \right\} = p^y \cdot (1-p)^{1-y} \text{ für } y \in \{0, 1\} \,.$$

1. Aufstellen der Likelihoodfunktion:

$$L(p) = \prod_{i=1}^{n} p_Y(x_i|p) = \prod_{i=1}^{n} \left(p^{x_i} \cdot (1-p)^{1-x_i}\right) = p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i}$$

bzw. — wenn $n_1 := \sum_{i=1}^n x_i$ die Anzahl der „Einsen" (Erfolge) in der Stichprobe angibt —

$$L(p) = p^{n_1} \cdot (1-p)^{n-n_1}$$

2. Aufstellen der log-Likelihoodfunktion:

$$\ln L(p) = n_1 \ln(p) + (n - n_1) \ln(1-p)$$

Beispiel: ML-Schätzung für Alternativverteilungen II

3. Ableiten und Nullsetzen der log-Likelihoodfunktion:

$$\begin{aligned} \frac{\partial \ln L}{\partial p} &= \frac{n_1}{p} - \frac{n - n_1}{1 - p} \overset{!}{=} 0 \\ \Leftrightarrow \quad n_1 - n_1 p &= np - n_1 p \\ \Rightarrow \quad \widehat{p} &= \frac{n_1}{n} \end{aligned}$$

Die 2. Ableitung $\frac{\partial^2 \ln L}{(\partial p)^2} = -\frac{n_1}{p^2} - \frac{n-n_1}{(1-p)^2}$ ist negativ für $0 < p < 1$, der Anteil der Erfolge in der Stichprobe $\widehat{p} = n_1/n$ ist also der ML-Schätzer.

Bemerkung:
Die Bestimmung des ML-Schätzers $\widehat{p}$ ist so nur zulässig, wenn $0 < p < 1$ gilt (sonst Multiplikation einer Gleichung mit 0). Allerdings gilt

- für $p = 0$ offensichtlich stets $n_1 = 0$, damit ist $\widehat{p} = \frac{n_1}{n} = 0$ für $p = 0$ ebenfalls der ML-Schätzer, denn für $n_1 = 0$ ist $L(p) = (1-p)^n$ maximal für $p = 0$;
- für $p = 1$ offensichtlich stets $n_1 = n$, damit ist $\widehat{p} = \frac{n}{n} = 1$ für $p = 1$ ebenfalls der ML-Schätzer, denn für $n_1 = n$ ist $L(p) = p^n$ maximal für $p = 1$.

$\Rightarrow \widehat{p} = \dfrac{n_1}{n}$ ist ML-Schätzer für Verteilungsannahme $Y \sim B(1, p), p \in [0, 1]$.

Beispiel: ML-Schätzung für Poissonverteilungen I

- Verteilungsannahme: $Y \sim \text{Pois}(\lambda)$ für $\lambda \in \Theta = \mathbb{R}_{++}$ mit

$$p_Y(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

für $k \in \mathbb{N}_0$.

1. Aufstellen der Likelihoodfunktion:

$$L(\lambda) = \prod_{i=1}^{n} p_Y(x_i|\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right)$$

(falls alle $x_i \in \mathbb{N}_0$)

2. Aufstellen der log-Likelihoodfunktion:

$$\ln L(\lambda) = \sum_{i=1}^{n} \left(x_i \ln(\lambda) - \ln(x_i!) - \lambda\right) = \left(\sum_{i=1}^{n} x_i\right) \ln(\lambda) - \left(\sum_{i=1}^{n} \ln(x_i!)\right) - n\lambda$$

Beispiel: ML-Schätzung für Poissonverteilungen II

3. Ableiten und Nullsetzen der log-Likelihoodfunktion:

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda} &= \frac{\sum_{i=1}^n x_i}{\lambda} - n \overset{!}{=} 0 \\ \Rightarrow \qquad \widehat{\lambda} &= \frac{\sum_{i=1}^n x_i}{n} = \overline{x} \end{aligned}$$

mit $\frac{\partial^2 \ln L}{(\partial\lambda)^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0$ für alle $\lambda > 0$, $\widehat{\lambda} = \overline{x}$ ist also der ML-Schätzer für λ.

- Aus Wahrscheinlichkeitsrechnung bekannt: $Y \sim \text{Pois}(\lambda) \Rightarrow E(Y) = \lambda$, also ergibt sich (hier) auch für den Schätzer nach der Momentenmethode offensichtlich $\widehat{\lambda} = \overline{X}$.
- Wird (ähnlich zur Anzahl n_1 der Erfolge in einer Stichprobe zu einer alternativverteilten Grundgesamtheit) statt der (expliziten) Stichprobenrealisation $x_1, \dots, x_n$ eine „Häufigkeitsverteilung" der in der Stichprobe aufgetretenen Werte angegeben, kann $\overline{x}$ mit der aus der deskriptiven Statistik bekannten „Formel" ausgerechnet werden.

Beispiel: ML-Schätzung bei diskreter Gleichverteilung

- Verteilungsannahme: für ein (unbekanntes) $M \in \mathbb{N}$ nimmt Y die Werte $\{1, \dots, M\}$ mit der gleichen Wahrscheinlichkeit von jeweils $1/M$ an, d.h.:

$$p_Y(k|M) = \begin{cases} \frac{1}{M} & \text{falls } k \in \{1, \dots, M\} \\ 0 & \text{falls } k \notin \{1, \dots, M\} \end{cases}$$

1. Aufstellen der Likelihoodfunktion:

$$\begin{aligned} L(M) &= \prod_{i=1}^n p_Y(x_i|M) = \begin{cases} \frac{1}{M^n} & \text{falls } x_i \in \{1, \dots, M\} \text{ für alle } i \\ 0 & \text{falls } x_i \notin \{1, \dots, M\} \text{ für mindestens ein } i \end{cases} \\ &= \begin{cases} \frac{1}{M^n} & \text{falls } \max\{x_1, \dots, x_n\} \leq M \\ 0 & \text{falls } \max\{x_1, \dots, x_n\} > M \end{cases} \quad (\text{gegeben } x_i \in \mathbb{N} \text{ für alle } i) \end{aligned}$$

2. Maximieren der Likelihoodfunktion:
   Offensichtlich ist $L(M)$ für $\max\{x_1, \dots, x_n\} \leq M$ streng monoton fallend in M, M muss also **unter Einhaltung der Bedingung** $\max\{x_1, \dots, x_n\} \leq M$ möglichst klein gewählt werden. Damit erhält man den ML-Schätzer als $\widehat{M} = \max\{x_1, \dots, x_n\}$.

Beurteilung von Schätzfunktionen

- *Bisher:* Zwei Methoden zur Konstruktion von Schätzfunktionen bekannt.
- *Problem:*

  Wie kann Güte/Qualität dieser Methoden bzw. der resultierenden Schätzfunktionen beurteilt werden?
- *Lösung:*

  Zu gegebener Schätzfunktion $\widehat{\theta}$ für θ: Untersuchung des **zufälligen** Schätzfehlers $\widehat{\theta} - \theta$ (bzw. dessen Verteilung)
- Naheliegende Forderung für „gute" Schätzfunktionen:

  Verteilung des Schätzfehler sollte möglichst „dicht" um 0 konzentriert sein (d.h. Verteilung von $\widehat{\theta}$ sollte möglichst „dicht" um θ konzentriert sein)
- Aber:
  - Was bedeutet das?
  - Wie vergleicht man zwei Schätzfunktionen $\widehat{\theta}$ und $\widetilde{\theta}$? Wann ist Schätzfunktion $\widehat{\theta}$ „besser" als $\widetilde{\theta}$ (und was bedeutet „besser")?
  - Was ist zu beachten, wenn Verteilung des Schätz**fehlers** noch vom zu schätzenden Parameter abhängt?

Bias, Erwartungstreue

- Eine offensichtlich gute Eigenschaft von Schätzfunktionen ist, wenn der zu schätzende (wahre) Parameter zumindest *im Mittel* getroffen wird, d.h. der *erwartete* Schätzfehler gleich Null ist:

Definition 3.4 (Bias, Erwartungstreue)

Seien W eine parametrische Verteilungsannahme mit Parameterraum Θ, $\widehat{\theta}$ eine Schätzfunktion für θ. Dann heißt

1. der erwartete Schätzfehler

   $$\text{Bias}(\widehat{\theta}) := \text{E}(\widehat{\theta} - \theta) = \text{E}(\widehat{\theta}) - \theta$$

   die **Verzerrung** oder der **Bias** von $\widehat{\theta}$,
2. die Schätzfunktion $\widehat{\theta}$ **erwartungstreu für** θ oder auch **unverzerrt für** θ, falls $\text{Bias}(\widehat{\theta}) = 0$ bzw. $\text{E}(\widehat{\theta}) = \theta$ für alle $\theta \in \Theta$ gilt.
3. Ist allgemeiner $g : \Theta \to \mathbb{R}$ eine (messbare) Abbildung, so betrachtet man auch Schätzfunktionen $\widehat{g(\theta)}$ für $g(\theta)$ und nennt diese **erwartungstreu für** $g(\theta)$, wenn $\text{E}(\widehat{g(\theta)} - g(\theta)) = 0$ für alle $\theta \in \Theta$ gilt.

Bemerkungen

- Obwohl Definition 3.4 auch für mehrdimensionale Parameterräume Θ geeignet ist („0“ entspricht dann ggf. dem Nullvektor), betrachten wir zur Vereinfachung im Folgenden meist nur noch **eindimensionale** Parameterräume $\Theta \subseteq \mathbb{R}$.
- Ist beispielsweise W als Verteilungsannahme für Y die Menge aller Alternativverteilungen $B(1,p)$ mit Parameter $p \in \Theta = [0,1]$, so ist der ML-Schätzer $\widehat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ bei Vorliegen einer Zufallsstichprobe $X_1, \ldots, X_n$ zu Y erwartungstreu für p, denn es gilt:

$$\begin{aligned} \mathrm{E}(\widehat{p}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) &\overset{\mathrm{E}\text{ linear}}{=} \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \\ &\overset{F_{X_i}=F_Y}{=} \frac{1}{n}\sum_{i=1}^n \mathrm{E}(Y) \\ &= \frac{1}{n}\cdot n \cdot p = p \text{ für alle } p \in [0,1] \end{aligned}$$

- Allgemeiner gilt, dass $\overline{X}$ bei Vorliegen einer Zufallsstichprobe stets erwartungstreu für $\mathrm{E}(Y)$ ist, denn es gilt analog zu oben:

$$\begin{aligned} \mathrm{E}(\overline{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) &\overset{\mathrm{E}\text{ linear}}{=} \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \\ &\overset{F_{X_i}=F_Y}{=} \frac{1}{n}\sum_{i=1}^n \mathrm{E}(Y) \\ &= \frac{1}{n}\cdot n \cdot \mathrm{E}(Y) = \mathrm{E}(Y) \end{aligned}$$

- Genauso ist klar, dass man für beliebiges k mit dem k-ten empirischen Moment $\overline{X^k}$ bei Vorliegen einer Zufallsstichprobe stets erwartungstreue Schätzer für das k-te theoretische Moment $\mathrm{E}(Y^k)$ erhält, denn es gilt:

$$\mathrm{E}(\overline{X^k}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n X_i^k\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i^k) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(Y^k) = \mathrm{E}(Y^k)$$

- Der nach der Methode der Momente erhaltene Schätzer

$$\widehat{\sigma^2} = \overline{X^2} - \overline{X}^2 \overset{\text{Verschiebungssatz}}{=} \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$$

für den Parameter σ^2 einer normalverteilten Zufallsvariable ist **nicht** erwartungstreu für σ^2.
Bezeichnet $\sigma^2 := \mathrm{Var}(Y)$ nämlich die (unbekannte) Varianz der Zufallsvariablen Y, so kann gezeigt werden, dass für $\widehat{\sigma^2}$ generell

$$\mathrm{E}(\widehat{\sigma^2}) = \frac{n-1}{n}\sigma^2$$

gilt. Einen erwartungstreuen Schätzer für σ^2 erhält man folglich mit der sogenannten **Stichprobenvarianz**

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{n}{n-1}\widehat{\sigma^2}\ ,$$

denn es gilt offensichtlich

$$\mathrm{E}(S^2) = \mathrm{E}\left(\frac{n}{n-1}\widehat{\sigma^2}\right) = \frac{n}{n-1}\mathrm{E}\left(\widehat{\sigma^2}\right) = \frac{n}{n-1}\cdot\frac{n-1}{n}\cdot\sigma^2 = \sigma^2\ .$$

Vergleich von Schätzfunktionen

- Beim Vergleich von Schätzfunktionen: **oft** Beschränkung auf erwartungstreue Schätzfunktionen
- In der Regel: viele erwartungstreue Schätzfunktionen denkbar.
- Für die Schätzung von $\mu := E(Y)$ beispielsweise alle *gewichteten* Mittel

$$\widehat{\mu}_{w_1,\ldots,w_n} := \sum_{i=1}^n w_i \cdot X_i$$

mit der Eigenschaft $\sum_{i=1}^n w_i = 1$ erwartungstreu für μ, denn es gilt dann offensichtlich

$$\mathrm{E}\left(\widehat{\mu}_{w_1,\ldots,w_n}\right) = \mathrm{E}\left(\sum_{i=1}^n w_i \cdot X_i\right) = \sum_{i=1}^n w_i\,\mathrm{E}(X_i) = \mathrm{E}(Y)\cdot\sum_{i=1}^n w_i = \mathrm{E}(Y) = \mu\ .$$

- Problem: Welche Schätzfunktion ist „die beste“?
- Übliche Auswahl (bei Beschränkung auf erwartungstreue Schätzfunktionen!): Schätzfunktionen mit geringerer **Streuung (Varianz)** bevorzugen.

Wirksamkeit, Effizienz

Definition 3.5 (Wirksamkeit, Effizienz)

Sei W eine parametrische Verteilungsannahme mit Parameterraum Θ.

1. Seien $\widehat{\theta}$ und $\widetilde{\theta}$ erwartungstreue Schätzfunktionen für θ. Dann heißt $\widehat{\theta}$ **mindestens so wirksam** wie $\widetilde{\theta}$, wenn
$$\mathrm{Var}(\widehat{\theta}) \leq \mathrm{Var}(\widetilde{\theta}) \text{ für alle } \theta \in \Theta$$
gilt. $\widehat{\theta}$ heißt **wirksamer** als $\widetilde{\theta}$, wenn *außerdem* $\mathrm{Var}(\widehat{\theta}) < \mathrm{Var}(\widetilde{\theta})$ für mindestens ein $\theta \in \Theta$ gilt.
2. Ist $\widehat{\theta}$ mindestens so wirksam wie alle (anderen) Schätzfunktionen einer Klasse mit erwartungstreuen Schätzfunktionen für θ, so nennt man $\widehat{\theta}$ **effizient** in dieser Klasse erwartungstreuer Schätzfunktionen.

- Die Begriffe „Wirksamkeit" und „Effizienz" betrachtet man analog zu Definition 3.5 ebenfalls, wenn Funktionen $g(\theta)$ eines Parameters θ Gegenstand der Schätzung sind.
- $\sqrt{\mathrm{Var}(\widehat{\theta})}$ wird auch **Standardfehler** oder **Stichprobenfehler** von $\widehat{\theta}$ genannt.

Beispiel: Effizienz

- Betrachte Klasse der (linearen) erwartungstreuen Schätzfunktionen
$$\widehat{\mu}_{w_1,\ldots,w_n} := \sum_{i=1}^{n} w_i \cdot X_i$$
mit $\sum_{i=1}^{n} w_i = 1$ für den Erwartungswert $\mu := E(Y)$ aus Folie 56.
- Für welche $w_1, \ldots, w_n$ erhält man (bei Vorliegen einer einfachen Stichprobe) die in dieser Klasse **effiziente** Schätzfunktion $\widehat{\mu}_{w_1,\ldots,w_n}$?

$\rightsquigarrow$ Suche nach den Gewichten $w_1, \ldots, w_n$ (mit $\sum_{i=1}^{n} w_i = 1$), für die $\mathrm{Var}(\widehat{\mu}_{w_1,\ldots,w_n})$ möglichst klein wird.

- Man kann zeigen, dass $\mathrm{Var}(\widehat{\mu}_{w_1,\ldots,w_n})$ minimal wird, wenn
$$w_i = \frac{1}{n} \text{ für alle } i \in \{1, \ldots, n\}$$
gewählt wird.
- Damit ist $\overline{X}$ also effizient in der Klasse der linearen erwartungstreuen Schätzfunktionen für den Erwartungswert μ einer Verteilung!

Mittlerer quadratischer Fehler (MSE)

- Wenn Erwartungstreue im Vordergrund steht, ist Auswahl nach minimaler Varianz der Schätzfunktion sinnvoll.
- Ist Erwartungstreue nicht das „übergeordnete" Ziel, verwendet man zur Beurteilung der Qualität von Schätzfunktionen häufig auch den sogenannten mittleren quadratischen Fehler (mean square error, MSE).

Definition 3.6 (Mittlerer quadratischer Fehler (MSE))

Sei W eine parametrische Verteilungsannahme mit Parameterraum Θ, $\widehat{\theta}$ eine Schätzfunktion für $\theta \in \Theta$. Dann heißt $\mathrm{MSE}(\widehat{\theta}) := \mathrm{E}\left[(\widehat{\theta} - \theta)^2\right]$ der **mittlere quadratische Fehler (mean square error, MSE)** von $\widehat{\theta}$.

- Mit dem (umgestellten) Varianzzerlegungssatz erhält man direkt
$$\mathrm{E}\left[(\widehat{\theta} - \theta)^2\right] = \underbrace{\mathrm{Var}(\widehat{\theta} - \theta)}_{=\mathrm{Var}(\widehat{\theta})} + \underbrace{\left[\mathrm{E}(\widehat{\theta} - \theta)\right]^2}_{=(\mathrm{Bias}(\widehat{\theta}))^2},$$
für erwartungstreue Schätzfunktionen stimmt der MSE einer Schätzfunktion also gerade mit der Varianz überein!

Konsistenz im quadratischen Mittel

- Basierend auf dem MSE ist ein „minimales" Qualitätskriterium für Schätzfunktionen etabliert.
- Das Kriterium fordert (im Prinzip), dass man den MSE durch Vergrößerung des Stichprobenumfangs beliebig klein bekommen muss.
- Zur Formulierung des Kriteriums müssen Schätzfunktionen $\widehat{\theta}_n$ für „variable" Stichprobengrößen $n \in \mathbb{N}$ betrachtet werden.

Definition 3.7 (Konsistenz im quadratischen Mittel)

Seien W eine parametrische Verteilungsannahme mit Parameterraum Θ, $\widehat{\theta}_n$ eine Schätzfunktion für $\theta \in \Theta$ zum Stichprobenumfang $n \in \mathbb{N}$.
Dann heißt die (Familie von) Schätzfunktion(en) $\widehat{\theta}_n$ **konsistent im quadratischen Mittel**, falls
$$\lim_{n\to\infty} \mathrm{MSE}(\widehat{\theta}_n) = \lim_{n\to\infty} \mathrm{E}\left[(\widehat{\theta}_n - \theta)^2\right] = 0$$
für alle $\theta \in \Theta$ gilt.

- Mit der (additiven) Zerlegung des MSE in Varianz und quadrierten Bias aus Folie 59 erhält man sofort:

Satz 3.8

Seien W eine parametrische Verteilungsannahme mit Parameterraum Θ, $\widehat{\theta}_n$ eine Schätzfunktion für $\theta \in \Theta$ zum Stichprobenumfang $n \in \mathbb{N}$. Dann ist die Familie $\widehat{\theta}_n$ von Schätzfunktionen genau dann konsistent im quadratischen Mittel für θ, wenn sowohl

1. $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n - \theta) = 0$ *bzw.* $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n) = \theta$ *als auch*
2. $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) = 0$

für alle $\theta \in \Theta$ gilt.

- Eigenschaft 1 aus Satz 3.8 wird auch **asymptotische Erwartungstreue** genannt; asymptotische Erwartungstreue ist offensichtlich schwächer als Erwartungstreue.
- Es gibt also auch (Familien von) Schätzfunktionen, die für einen Parameter θ zwar konsistent im quadratischen Mittel sind, aber nicht erwartungstreu.

Beispiel: Konsistenz im quadratischen Mittel

- Voraussetzung (wie üblich): $X_1, \ldots, X_n$ einfache Stichprobe zu Y.
- Bekannt: Ist $\mu := \mathrm{E}(Y)$ der unbekannte Erwartungswert der interessierenden Zufallsvariable Y, so ist $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ für alle $n \in \mathbb{N}$ erwartungstreu.
- Ist $\sigma^2 := \mathrm{Var}(Y)$ die Varianz von Y, so erhält man für die Varianz von $\overline{X}_n$ (vgl. Beweis der Effizienz von $\overline{X}$ unter allen linearen erwartungstreuen Schätzfunktionen für μ):

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathrm{Var}(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}$$

- Es gilt also $\lim_{n\to\infty} \mathrm{Var}(\overline{X}_n) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$, damit folgt zusammen mit der Erwartungstreue, dass $\overline{X}_n$ konsistent im quadratischen Mittel für μ ist.

Verteilung des Stichprobenmittels $\overline{X}$

- **Bisher:** Interesse meist an einigen *Momenten* (Erwartungswert und Varianz) von Schätzfunktionen, insbesondere des Stichprobenmittels $\overline{X}$.
- Bereits bekannt: Ist $\mu := E(Y)$, $\sigma^2 := \mathrm{Var}(Y)$ und $X_1, \ldots, X_n$ eine einfache Stichprobe zu Y, so gilt

$$\mathrm{E}(\overline{X}) = \mu \qquad \text{sowie} \qquad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \ .$$

- Damit Aussagen über Erwartungstreue, Wirksamkeit, Konsistenz möglich.
- **Jetzt:** Interesse an ganzer **Verteilung** von Schätzfunktionen, insbesondere $\overline{X}$.
- Verteilungsaussagen entweder
  - auf Grundlage des Verteilungstyps von Y aus der Verteilungsannahme in speziellen Situationen **exakt** möglich oder
  - auf Grundlage des zentralen Grenzwertsatzes (bei genügend großem Stichprobenumfang!) allgemeiner **näherungsweise (approximativ)** möglich.
- Wir unterscheiden im Folgenden nur zwischen:
  - Y normalverteilt $\rightsquigarrow$ Verwendung der exakten Verteilung von $\overline{X}$.
  - Y nicht normalverteilt $\rightsquigarrow$ Verwendung der Näherung der Verteilung von $\overline{X}$ aus dem zentralen Grenzwertsatz.

Aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“:

1. Gilt $Y \sim N(\mu, \sigma^2)$, so ist $\overline{X}$ **exakt** normalverteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$, es gilt also
$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \ .$$
2. Ist Y beliebig verteilt mit $\mathrm{E}(Y) =: \mu$ und $\mathrm{Var}(Y) =: \sigma^2$, so rechtfertigt der zentrale Grenzwertsatz **für ausreichend große Stichprobenumfänge** n die Näherung der tatsächlichen Verteilung von $\overline{X}$ durch eine Normalverteilung mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$ (wie oben!), man schreibt dann auch
$$\overline{X} \overset{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$
und sagt „$\overline{X}$ ist approximativ (näherungsweise) $N\left(\mu, \frac{\sigma^2}{n}\right)$-verteilt“.

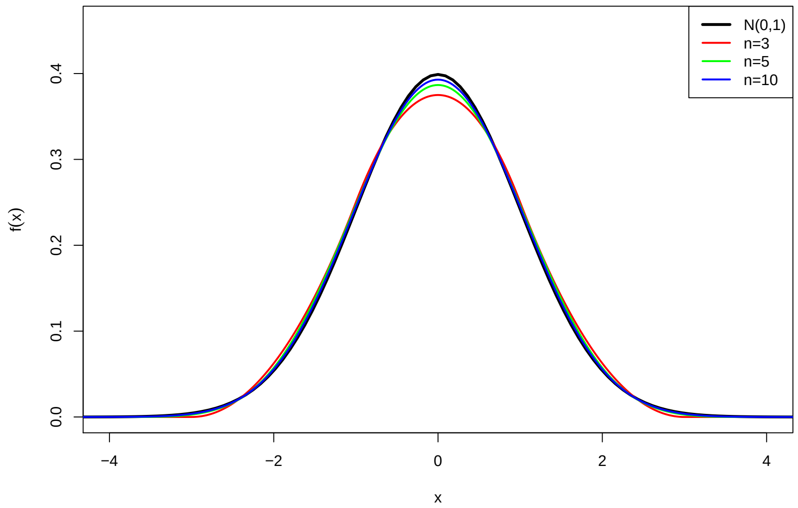
Der Standardabweichung $\mathrm{Sd}(\overline{X}) = \sqrt{\mathrm{Var}(\overline{X})}$ von $\overline{X}$ (also der Standardfehler der Schätzfunktion $\overline{X}$ für μ) wird häufig mit $\sigma_{\overline{X}} := \frac{\sigma}{\sqrt{n}}$ abgekürzt.

- Die Qualität der Näherung der Verteilung im Fall ❷ wird mit zunehmendem Stichprobenumfang höher, hängt aber **ganz entscheidend** vom Verteilungstyp (und sogar der konkreten Verteilung) von Y ab!
- Pauschale Kriterien an den Stichprobenumfang n („Daumenregeln", z.B. $n \geq 30$) finden sich häufig in der Literatur, sind aber nicht ganz unkritisch.
- Verteilungseigenschaft $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ bzw. $\overline{X} \overset{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ wird meistens (äquivalent!) in der (auch aus dem zentralen Grenzwertsatz bekannten) Gestalt

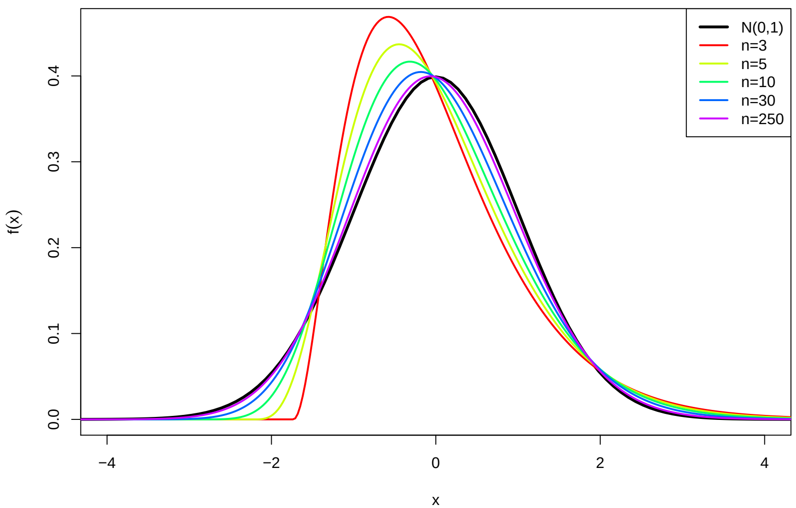
$$\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1) \qquad \text{bzw.} \qquad \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \overset{\bullet}{\sim} N(0,1)$$

  verwendet, da dann Verwendung von Tabellen zur Standardnormalverteilung möglich.
- Im Folgenden: Einige Beispiele für Qualität von Näherungen durch Vergleich der Dichtefunktion der Standardnormalverteilungsapproximation mit der tatsächlichen Verteilung von $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ für unterschiedliche Stichprobenumfänge n.

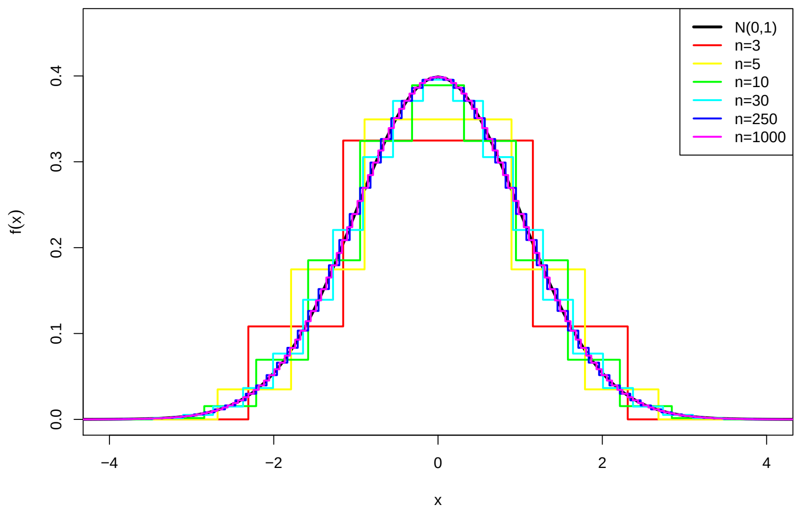
Beispiel: Näherung, falls $Y \sim \text{Unif}(20, 50)$



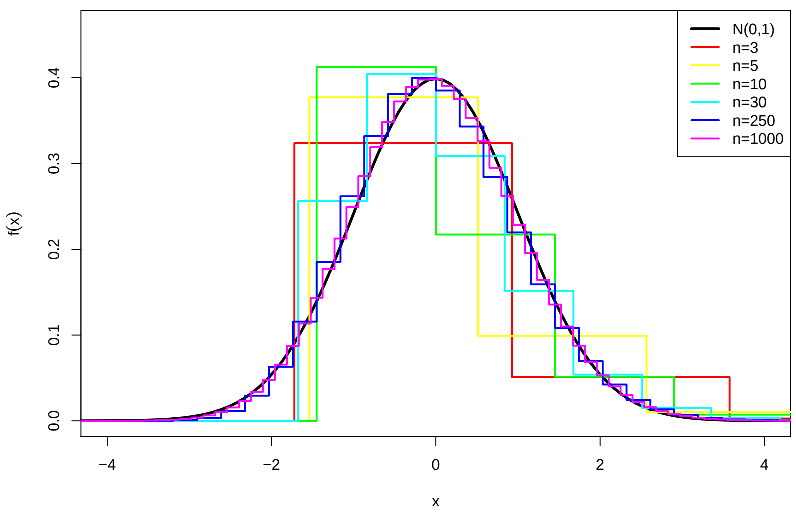
Beispiel: Näherung, falls $Y \sim \text{Exp}(2)$



Beispiel: Näherung, falls $Y \sim B(1, 0.5)$



Beispiel: Näherung, falls $Y \sim B(1, 0.05)$



Schwankungsintervalle für $\overline{X}$

- *Eine* Verwendungsmöglichkeit für Verteilung von $\overline{X}$:

  Berechnung von (festen) Intervallen mit der Eigenschaft, dass die Stichprobenziehung mit einer vorgegebenen Wahrscheinlichkeit zu einer Realisation von $\overline{X}$ führt, die in dieses berechnete Intervall fällt.

  Solche Intervalle heißen **Schwankungsintervalle**.
- Gesucht sind also Intervallgrenzen $g_u < g_o$ von Intervallen $[g_u, g_o]$ mit

$$P_{\overline{X}}([g_u, g_o]) = P\{\overline{X} \in [g_u, g_o]\} \overset{!}{=} p_S$$

  für eine vorgegebene Wahrscheinlichkeit $p_S \in (0, 1)$.
- Aus bestimmten Gründen (die später verständlich werden) gibt man nicht p_S vor, sondern die Gegenwahrscheinlichkeit $\alpha := 1 - p_S$, d.h. man fordert

$$P_{\overline{X}}([g_u, g_o]) = P\{\overline{X} \in [g_u, g_o]\} \overset{!}{=} 1 - \alpha$$

  für ein vorgegebenes $\alpha \in (0, 1)$.
  $1 - \alpha$ wird dann auch **Sicherheitswahrscheinlichkeit** genannt.

- Eindeutigkeit für die Bestimmung von g_u und g_o erreicht man durch die Forderung von **Symmetrie** in dem Sinn, dass die untere bzw. obere Grenze des Intervalls jeweils mit einer Wahrscheinlichkeit von $\alpha/2$ unter- bzw. überschritten werden soll, d.h. man fordert genauer

$$P\{\overline{X} < g_u\} \overset{!}{=} \frac{\alpha}{2} \qquad \text{und} \qquad P\{\overline{X} > g_o\} \overset{!}{=} \frac{\alpha}{2} .$$

- Unter Verwendung der Verteilungseigenschaft

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0, 1) \qquad \text{bzw.} \qquad \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \overset{\bullet}{\sim} N(0, 1)$$

  erhält man also exakt bzw. näherungsweise

$$\begin{aligned} P\{\overline{X} < g_u\} &= P\left\{\frac{\overline{X} - \mu}{\sigma}\sqrt{n} < \frac{g_u - \mu}{\sigma}\sqrt{n}\right\} \overset{!}{=} \frac{\alpha}{2} \\ \Leftrightarrow \qquad \frac{g_u - \mu}{\sigma}\sqrt{n} &= \Phi^{-1}\left(\frac{\alpha}{2}\right) \\ \Rightarrow \qquad g_u &= \mu + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(\frac{\alpha}{2}\right) \end{aligned}$$

  als untere Intervallgrenze.

- Analog erhält man exakt bzw. näherungsweise

$$\begin{aligned} P\{\overline{X} > g_o\} &= P\left\{\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \frac{g_o - \mu}{\sigma}\sqrt{n}\right\} \overset{!}{=} \frac{\alpha}{2} \\ \Leftrightarrow \qquad \frac{g_o - \mu}{\sigma}\sqrt{n} &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\ \Rightarrow \qquad g_o &= \mu + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) . \end{aligned}$$

  als die obere Intervallgrenze.
- Als Abkürzung für p-Quantile der Standardnormalverteilung (also Funktionswerte von Φ^{-1} an der Stelle $p \in (0, 1)$) verwenden wir:

$$N_p := \Phi^{-1}(p)$$

- Man erhält also insgesamt als symmetrisches Schwankungsintervall für $\overline{X}$ exakt bzw. näherungsweise das Intervall

$$\left[\mu + \frac{\sigma}{\sqrt{n}} \cdot N_{\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] .$$

Bemerkungen

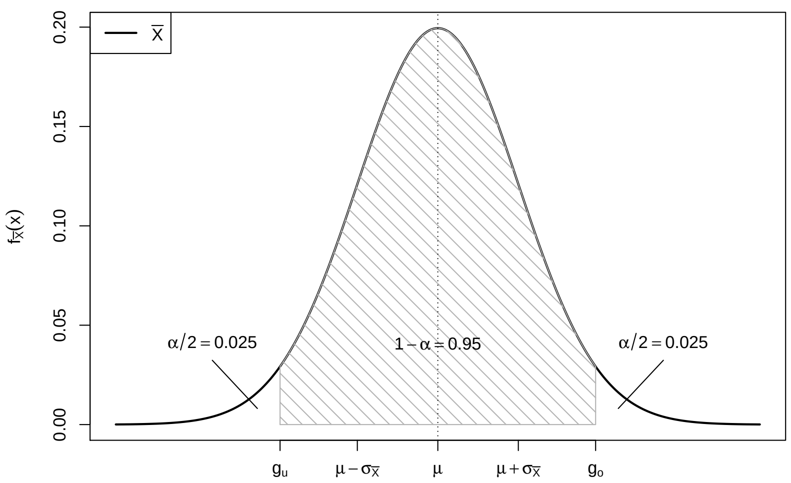
- Die bekannte Symmetrieeigenschaft
$$\Phi(x) = 1 - \Phi(-x) \qquad \text{bzw.} \qquad \Phi(-x) = 1 - \Phi(x)$$
für alle $x \in \mathbb{R}$ überträgt sich auf die Quantile N_p der Standardnormalverteilung in der Form
$$N_p = -N_{1-p} \qquad \text{bzw.} \qquad N_{1-p} = -N_p$$
für alle $p \in (0, 1)$.
- Üblicherweise sind nur die Quantile für $p \geq \frac{1}{2}$ in Tabellen enthalten. Man schreibt daher das Schwankungsintervall meist in der Form
$$\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] .$$
In dieser Gestalt wird (noch klarer) deutlich, dass symmetrische Schwankungsintervalle für $\overline{X}$ ebenfalls (!) stets symmetrisch um μ sind.
- In der Literatur werden anstelle der Abkürzung N_p für die Quantile der Standardnormalverteilung häufig auch die Abkürzungen z_p oder λ_p verwendet.
- Geläufige Sicherheitswahrscheinlichkeiten sind z.B. $1 - \alpha \in \{0.90, 0.95, 0.99\}$.

Beispiel: Schwankungsintervall

- *Aufgabenstellung:*
  - Es gelte $Y \sim N(50, 10^2)$.
  - Zu Y liege eine einfache Stichprobe $X_1, \ldots, X_{25}$ der Länge $n = 25$ vor.
  - Gesucht ist ein (symmetrisches) Schwankungsintervall für $\overline{X}$ zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0.95$.
- *Lösung:*
  - Es gilt also $\mu := E(Y) = 50$, $\sigma^2 := \mathrm{Var}(Y) = 10^2$, $n = 25$ und $\alpha = 0.05$.
  - Zur Berechnung des Schwankungsintervalls
$$\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$$
benötigt man also nur noch das $1 - \alpha/2 = 0.975$-Quantil $N_{0.975}$ der Standardnormalverteilung. Dies erhält man mit geeigneter Software (oder aus geeigneten Tabellen) als $N_{0.975} = 1.96$.
  - Insgesamt erhält man also das Schwankungsintervall
$$\left[50 - \frac{10}{\sqrt{25}} \cdot 1.96, 50 + \frac{10}{\sqrt{25}} \cdot 1.96\right] = [46.08, 53.92] .$$
  - Die Ziehung einer Stichprobenrealisation führt also mit einer Wahrscheinlichkeit von 95% zu einer Realisation $\overline{x}$ von $\overline{X}$ im Intervall $[46.08, 53.92]$.

Beispiel: Schwankungsintervall (Grafische Darstellung)

Im Beispiel: $\overline{X} \sim N\left(50, \frac{10^2}{25}\right)$



Konfidenzintervalle

- Schwankungsintervalle für $\overline{X}$ zu gegebenem Erwartungswert μ und gegebener Varianz σ^2 von Y eher theoretisch interessant.
- In praktischen Anwendungen der schließenden Statistik: μ (und eventuell auch σ^2) unbekannt!
- Ziel ist es, über die (bereits diskutierte) Parameterpunktschätzung durch $\overline{X}$ hinaus *mit Hilfe der Verteilung von* $\overline{X}$ eine Intervallschätzung von μ zu konstruieren, die bereits Information über die Güte der Schätzung enthält.
- Ansatz zur Konstruktion dieser Intervallschätzer ähnlich zum Ansatz bei der Konstruktion von (symmetrischen) Schwankungsintervallen.
- Idee: Verwende die Kenntnis der Verteilung von $\overline{X}$ (abhängig vom unbekannten μ), um zufällige (von der Stichprobenrealisation abhängige) Intervalle zu konstruieren, die den wahren Erwartungswert μ mit einer vorgegebenen Wahrscheinlichkeit überdecken.
- Konfidenzintervalle nicht nur für den Erwartungswert μ einer Verteilung möglich; hier allerdings Beschränkung auf Konfidenzintervalle für μ.

Konfidenzintervalle für μ bei bekannter Varianz σ^2

- Für die (festen!) Schwankungsintervalle $\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$ für $\overline{X}$ zur Sicherheitswahrscheinlichkeit $1-\alpha$ auf Grundlage der exakten oder näherungsweise verwendeten Standardnormalverteilung der Größe $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ gilt nach Konstruktion

$$P\left\{\overline{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha\ .$$

- Idee: Auflösen dieser Wahrscheinlichkeitsaussage nach μ, das heißt, Suche von **zufälligen** Intervallgrenzen $\mu_u < \mu_o$ mit der Eigenschaft

$$P\{\mu \in [\mu_u, \mu_o]\} = P\{\mu_u \leq \mu \leq \mu_o\} \stackrel{!}{=} 1-\alpha\ .$$

(bzw. genauer $P\{\mu < \mu_u\} \stackrel{!}{=} \frac{\alpha}{2}$ und $P\{\mu > \mu_o\} \stackrel{!}{=} \frac{\alpha}{2}$).

- Solche Intervalle $[\mu_u, \mu_o]$ nennt man dann **Konfidenzintervalle für** μ **zum Konfidenzniveau (zur Vertrauenswahrscheinlichkeit)** $1-\alpha$.

Man erhält

$$\begin{aligned}
& P\left\{\overline{X} \in \left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \overline{X} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{-\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq -\mu \leq -\overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \geq \mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha \\
\Leftrightarrow\quad & P\left\{\mu \in \left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = 1-\alpha
\end{aligned}$$

und damit das Konfidenzintervall

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$$

zum Konfidenzniveau $1-\alpha$ für μ.

- In der resultierenden Wahrscheinlichkeitsaussage

$$P\left\{\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

sind die **Intervallgrenzen**

$$\mu_u = \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}} \qquad \text{und} \qquad \mu_o = \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}$$

des Konfidenzintervalls **zufällig** (nicht etwa μ!).

- Ziehung einer Stichprobenrealisation liefert also Realisationen der Intervallgrenzen und damit ein konkretes Konfidenzintervall, welches den wahren (unbekannten) Erwartungswert μ entweder überdeckt oder nicht.
- Die Wahrscheinlichkeitsaussage für Konfidenzintervalle zum Konfidenzniveau $1-\alpha$ ist also so zu verstehen, dass man bei der Ziehung der Stichprobe mit einer Wahrscheinlichkeit von $1-\alpha$ ein Stichprobenergebnis erhält, welches zu einem realisierten Konfidenzintervall führt, das den wahren Erwartungswert überdeckt.

Beispiel: Konfidenzintervall bei bekanntem σ^2

- Die Zufallsvariable Y sei normalverteilt mit unbekanntem Erwartungswert und bekannter Varianz $\sigma^2 = 2^2$.
- Gesucht: Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha = 0.99$.
- Als Realisation $x_1, \ldots, x_{16}$ einer einfachen Stichprobe $X_1, \ldots, X_{16}$ vom Umfang $n = 16$ zu Y liefere die Stichprobenziehung

  18.75, 20.37, 18.33, 23.19, 20.66, 18.36, 20.97, 21.48, 21.15, 19.39, 23.02, 20.78, 18.76, 15.57, 22.25, 19.91 ,

  was zur Realisationen $\overline{x} = 20.184$ von $\overline{X}$ führt.
- Als Realisation des Konfidenzintervalls für μ zum Konfidenzniveau $1-\alpha = 0.99$ erhält man damit insgesamt

$$\begin{aligned}
& \left[\overline{x} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] \\
= & \left[20.184 - \frac{2}{\sqrt{16}} \cdot 2.576, 20.184 + \frac{2}{\sqrt{16}} \cdot 2.576\right] \\
= & [18.896, 21.472]\ .
\end{aligned}$$

Verteilung von $\overline{X}$ bei unbekanntem σ^2

- Wie kann man vorgehen, falls die Varianz σ^2 von Y unbekannt ist?
- Naheliegender Ansatz: Ersetzen von σ^2 durch eine geeignete Schätzfunktion.
- Erwartungstreue Schätzfunktion für σ^2 bereits bekannt:

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

- Ersetzen von σ durch $S = \sqrt{S^2}$ möglich, Verteilung ändert sich aber:

Satz 5.1

Seien $Y \sim N(\mu, \sigma^2)$, $X_1, \ldots, X_n$ eine einfache Stichprobe zu Y. Dann gilt mit $S := \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1) \ ,$$

*wobei $t(n-1)$ die t-**Verteilung mit** $n-1$ **Freiheitsgraden** bezeichnet.*

Die Familie der $t(n)$-Verteilungen

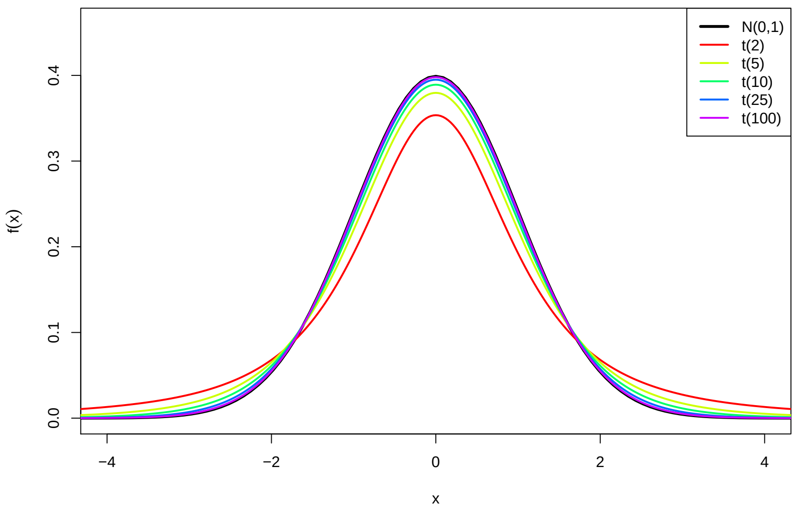
- Die Familie der $t(n)$-Verteilungen mit $n > 0$ ist eine spezielle Familie stetiger Verteilungen. Der Parameter n wird meist „Anzahl der Freiheitsgrade" („degrees of freedom") genannt.
- t-Verteilungen werden (vor allem in englischsprachiger Literatur) oft auch als „Student's t distribution" bezeichnet; „Student" war das Pseudonym, unter dem William Gosset die erste Arbeit zur t-Verteilung in englischer Sprache veröffentlichte.
- $t(n)$-Verteilungen sind für alle $n > 0$ symmetrisch um 0. Entsprechend gilt für p-Quantile der $t(n)$-Verteilung, die wir im Folgendem mit $t_{n;p}$ abkürzen, analog zu Standardnormalverteilungsquantilen

$$t_{n;p} = -t_{n;1-p} \qquad \text{bzw.} \qquad t_{n;1-p} = -t_{n;p}$$

für alle $p \in (0, 1)$
- Für wachsendes n nähert sich die $t(n)$-Verteilung der Standardnormalverteilung an.

Grafische Darstellung einiger $t(n)$-Verteilungen

für $n \in \{2, 5, 10, 25, 100\}$



- Konstruktion von Konfidenzintervallen für μ bei unbekannter Varianz $\sigma^2 = \text{Var}(Y)$ ganz analog zur Situation mit bekannter Varianz, lediglich
  1. Ersetzen von σ durch $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$
  2. Ersetzen von $N_{1-\frac{\alpha}{2}}$ durch $t_{n-1;1-\frac{\alpha}{2}}$

  erforderlich.
- Resultierendes Konfidenzintervall:

$$\left[\overline{X} - \frac{S}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \overline{X} + \frac{S}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

- Benötigte Quantile $t_{n-1;1-\frac{\alpha}{2}}$ können ähnlich wie bei der Standardnormalverteilung z.B. mit der Statistik-Software **R** ausgerechnet werden oder aus geeigneten Tabellen abgelesen werden.
- Mit **R** erhält man z.B. $t_{15;0.975}$ durch

```
> qt(0.975,15)
[1] 2.13145
```

- Mit zunehmendem n werden die Quantile der $t(n)$-Verteilungen betragsmäßig kleiner und nähern sich den Quantilen der Standardnormalverteilung an.

Quantile der t-Verteilungen: $t_{n;p}$

$n \backslash p$	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	1.074	1.341	1.753	2.131	2.602	2.947	4.073
20	1.064	1.325	1.725	2.086	2.528	2.845	3.850
25	1.058	1.316	1.708	2.060	2.485	2.787	3.725
30	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	1.042	1.290	1.660	1.984	2.364	2.626	3.390
200	1.039	1.286	1.653	1.972	2.345	2.601	3.340
500	1.038	1.283	1.648	1.965	2.334	2.586	3.310
1000	1.037	1.282	1.646	1.962	2.330	2.581	3.300
5000	1.037	1.282	1.645	1.960	2.327	2.577	3.292

Beispiel: Konfidenzintervall bei unbekanntem σ^2

- Die Zufallsvariable Y sei normalverteilt mit unbekanntem Erwartungswert und unbekannter Varianz.
- Gesucht: Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha = 0.95$.
- Als Realisation $x_1, \ldots, x_9$ einer einfachen Stichprobe $X_1, \ldots, X_9$ vom Umfang $n = 9$ zu Y liefere die Stichprobenziehung

  28.12, 30.55, 27.49, 34.79, 30.99, 27.54, 31.46, 32.21, 31.73 ,

  was zur Realisationen $\overline{x} = 30.542$ von $\overline{X}$ und zur Realisation $s = 2.436$ von $S = \sqrt{S^2}$ führt.
- Als Realisation des Konfidenzintervalls für μ zum Konfidenzniveau $1 - \alpha = 0.95$ erhält man damit insgesamt

$$
\begin{aligned}
&\left[\overline{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \overline{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right] \\
&= \left[30.542 - \frac{2.436}{\sqrt{9}} \cdot 2.306, 30.542 + \frac{2.436}{\sqrt{9}} \cdot 2.306\right] \\
&= [28.67, 32.414] \ .
\end{aligned}
$$

Konfidenzintervalle, falls Y nicht normalverteilt

1. Ist Y nicht normalverteilt, aber die **Varianz** σ^2 von Y **bekannt**, so verwendet man wie bei der Berechnung der Schwankungsintervalle näherungsweise (durch den zentralen Grenzwertsatz gerechtfertigt!) die Standardnormalverteilung als Näherung der Verteilung von $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ und erhält so **approximative (näherungsweise)** Konfidenzintervalle

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$$

   zum (Konfidenz-)Niveau $1 - \alpha$.

2. Ist Y nicht normalverteilt und die **Varianz** von Y unbekannt, so verwendet man nun analog als Näherung der Verteilung von $\frac{\overline{X}-\mu}{S}\sqrt{n}$ die $t(n-1)$-Verteilung und erhält so **approximative (näherungsweise)** Konfidenzintervalle

$$\left[\overline{X} - \frac{S}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \overline{X} + \frac{S}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

   zum (Konfidenz-)Niveau $1 - \alpha$.

Spezialfall: Konfidenzintervalle für p, falls $Y \sim B(1, p)$

- Gilt $Y \sim B(1, p)$ für einen unbekannten Parameter $p \in [0, 1]$, so können Konfidenzintervalle wegen $p = E(Y) = \mu$ näherungsweise ebenfalls mit Hilfe der Näherung 2 aus Folie 87 bestimmt werden.
- In der „Formel“ für die Berechnung der Konfidenzintervalle ersetzt man üblicherweise $\overline{X}$ wieder durch die in dieser Situation geläufigere (gleichbedeutende!) Notation $\widehat{p}$.
- Die (notwendige) Berechnung von $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ gestaltet sich hier besonders einfach. Man kann zeigen, dass $S^2 = \frac{n}{n-1}\widehat{p}(1-\widehat{p})$ gilt.
- Man erhält so die *von der Stichprobe nur noch über* $\widehat{p}$ *abhängige* Darstellung

$$\left[\widehat{p} - \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}} \cdot t_{n-1;1-\frac{\alpha}{2}}, \widehat{p} + \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n-1}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$$

  für *approximative* Konfidenzintervalle für p zum Niveau $1 - \alpha$.
- Die Güte der Näherung hängt von n und p ab. Je größer n, desto besser; je näher p an $\frac{1}{2}$, desto besser.

Hypothesentests

- *Bisher* wichtigstes betrachtetes Anwendungsbeispiel der schließenden Statistik:

  Punkt- bzw. Intervallschätzung des unbekannten Mittelwerts
- Hierzu: Verwendung der
  1. theoretischen Information über Verteilung von $\overline{X}$
  2. empirischen Information aus Stichprobenrealisation $\overline{x}$ von $\overline{X}$

  zur Konstruktion einer
  - Punktschätzung (inkl. Beurteilung der Genauigkeit des Schätzers!)
  - Intervallschätzung, bei der jede Stichprobenziehung mit einer vorgegebenen Chance ein realisiertes (Konfidenz-)Intervall liefert, welches den (wahren) Mittelwert enthält.
- Nächste Anwendung: **Hypothesentests**:

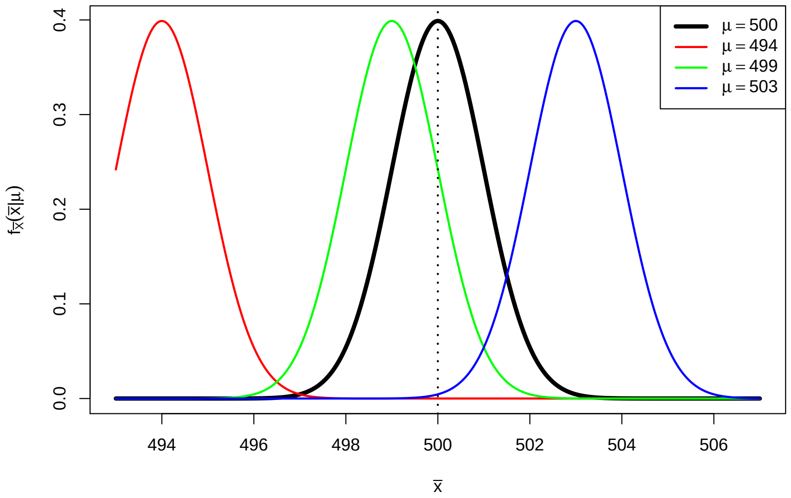
  Entscheidung, ob die unbekannte, wahre Verteilung von Y zu einer vorgegebenen Teilmenge der Verteilungsannahme W gehört oder nicht.
- Zunächst: Illustration der Vorgehensweise am Beispiel einer Entscheidung über den Mittelwert der Verteilung.

Einführendes Beispiel

- Interessierende Zufallsvariable Y:
  Von einer speziellen Abfüllmaschine abgefüllte Inhaltsmenge von Müslipackungen mit Soll-Inhalt $\mu_0 = 500$ (in $[g]$).
- Verteilungsannahme:
  $Y \sim N(\mu, 4^2)$ mit unbekanntem Erwartungswert $\mu = E(Y)$.
- Es liege eine Realisation $x_1, \ldots, x_{16}$ einer einfachen Stichprobe $X_1, \ldots, X_{16}$ vom Umfang $n = 16$ zu Y vor.
- **Ziel:** Verwendung der Stichprobeninformation (über $\overline{X}$ bzw. $\overline{x}$), um zu **entscheiden**, ob die tatsächliche mittlere Füllmenge (also der wahre, unbekannte Parameter μ) mit dem Soll-Inhalt $\mu_0 = 500$ übereinstimmt oder nicht.
- Offensichtlich gilt:
  - $\overline{X}$ schwankt um den wahren Mittelwert μ; selbst wenn $\mu = 500$ gilt, wird $\overline{X}$ praktisch nie genau den Wert $\overline{x} = 500$ annehmen!
  - Realisationen $\overline{x}$ „in der Nähe“ von 500 sprechen eher dafür, dass $\mu = 500$ gilt.
  - Realisationen $\overline{x}$ „weit weg“ von 500 sprechen eher dagegen, dass $\mu = 500$ gilt.
- Also: Entscheidung für Hypothese $\mu = 500$, wenn $\overline{x}$ nahe bei 500, und gegen $\mu = 500$ (also für $\mu \neq 500$), wenn $\overline{x}$ weit weg von 500.
- **Aber:** Wo ist die Grenze zwischen „in der Nähe“ und „weit weg“?

Verteilungen von $\overline{X}$

für verschiedene Parameter μ bei $\sigma = 4$ und $n = 16$



Entscheidungsproblem

- Fällen einer Entscheidung zwischen $\mu = 500$ und $\mu \neq 500$ führt zu *genau einer* der folgenden vier verschiedenen Situationen:

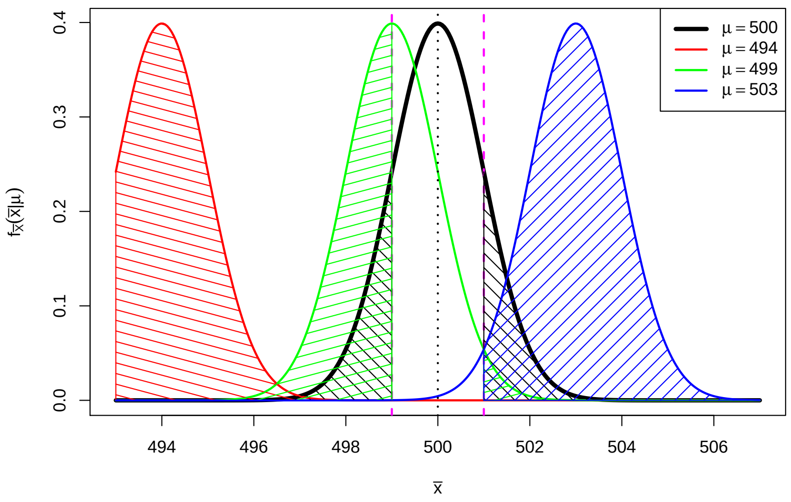
	Tatsächliche Situation: $\mu = 500$	Tatsächliche Situation: $\mu \neq 500$
Entscheidung für $\mu = 500$	richtige Entscheidung	Fehler 2. Art
Entscheidung für $\mu \neq 500$	Fehler 1. Art	richtige Entscheidung

- Wünschenswert:
  Sowohl „Fehler 1. Art“ als auch „Fehler 2. Art“ möglichst selten begehen.
- **Aber:** Zielkonflikt vorhanden:
  Je näher Grenze zwischen „in der Nähe“ und „weit weg“ an $\mu_0 = 500$, desto
  - seltener Fehler 2. Art
  - häufiger Fehler 1. Art

  und umgekehrt für fernere Grenzen zwischen „in der Nähe“ und „weit weg“.

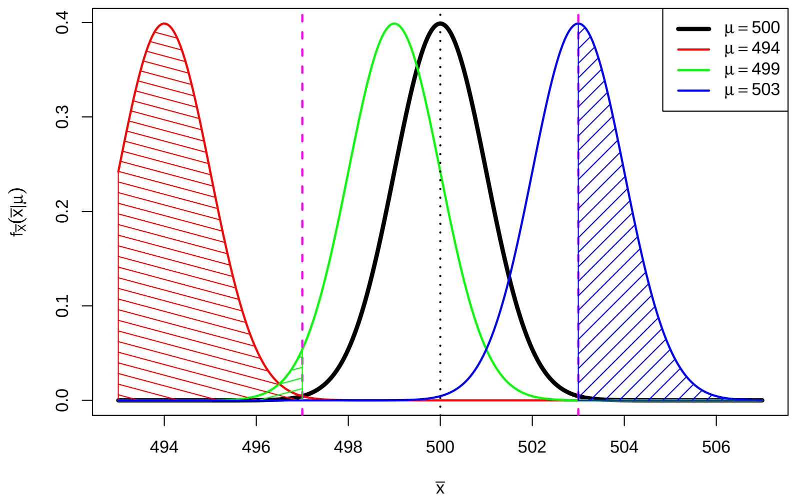
Beispiel für „nahe“ Grenze

Für $\mu \neq 500$ (gegen $\mu = 500$) entscheiden, wenn Abstand zwischen $\overline{x}$ und 500 größer als 1



Beispiel für „ferne“ Grenze

Für $\mu \neq 500$ (gegen $\mu = 500$) entscheiden, wenn Abstand zwischen $\overline{x}$ und 500 größer als 3



- Unmöglich, Wahrscheinlichkeiten der Fehler 1. Art und 2. Art gleichzeitig für alle möglichen Situationen (also alle μ) zu verringern.
- Übliche Vorgehensweise: **Fehler(wahrscheinlichkeit) 1. Art kontrollieren!**
- Also: Vorgabe einer *kleinen* Schranke α (**„Signifikanzniveau“**) für die Wahrscheinlichkeit, mit der man einen Fehler 1. Art begehen darf.
- Festlegung der Grenze zwischen „in der Nähe“ und „weit weg“ so, dass man den Fehler 1. Art nur mit Wahrscheinlichkeit α begeht, also die Realisation $\overline{x}$ **bei Gültigkeit von** $\mu = 500$ nur mit einer Wahrscheinlichkeit von α jenseits der Grenzen liegt, bis zu denen man sich für $\mu = 500$ entscheidet!
- Damit liefert aber das Schwankungsintervall für $\overline{X}$ zur Sicherheitswahrscheinlichkeit $1-\alpha$

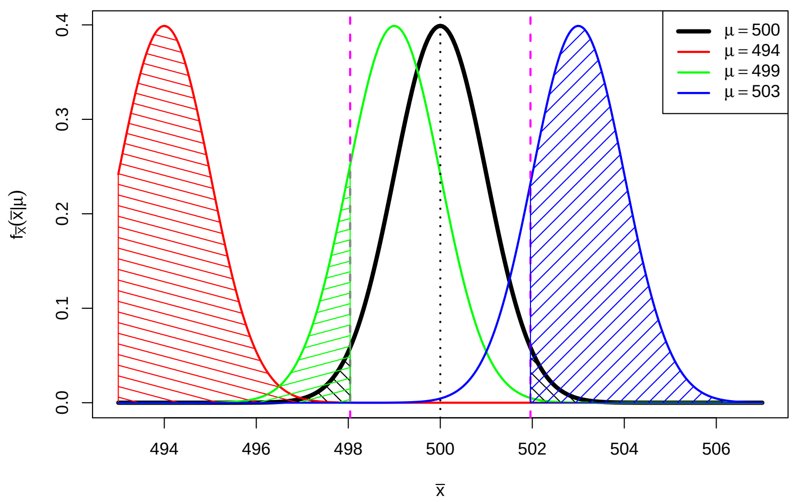
$$\left[\mu - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]$$

für $\mu = \mu_0 = 500$ **(!)** gerade solche Grenzen, denn es gilt im Fall $\mu = \mu_0 = 500$

$$P\left\{\overline{X} \notin \left[\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right]\right\} = \alpha\ .$$

Beispiel für Grenze zum Signifikanzniveau $\alpha = 0.05$

Grenzen aus Schwankungsintervall zur Sicherheitswahrscheinlichkeit $1-\alpha = 0.95$



- Bei einem Signifikanzniveau von $\alpha = 0.05$ entscheidet man sich also **für** $\mu = \mu_0 = 500$ genau dann, wenn die Realisation $\overline{x}$ von $\overline{X}$ im Intervall

$$\left[500 - \frac{4}{\sqrt{16}} \cdot N_{0.975}, 500 + \frac{4}{\sqrt{16}} \cdot N_{0.975}\right] = [498.04, 501.96] \ ,$$

dem sog. **Annahmebereich** des Hypothesentests, liegt.
- Entsprechend fällt die Entscheidung für $\mu \neq 500$ (bzw. **gegen** $\mu = 500$) aus, wenn die Realisation $\overline{x}$ von $\overline{X}$ in der Menge

$$(-\infty, 498.04) \cup (501.96, \infty) \ ,$$

dem sog. **Ablehnungsbereich** oder **kritischen Bereich** des Hypothesentests, liegt.
- Durch Angabe eines dieser Bereiche ist die Entscheidungsregel offensichtlich schon vollständig spezifiziert!
- Statt Entscheidungsregel auf Grundlage der Realisation $\overline{x}$ von $\overline{X}$ (unter Verwendung der Eigenschaft $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$) üblicher:
Äquivalente Entscheidungsregel auf Basis der sog. **Testgröße** oder **Teststatistik** $N := \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ unter Verwendung der Eigenschaft

$$\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0,1) \quad \text{falls} \quad \mu = \mu_0 \ .$$

- Die Verteilungeigenschaft von $N = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ für $\mu = \mu_0$ aus Folie 97 erhält man aus der allgemeineren Verteilungsaussage

$$\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \sim N\left(\frac{\mu - \mu_0}{\sigma}\sqrt{n}, 1\right) \ ,$$

die man wiederum aus der Verteilung $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ durch Anwendung der aus der Wahrscheinlichkeitsrechnung bekannten Transformationsregeln ableiten kann. Damit: $F_N(x) = P\{N \leq x\} = \Phi\left(x - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right)$
- Man rechnet außerdem leicht nach:

$$\overline{X} \in \left[\mu_0 - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] \Leftrightarrow \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \in \left[-N_{1-\frac{\alpha}{2}}, N_{1-\frac{\alpha}{2}}\right]$$

- Als Annahmebereich A für die Testgröße $N = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ erhält man also $\left[-N_{1-\frac{\alpha}{2}}, N_{1-\frac{\alpha}{2}}\right]$, als kritischen Bereich K entsprechend

$$K = \mathbb{R}\backslash A = \left(-\infty, -N_{1-\frac{\alpha}{2}}\right) \cup \left(N_{1-\frac{\alpha}{2}}, \infty\right)$$

und damit eine Formulierung der Entscheidungsregel auf Grundlage von N.

- In Abhängigkeit des tatsächlichen Erwartungswerts μ von Y kann so die Wahrscheinlichkeit für die Ablehnung der Hypothese $\mu = \mu_0$ berechnet werden:

$$\begin{aligned} P\{N \in K\} &= P\left\{N \in (-\infty, -N_{1-\frac{\alpha}{2}}) \cup (N_{1-\frac{\alpha}{2}}, \infty)\right\} \\ &= P\left\{N < -N_{1-\frac{\alpha}{2}}\right\} + P\left\{N > N_{1-\frac{\alpha}{2}}\right\} \\ &= \Phi\left(-N_{1-\frac{\alpha}{2}} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) + 1 - \Phi\left(N_{1-\frac{\alpha}{2}} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) \end{aligned}$$

- Im Beispiel erhält man damit die folgenden Wahrscheinlichkeiten für Annahme bzw. Ablehnung der Hypothese $\mu = \mu_0 = 500$ zu den betrachteten Szenarien (also unterschiedlichen wahren Parametern μ):

	Wahrscheinlichkeit der Annahme von $\mu = 500$ $P\{N \in A\}$	Wahrscheinlichkeit der Ablehnung von $\mu = 500$ $P\{N \in K\}$
$\mu = 500$	**0.95**	0.05
$\mu = 494$	0	**1**
$\mu = 499$	0.8299	**0.1701**
$\mu = 503$	0.1492	**0.8508**

(Fettgedruckte Wahrscheinlichkeiten entsprechen korrekter Entscheidung.)

Grundbegriffe: Hypothesen

- Bekannt: Hypothesentests sind Entscheidungsregeln für die Fragestellung
„Liegt die (unbekannte) Verteilung Q_Y von Y in einer bestimmten **Teilmenge** der Verteilungsannahme W oder nicht?“
- Zur präzisen Formulierung der Fragestellung: Angabe der interessierenden Teilmenge W_0 von Verteilungen mit $\emptyset \neq W_0 \subsetneq W$
- Man nennt dann die Hypothese $Q_Y \in W_0$ auch **Nullhypothese** und schreibt $H_0 : Q_Y \in W_0$. Die Verletzung der Nullhypothese entspricht dem Eintreten der sog. **Gegenhypothese** oder **Alternative** $Q_Y \in W_1 := W \backslash W_0$; man schreibt auch $H_1 : Q_Y \in W_1$.
- Formulierung *prinzipiell* immer in zwei Varianten möglich, da W_0 und W_1 vertauscht werden können. Welche der beiden Varianten gewählt wird, ist allerdings wegen der Asymmetrie in den Wahrscheinlichkeiten für Fehler 1. und 2. Art **nicht unerheblich** (später mehr!).
- Eine Hypothese heißt **einfach**, wenn die zugehörige Teilmenge von W einelementig ist, **zusammengesetzt** sonst.
- Im Beispiel: $W = \{N(\mu, 4^2) \,|\, \mu \in \mathbb{R}\}$, $W_0 = \{N(500, 4^2)\}$.
H_0 ist also einfach, H_1 zusammengesetzt.

Hypothesen bei parametrischen Verteilungsannahmen

- Ist W eine parametrische Verteilungsannahme mit Parameterraum Θ, so existiert offensichtlich immer auch eine (äquivalente) Darstellung von H_0 und H_1 in der Gestalt

$$H_0 : \theta \in \Theta_0 \qquad \text{gegen} \qquad H_1 : \theta \in \Theta_1 := \Theta \backslash \Theta_0$$

  für eine Teilmenge Θ_0 des Parameterraums Θ mit $\emptyset \neq \Theta_0 \subsetneq \Theta$.
- Im Beispiel: $W = \{N(\mu, 4^2) \,|\, \mu \in \Theta = \mathbb{R}\}, \Theta_0 = \{500\}$
- Hypothesenformulierung damit z.B. in der folgenden Form möglich:

$$H_0 : \mu = \mu_0 = 500 \qquad \text{gegen} \qquad H_1 : \mu \neq \mu_0 = 500$$

- Hypothesentests bei parametrischer Verteilungsannahme heißen auch **parametrische (Hypothesen-)Tests**.
- Parametrische Tests heißen (für $\Theta \subseteq \mathbb{R}$) **zweiseitig**, wenn Θ_1 **links und rechts** von Θ_0 liegt, **einseitig** sonst (Im Beispiel: zweiseitiger Test).
- Einseitige Tests heißen **linksseitig**, wenn Θ_1 links von Θ_0 liegt, **rechtsseitig** sonst.

Teststatistik und Ablehnungsbereich

- Nach Präzisierung der Fragestellung in den Hypothesen benötigt man nun eine geeignete **Entscheidungsregel**, die *im Prinzip* jeder möglichen Stichprobenrealisation (aus dem Stichprobenraum $\mathcal{X}$) eine Entscheidung **entweder** für H_0 **oder** für H_1 zuordnet.
- In der Praxis Entscheidung (fast) immer in 3 Schritten:
  1. „Zusammenfassung" der für die Entscheidungsfindung relevanten Stichprobeninformation mit einer geeigneten Stichprobenfunktion, der sog. **Teststatistik** oder **Testgröße** T.
  2. Angabe eines **Ablehnungsbereichs** bzw. **kritischen Bereichs** K, in den die Teststatistik **bei Gültigkeit von** H_0 nur mit einer typischerweise kleinen Wahrscheinlichkeit (durch eine obere Grenze α beschränkt) fallen darf.
  3. Entscheidung **gegen** H_0 bzw. für H_1, falls realisierter Wert der Teststatistik in den **Ablehnungsbereich** bzw. **kritischen Bereich** K fällt, also $T \in K$ gilt (für H_0, falls $T \notin K$).
- Konstruktion des kritischen Bereichs K in Schritt 2 gerade so, dass **Wahrscheinlichkeit für Fehler 1. Art beschränkt** bleibt durch ein vorgegebenes **Signifikanzniveau** (auch **„Irrtumswahrscheinlichkeit"**) α.
- Konstruktion meist so, dass Niveau α gerade eben eingehalten wird (also **kleinste** obere Schranke für die Fehlerwahrscheinlichkeit 1. Art ist).

- Für Konstruktion des kritischen Bereichs wesentlich:
  Analyse der Verteilung der Teststatistik, insbesondere falls H_0 gilt!
- **Im Beispiel:**
  1. Teststatistik: $N = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$
     Verteilung: $N \sim N\left(\frac{\mu-\mu_0}{\sigma}\sqrt{n}, 1\right)$, also insbesondere
     $N \sim N(0,1)$ falls H_0 (also $\mu = \mu_0$) gilt.
  2. Kritischer Bereich: $K = \left(-\infty, -N_{1-\frac{\alpha}{2}}\right) \cup \left(N_{1-\frac{\alpha}{2}}, \infty\right)$

  Wahrscheinlichkeit der Ablehnung von H_0 (abhängig vom Parameter μ):

$$P\{N \in K\} = \Phi\left(-N_{1-\frac{\alpha}{2}} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right) + 1 - \Phi\left(N_{1-\frac{\alpha}{2}} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right)$$

- Die Zuordnung $G : \Theta \to \mathbb{R}; G(\theta) = P\{T \in K\}$ heißt (allgemein) auch **Gütefunktion** des Tests. Im Beispiel also:

$$G(\mu) = \Phi\left(-N_{1-\frac{\alpha}{2}} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right) + 1 - \Phi\left(N_{1-\frac{\alpha}{2}} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right)$$

- Mit der Gütefunktion können also offensichtlich
  - Fehlerwahrscheinlichkeiten 1. Art (für $\theta \in \Theta_0$) direkt durch $G(\theta)$ und
  - Fehlerwahrscheinlichkeiten 2. Art (für $\theta \in \Theta_1$) durch $1 - G(\theta)$

  berechnet werden!

- Berechnung der Eintrittswahrscheinlichkeiten EW mit Gütefunktion $G(\theta)$:

	Tatsächliche Situation: $\theta \in \Theta_0$ (H_0 wahr)	Tatsächliche Situation: $\theta \notin \Theta_0$ (H_0 falsch)
Entscheidung für H_0 ($\theta \in \Theta_0$)	richtige Entscheidung $EW : 1 - G(\theta)$	Fehler 2. Art $EW : 1 - G(\theta)$
Entscheidung gegen H_0 ($\theta \notin \Theta_0$)	Fehler 1. Art $EW : G(\theta)$	richtige Entscheidung $EW : G(\theta)$

- Welche Teststatistik geeignet ist und wie die Teststatistik dann verteilt ist, hängt nicht nur von der Problemformulierung (Hypothesen), sondern oft auch von der Verteilungsannahme ab!
- Test aus Beispiel zum Beispiel *exakt* anwendbar, falls $Y \sim N(\mu, \sigma^2)$ **mit bekannter Varianz**, und *approximativ* anwendbar, wenn Y beliebig verteilt ist **mit bekannter Varianz** (Güte der Näherung abhängig von n sowie Verteilung von Y!)
- Test aus Beispiel heißt auch „zweiseitiger Gauß-Test für den Mittelwert einer Zufallsvariablen mit bekannter Varianz".

Zweiseitiger Gauß-Test

für den Mittelwert einer Zufallsvariablen mit bekannter Varianz

Anwendung

- als **exakter** Test, falls Y normalverteilt und $\text{Var}(Y) = \sigma^2$ bekannt,
- als **approximativer** Test, falls Y beliebig verteilt mit bekannter Varianz σ^2.

„Testrezept" des **zweiseitigen Tests**:

1. Hypothesen: $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ für ein vorgegebenes $\mu_0 \in \mathbb{R}$.
2. Teststatistik:
$$N := \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \text{ mit } N \sim N(0,1) \text{ (bzw. } N \overset{\bullet}{\sim} N(0,1)\text{), falls } H_0 \text{ gilt } (\mu = \mu_0).$$
3. Kritischer Bereich zum Signifikanzniveau α:
$$K = \left(-\infty, -N_{1-\frac{\alpha}{2}}\right) \cup \left(N_{1-\frac{\alpha}{2}}, \infty\right)$$
4. Berechnung der realisierten Teststatistik N
5. Entscheidung: H_0 ablehnen $\Leftrightarrow$ $N \in K$.

Beispiel: Qualitätskontrolle (Länge von Stahlstiften)

- Untersuchungsgegenstand: Weicht die mittlere Länge der von einer bestimmten Maschine produzierten Stahlstifte von der Solllänge $\mu_0 = 10$ (in [cm]) ab, so dass die Produktion gestoppt werden muss?
- Annahmen: Für Länge Y der produzierten Stahlstifte gilt: $Y \sim N(\mu, 0.4^2)$
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 64$ zu Y liefert Stichprobenmittel $\overline{x} = 9.7$.
- Gewünschtes Signifikanzniveau (max. Fehlerwahrscheinlichkeit 1. Art): $\alpha = 0.05$

Geeigneter Test:
(Exakter) Gauß-Test für den Mittelwert bei bekannter Varianz

1. Hypothesen: $H_0 : \mu = \mu_0 = 10 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0 = 10$
2. Teststatistik: $N = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \sim N(0,1)$, falls H_0 gilt $(\mu = \mu_0)$
3. Kritischer Bereich zum Niveau $\alpha = 0.05$:
$K = (-\infty, -N_{0.975}) \cup (N_{0.975}, \infty) = (-\infty, -1.96) \cup (1.96, \infty)$
4. Realisierter Wert der Teststatistik: $N = \frac{9.7-10}{0.4}\sqrt{64} = -6$
5. Entscheidung: $N \in K \rightsquigarrow H_0$ wird abgelehnt und die Produktion gestoppt.

Einseitige Gauß-Tests

Wahl der Hypothesen

- Bei zweiseitigem Test: Hypothesentest zu
$$H_0 : \mu \neq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu = \mu_0$$
zwar konstruierbar, aber ohne praktische Bedeutung.
- Neben zweiseitigem Test zwei (symmetrische) einseitige Varianten:
$$H_0 : \mu \leq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0$$
$$H_0 : \mu \geq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu < \mu_0$$
- Konstruktion der Tests beschränkt Wahrscheinlichkeit, H_0 **fälschlicherweise abzulehnen**. Entscheidung zwischen beiden Varianten daher wie folgt:
  - H_0 : **Nullhypothese** ist in der Regel die Aussage, die von vornherein als glaubwürdig gilt und die man beibehält, wenn das Stichprobenergebnis bei Gültigkeit von H_0 nicht sehr untypisch bzw. überraschend ist.
  - H_1 : **Gegenhypothese** ist in der Regel die Aussage, die man statistisch absichern möchte und für deren Akzeptanz man hohe Evidenz fordert.
  Die Entscheidung für H_1 hat typischerweise erhebliche Konsequenzen, so dass man das Risiko einer fälschlichen Ablehnung von H_0 zugunsten von H_1 kontrollieren will.

- Auch für einseitige Tests fasst Teststatistik
$$N = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \quad \text{mit} \quad N \sim N\left(\frac{\mu - \mu_0}{\sigma}\sqrt{n}, 1\right)$$
die empirische Information über den Erwartungswert μ geeignet zusammen.
- Allerdings gilt nun offensichtlich
  - im Falle des **rechtsseitigen** Tests von
  $$H_0 : \mu \leq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 \ ,$$
  dass **große (insbesondere positive)** Realisationen von N gegen H_0 und für H_1 sprechen, sowie
  - im Falle des **linksseitigen** Tests von
  $$H_0 : \mu \geq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu < \mu_0 \ ,$$
  dass **kleine (insbesondere negative)** Realisationen von N gegen H_0 und für H_1 sprechen.
- Noch nötig zur Konstruktion der Tests:
Geeignetes Verfahren zur Wahl der **kritischen Bereiche** so, dass Wahrscheinlichkeit für Fehler 1. Art durch vorgegebenes Signifikanzniveau α beschränkt bleibt.

Kritischer Bereich (rechtsseitiger Test)

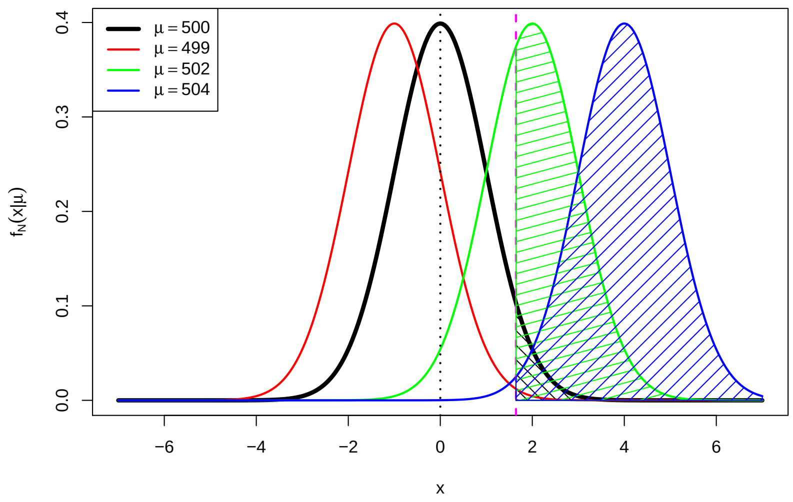
- Für **rechtsseitigen** Test muss also zur Konstruktion des kritischen Bereichs ein **kritischer Wert** bestimmt werden, den die Teststatistik N im Fall der Gültigkeit von H_0 **maximal** mit einer Wahrscheinlichkeit von α überschreitet.
- Gesucht ist also ein Wert k_α mit $P\{N \in (k_\alpha, \infty)\} \le \alpha$ **für alle** $\mu \le \mu_0$.
- Offensichtlich wird $P\{N \in (k_\alpha, \infty)\}$ mit wachsendem μ größer, es genügt also, die Einhaltung der Bedingung $P\{N \in (k_\alpha, \infty)\} \le \alpha$ für das **größtmögliche** μ mit der Eigenschaft $\mu \le \mu_0$, also $\mu = \mu_0$, zu gewährleisten.
- Um die Fehlerwahrscheinlichkeit 2. Art unter Einhaltung der Bedingung an die Fehlerwahrscheinlichkeit 1. Art möglichst klein zu halten, wird k_α gerade so gewählt, dass $P\{N \in (k_\alpha, \infty)\} = \alpha$ für $\mu = \mu_0$ gilt.
- Wegen $N \sim N(0,1)$ für $\mu = \mu_0$ erhält man hieraus

$$\begin{aligned} & & P\{N \in (k_\alpha, \infty)\} &= \alpha \\ &\Leftrightarrow & 1 - P\{N \in (-\infty, k_\alpha)\} &= \alpha \\ &\Leftrightarrow & \Phi(k_\alpha) &= 1 - \alpha \\ &\Leftrightarrow & k_\alpha &= N_{1-\alpha} \end{aligned}$$

und damit insgesamt den kritischen Bereich $K = (N_{1-\alpha}, \infty)$ für den rechtsseitigen Test.

Beispiel für Verteilungen von N

Rechtsseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.05$



Rechtsseitiger Gauß-Test

für den Mittelwert einer Zufallsvariablen mit bekannter Varianz

Anwendung

- als **exakter** Test, falls Y normalverteilt und $\text{Var}(Y) = \sigma^2$ bekannt,
- als **approximativer** Test, falls Y beliebig verteilt mit bekannter Varianz σ^2.

„Testrezept" des **rechtsseitigen Tests**:

1. Hypothesen: $H_0 : \mu \le \mu_0$ gegen $H_1 : \mu > \mu_0$ für ein vorgegebenes $\mu_0 \in \mathbb{R}$.
2. Teststatistik:
$$N := \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \text{ mit } N \sim N(0,1) \ (N \overset{\bullet}{\sim} N(0,1)), \text{ falls } H_0 \text{ gilt (mit } \mu = \mu_0).$$
3. Kritischer Bereich zum Signifikanzniveau α:
$$K = (N_{1-\alpha}, \infty)$$
4. Berechnung der realisierten Teststatistik N
5. Entscheidung: H_0 ablehnen $\Leftrightarrow$ $N \in K$.

Kritischer Bereich (linksseitiger Test)

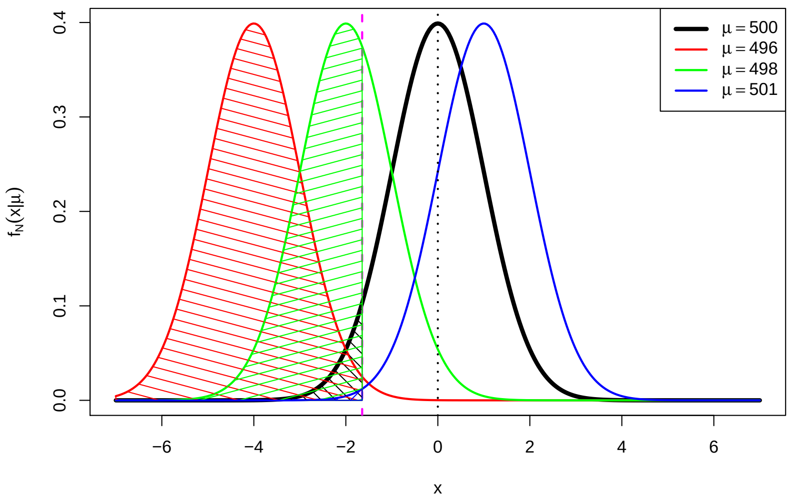
- Für **linksseitigen** Test muss zur Konstruktion des kritischen Bereichs ein **kritischer Wert** bestimmt werden, den die Teststatistik N im Fall der Gültigkeit von H_0 **maximal** mit einer Wahrscheinlichkeit von α unterschreitet.
- Gesucht ist also ein Wert k_α mit $P\{N \in (-\infty, k_\alpha)\} \le \alpha$ **für alle** $\mu \ge \mu_0$.
- Offensichtlich wird $P\{N \in (-\infty, k_\alpha)\}$ mit fallendem μ größer, es genügt also, die Einhaltung der Bedingung $P\{N \in (-\infty, k_\alpha)\} \le \alpha$ für das **kleinstmögliche** μ mit $\mu \ge \mu_0$, also $\mu = \mu_0$, zu gewährleisten.
- Um die Fehlerwahrscheinlichkeit 2. Art unter Einhaltung der Bedingung an die Fehlerwahrscheinlichkeit 1. Art möglichst klein zu halten, wird k_α gerade so gewählt, dass $P\{N \in (-\infty, k_\alpha)\} = \alpha$ für $\mu = \mu_0$ gilt.
- Wegen $N \sim N(0,1)$ für $\mu = \mu_0$ erhält man hieraus

$$\begin{aligned} & & P\{N \in (-\infty, k_\alpha)\} &= \alpha \\ &\Leftrightarrow & \Phi(k_\alpha) &= \alpha \\ &\Leftrightarrow & k_\alpha &= N_\alpha \\ &\Leftrightarrow & k_\alpha &= -N_{1-\alpha} \end{aligned}$$

und damit insgesamt den kritischen Bereich $K = (-\infty, -N_{1-\alpha})$ für den linksseitigen Test.

Beispiel für Verteilungen von N

Linksseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.05$



Linksseitiger Gauß-Test

für den Mittelwert einer Zufallsvariablen mit bekannter Varianz

Anwendung

- als **exakter** Test, falls Y normalverteilt und $\text{Var}(Y) = \sigma^2$ bekannt,
- als **approximativer** Test, falls Y beliebig verteilt mit bekannter Varianz σ^2.

„Testrezept" des **linksseitigen Tests**:

1. Hypothesen: $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ für ein vorgegebenes $\mu_0 \in \mathbb{R}$.
2. Teststatistik:
$$N := \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \text{ mit } N \sim N(0,1) \;(N \overset{\bullet}{\sim} N(0,1)), \text{ falls } H_0 \text{ gilt (mit } \mu = \mu_0).$$
3. Kritischer Bereich zum Signifikanzniveau α:
$$K = (-\infty, -N_{1-\alpha})$$
4. Berechnung der realisierten Teststatistik N
5. Entscheidung: H_0 ablehnen $\Leftrightarrow$ $N \in K$.

Gütefunktionen einseitiger Gauß-Tests

- Gütefunktion allgemein: $G(\theta) = P\{T \in K\}$
- Für **rechtsseitigen** Gauß-Test:
  - $G(\mu) = P\{N \in (N_{1-\alpha}, \infty)\}$
  - Mit $N \sim N\left(\frac{\mu-\mu_0}{\sigma}\sqrt{n}, 1\right)$ erhält man

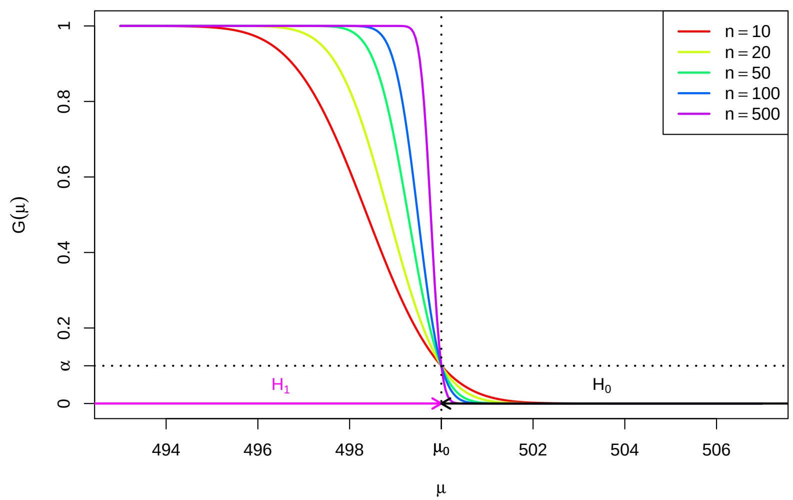
$$\begin{aligned} P\{N \in (N_{1-\alpha}, \infty)\} &= 1 - P\{N \leq N_{1-\alpha}\} \\ &= 1 - \Phi\left(N_{1-\alpha} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{\mu-\mu_0}{\sigma}\sqrt{n} - N_{1-\alpha}\right) \end{aligned}$$

- Für **linksseitigen** Gauß-Test:
  - $G(\mu) = P\{N \in (-\infty, -N_{1-\alpha})\}$
  - Mit $N \sim N\left(\frac{\mu-\mu_0}{\sigma}\sqrt{n}, 1\right)$ erhält man hier

$$\begin{aligned} P\{N \in (-\infty, -N_{1-\alpha})\} &= P\{N < -N_{1-\alpha}\} \\ &= \Phi\left(-N_{1-\alpha} - \frac{\mu-\mu_0}{\sigma}\sqrt{n}\right) \end{aligned}$$

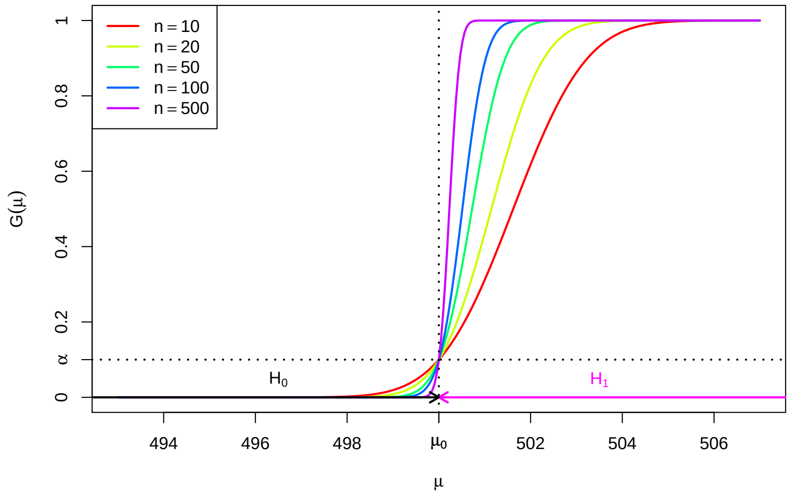
Beispiel für Gütefunktionen

Linksseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.10$



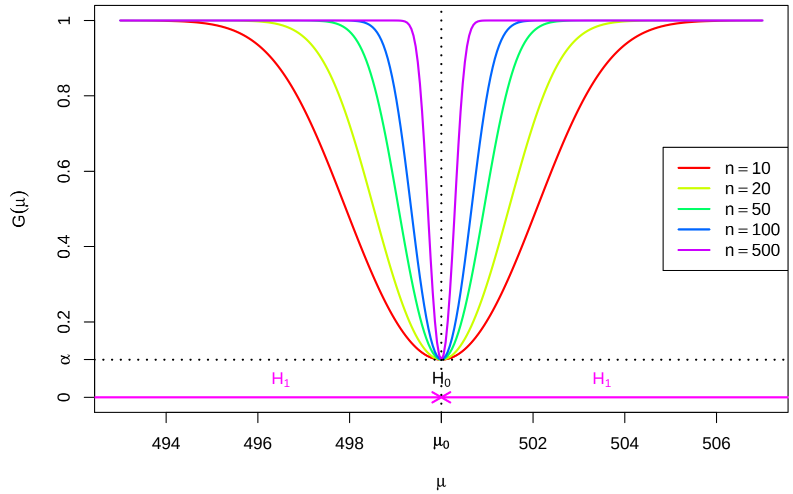
Beispiel für Gütefunktionen

Rechtsseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.10$



Beispiel für Gütefunktionen

Zweiseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.10$



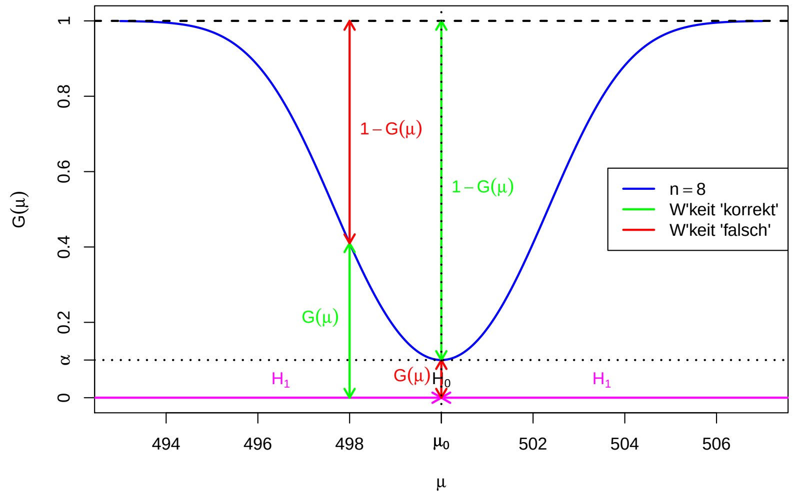
Gütefunktion und Fehlerwahrscheinlichkeiten

für Gauß-Tests auf den Mittelwert bei bekannter Varianz

- Entscheidungsregel (nicht nur) bei Gauß-Tests stets: H_0 ablehnen $\Leftrightarrow$ $N \in K$
- Gütefunktion $G(\mu)$ gibt also für Gauß-Tests auf den Mittelwert bei bekannter Varianz zu jedem möglichen wahren Mittelwert μ die Wahrscheinlichkeit an, eine Stichprobenrealisation zu erhalten, die zu einer Entscheidung **gegen** H_0 führt.
- Dies kann — abhängig davon, ob für μ H_0 oder H_1 zutreffend ist — also die Wahrscheinlichkeit einer falschen bzw. richtigen Entscheidung sein (vgl. Folie 104).
- Gängige Abkürzung
  - für Fehlerwahrscheinlichkeiten 1. Art: $\alpha(\mu)$ für $\mu \in \Theta_0$,
  - für Fehlerwahrscheinlichkeiten 2. Art: $\beta(\mu)$ für $\mu \in \Theta_1$.
- Für $\mu \in \Theta_0$ (also bei Gültigkeit der Nullhypothese für μ) gilt also:
  - Fehlerwahrscheinlichkeit 1. Art: $\alpha(\mu) = G(\mu)$
  - Wahrscheinlichkeit richtiger Entscheidung: $1 - G(\mu)$
- Für $\mu \in \Theta_1$ (also bei Verletzung der Nullhypothese für μ) erhält man:
  - Fehlerwahrscheinlichkeit 2. Art: $\beta(\mu) = 1 - G(\mu)$
  - Wahrscheinlichkeit richtiger Entscheidung: $G(\mu)$

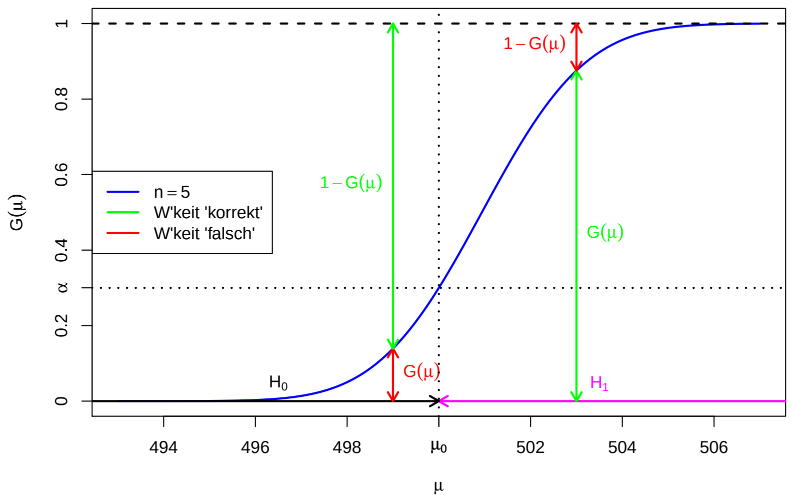
Gütefunktion und Fehlerwahrscheinlichkeiten

Zweiseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.10$



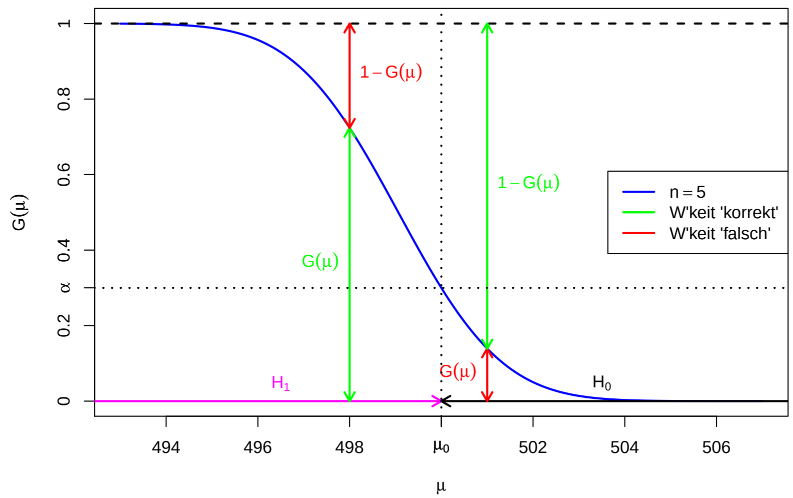
Gütefunktion und Fehlerwahrscheinlichkeiten

Rechtsseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.30$



Gütefunktion und Fehlerwahrscheinlichkeiten

Linksseitiger Test ($\mu_0 = 500$) zum Signifikanzniveau $\alpha = 0.30$



Interpretation von Testergebnissen I

- Durch die Asymmetrie in den Fehlerwahrscheinlichkeiten 1. und 2. Art ist Vorsicht bei Interpretation von Testergebnissen geboten!
- Es besteht ein großer Unterschied zwischen dem Aussagegehalt einer **Ablehnung** von H_0 und dem Aussagegehalt einer **Annahme** von H_0:
  - Fällt die Testentscheidung **gegen** H_0 aus, so hat man — sollte H_0 tatsächlich **erfüllt** sein — wegen der Beschränkung der Fehlerwahrscheinlichkeit 1. Art durch das Signifikanzniveau α nur mit einer typischerweise geringen Wahrscheinlichkeit $\leq \alpha$ eine Stichprobenrealisation erhalten, die **fälschlicherweise** zur **Ablehnung von** H_0 geführt hat.
    Aber: Vorsicht vor „Über"interpretation als Evidenz für Gültigkeit von H_1: Aussagen der Form *„Wenn H_0 abgelehnt wird, dann gilt H_1 mit Wahrscheinlichkeit von mindestens $1 - \alpha$"* sind unsinnig!
  - Fällt die Testentscheidung jedoch **für** H_0 aus, so ist dies ein vergleichsweise meist schwächeres „Indiz" für die Gültigkeit von H_0, da die Fehlerwahrscheinlichkeit 2. Art nicht kontrolliert ist und typischerweise große Werte (bis $1 - \alpha$) annehmen kann. Gilt also tatsächlich H_1, ist es dennoch mit einer sehr großen Wahrscheinlichkeit möglich, eine Stichprobenrealisation zu erhalten, die **fälschlicherweise nicht** zur **Ablehnung von** H_0 führt.

  Aus diesem Grund sagt man auch häufig statt „H_0 wird angenommen" eher „H_0 kann nicht verworfen werden".

Interpretation von Testergebnissen II

- Die Ablehnung von H_0 als Ergebnis eines statistischen Tests wird häufig als
  - **signifikante Veränderung** (zweiseitiger Test),
  - **signifikante Verringerung** (linksseitiger Test) oder
  - **signifikante Erhöhung** (rechtsseitiger Test)

  einer Größe bezeichnet. Konstruktionsbedingt kann das Ergebnis einer statistischen Untersuchung — auch im Fall einer Ablehnung von H_0 — aber **niemals** als zweifelsfreier Beweis für die Veränderung/Verringerung/Erhöhung einer Größe dienen!
- Weiteres Problem: Aussagen über die Fehlerwahrscheinlichkeiten 1. und 2. Art gelten nur perfekt, wenn alle Voraussetzungen erfüllt sind, also wenn
  - Verteilungsannahmen erfüllt sind (Vorsicht bei „approximativen" Tests) und
  - tatsächlich eine **einfache Stichprobe** vorliegt!
- Vorsicht vor „Publication Bias":
  - Bei einem Signifikanzniveau von $\alpha = 0.05$ resultiert im Mittel 1 von 20 statistischen Untersuchungen, bei denen H_0 wahr ist, konstruktionsbedingt in einer Ablehnung von H_0.
  - Gefahr von Fehlinterpretationen, wenn die Untersuchungen, bei denen H_0 nicht verworfen wurde, verschwiegen bzw. nicht publiziert werden!

Interpretation von Testergebnissen III

„signifikant" vs. „deutlich"

- Ein „signifikanter" Unterschied ist noch lange kein „deutlicher" Unterschied!
- Problem: „Fluch des großen Stichprobenumfangs"
- Beispiel: Abfüllmaschine soll Flaschen mit 1000 ml Inhalt abfüllen.
  - Abfüllmenge schwankt zufällig, Verteilung sei Normalverteilung mit bekannter Standardabweichung $\sigma = 0.5$ ml, d.h. in ca. 95% der Fälle liegt Abfüllmenge im Bereich ± 1 ml um den (tatsächlichen) Mittelwert.
  - Statistischer Test zum Niveau $\alpha = 0.05$ zur Überprüfung, ob mittlere Abfüllmenge (Erwartungswert) von 1000 ml abweicht.
- Tatsächlicher Mittelwert sei 1000.1 ml, Test auf Grundlage von 500 Flaschen.
- Wahrscheinlichkeit, die Abweichung von 0.1 ml zu erkennen (Berechnung mit Gütefunktion, siehe Folie 103): 99.4%
- Systematische Abweichung der Abfüllmenge von 0.1 ml zwar mit hoher Wahrscheinlichkeit (99.4%) signifikant, im Vergleich zur (ohnehin vorhandenen) zufälligen Schwankung mit $\sigma = 0.5$ ml aber keinesfalls deutlich!

Fazit: „Durch wissenschaftliche Studien belegte signifikante Verbesserungen" können vernachlässigbar klein sein ($\rightsquigarrow$ Werbung...)

Der p-Wert

- Hypothesentests „komprimieren" Stichprobeninformation zur Entscheidung zwischen H_0 und H_1 zu einem vorgegebenen Signifikanzniveau α.
- Testentscheidung hängt von α **ausschließlich** über kritischen Bereich K ab!
- Genauere Betrachtung offenbart: Abhängigkeit zwischen α und K ist **monoton** im Sinne der Teilmengenbeziehung.
  - Gilt $\widetilde{\alpha} < \alpha$ und bezeichnen $K_{\widetilde{\alpha}}$ und K_α die zugehörigen kritischen Bereiche, so gilt für alle bisher betrachteten Gauß-Tests $K_{\widetilde{\alpha}} \subsetneq K_\alpha$.
  - Unmittelbare Folge ist, dass Ablehnung von H_0 zum Signifikanzniveau $\widetilde{\alpha}$ mit $\widetilde{\alpha} < \alpha$ automatisch eine Ablehnung von H_0 zum Niveau α zur Folge hat (auf Basis derselben Stichprobeninformation)!
  - Außerdem wird K_α für $\alpha \to 0$ beliebig klein und für $\alpha \to 1$ beliebig groß, so dass man für jede Realisation T der Teststatistik sowohl Signifikanzniveaus α mit $T \in K_\alpha$ wählen kann, als auch solche mit $T \notin K_\alpha$.
- Zusammenfassend kann man also zu jeder Realisation T der Teststatistik das kleinste Signifikanzniveau α mit $T \in K_\alpha$ bestimmen (bzw. das größte Signifikanzniveau α mit $T \notin K_\alpha$). Dieses Signifikanzniveau heißt ***p*-Wert** oder **empirisches (marginales) Signifikanzniveau**.
- Mit der Information des p-Werts kann der Test also für **jedes beliebige Signifikanzniveau** α entschieden werden!

p-Wert bei Gauß-Tests

auf den Mittelwert bei bekannter Varianz

- Der Wechsel zwischen „$N \in K_\alpha$" und „$N \notin K_\alpha$" findet bei den diskutierten Gauß-Tests offensichtlich dort statt, wo die realisierte Teststatistik N gerade mit (einer) der Grenze(n) des kritischen Bereichs übereinstimmt, d.h.
  - bei rechtsseitigen Tests mit $K_\alpha = (N_{1-\alpha}, \infty)$ für $N = N_{1-\alpha}$,
  - bei linksseitigen Tests mit $K_\alpha = (-\infty, -N_{1-\alpha})$ für $N = -N_{1-\alpha}$,
  - bei zweiseitigen Tests mit $K_\alpha = (-\infty, -N_{1-\frac{\alpha}{2}}) \cup (N_{1-\frac{\alpha}{2}}, \infty)$ für

$$N = \begin{cases} -N_{1-\frac{\alpha}{2}} & \text{falls } N < 0 \\ N_{1-\frac{\alpha}{2}} & \text{falls } N \geq 0 \end{cases} .$$

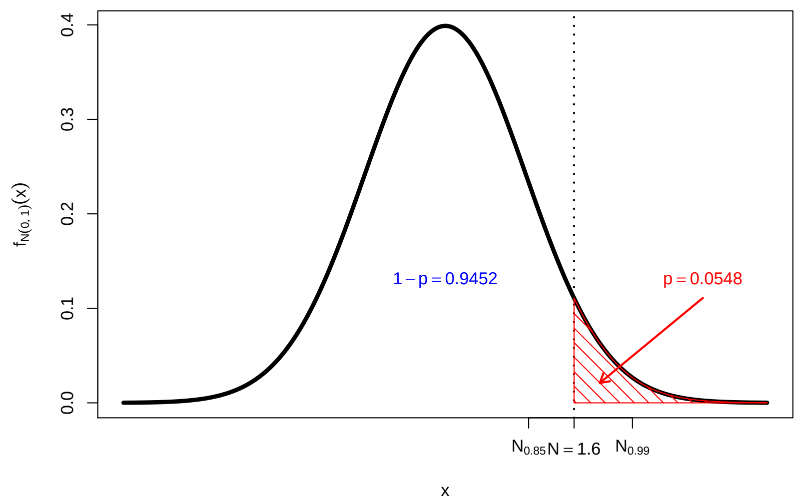
- Durch Auflösen nach α erhält man
  - für rechtsseitige Tests den p-Wert $1 - \Phi(N)$,
  - für linksseitige Tests den p-Wert $\Phi(N)$,
  - für zweiseitige Tests den p-Wert

$$\left.\begin{array}{ll} 2 \cdot \Phi(N) = 2 \cdot (1 - \Phi(-N)) & \text{falls } N < 0 \\ 2 \cdot (1 - \Phi(N)) & \text{falls } N \geq 0 \end{array}\right\} = 2 \cdot (1 - \Phi(|N|))$$

  sowie die alternative Darstellung $2 \cdot \min\{\Phi(N), 1 - \Phi(N)\}$.

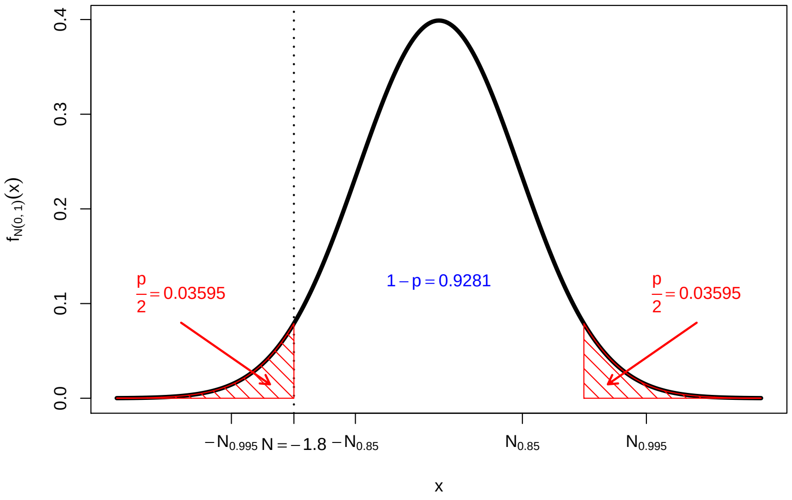
Beispiel: p-Werte bei rechtsseitigem Gauß-Test (Grafik)

Realisierte Teststatistik $N = 1.6$, p-Wert: 0.0548



Beispiel: p-Werte bei zweiseitigem Gauß-Test (Grafik)

Realisierte Teststatistik $N = -1.8$, p-Wert: 0.0719



Entscheidung mit p-Wert

- Offensichtlich erhält man auf der Grundlage des p-Werts p zur beobachteten Stichprobenrealisation die einfache Entscheidungsregel
$$H_0 \text{ ablehnen} \quad \Leftrightarrow \quad p < \alpha$$
für Hypothesentests zum Signifikanzniveau α.
- Sehr niedrige p-Werte bedeuten also, dass man beim zugehörigen Hypothesentest H_0 auch dann ablehnen würde, wenn man die maximale Fehlerwahrscheinlichkeit 1. Art sehr klein wählen würde.
- Kleinere p-Werte liefern also stärkere Indizien für die Gültigkeit von H_1 als größere, **aber** (wieder) Vorsicht vor Überinterpretation: Aussagen der Art *„Der p-Wert gibt die Wahrscheinlichkeit für die Gültigkeit von H_0 an"* sind unsinnig!

Warnung!

Bei der Entscheidung von statistischen Tests mit Hilfe des p-Werts ist es **unbedingt** erforderlich, das Signifikanzniveau α **vor** Berechnung des p-Werts festzulegen, um nicht der Versuchung zu erliegen, α im Nachhinein so zu wählen, dass man die „bevorzugte" Testentscheidung erhält!

Tests und Konfidenzintervalle

- Enger Zusammenhang zwischen zweiseitigem Gauß-Test und (symmetrischen) Konfidenzintervallen für den Erwartungswert bei bekannter Varianz.
- Für Konfidenzintervalle zur Vertrauenswahrscheinlichkeit $1-\alpha$ gilt:

$$\begin{aligned}
& \widetilde{\mu} \in \left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] \\
\Leftrightarrow \quad & \widetilde{\mu} - \overline{X} \in \left[-\frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}, \frac{\sigma}{\sqrt{n}} \cdot N_{1-\frac{\alpha}{2}}\right] \\
\Leftrightarrow \quad & \frac{\widetilde{\mu} - \overline{X}}{\sigma}\sqrt{n} \in \left[-N_{1-\frac{\alpha}{2}}, N_{1-\frac{\alpha}{2}}\right] \\
\Leftrightarrow \quad & \frac{\overline{X} - \widetilde{\mu}}{\sigma}\sqrt{n} \in \left[-N_{1-\frac{\alpha}{2}}, N_{1-\frac{\alpha}{2}}\right]
\end{aligned}$$

- Damit ist $\widetilde{\mu}$ also **genau dann** im Konfidenzintervall zur Sicherheitswahrscheinlichkeit $1-\alpha$ enthalten, **wenn** ein zweiseitiger Gauß-Test zum Signifikanzniveau α die Nullhypothese $H_0 : \mu = \widetilde{\mu}$ **nicht** verwerfen würde.
- Vergleichbarer Zusammenhang auch in anderen Situationen.

Zusammenfassung: Gauß-Test für den Mittelwert

bei bekannter Varianz

Anwendungs- voraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ unbekannt, σ^2 bekannt approximativ: $E(Y) = \mu \in \mathbb{R}$ unbekannt, $\text{Var}(Y) = \sigma^2$ bekannt $X_1, \ldots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$
Teststatistik	$N = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$		
Verteilung (H_0)	N für $\mu = \mu_0$ (näherungsweise) $N(0,1)$-verteilt		
Benötigte Größen	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$		
Kritischer Bereich zum Niveau α	$(-\infty, -N_{1-\frac{\alpha}{2}}) \cup (N_{1-\frac{\alpha}{2}}, \infty)$	$(N_{1-\alpha}, \infty)$	$(-\infty, -N_{1-\alpha})$
p-Wert	$2 \cdot (1 - \Phi(\lvert N \rvert))$	$1 - \Phi(N)$	$\Phi(N)$

Approximativer Gauß-Test für Anteilswert p

- Wichtiger Spezialfall des (approximativen) Gauß-Tests für den Mittelwert einer Zufallsvariablen mit bekannter Varianz:
  Approximativer Gauß-Test für den Anteilswert p einer alternativverteilten Zufallsvariablen
- *Erinnerung:* Für alternativverteilte Zufallsvariablen $Y \sim B(1, p)$ war Konfidenzintervall für Anteilswert p ein Spezialfall für Konfidenzintervalle für Mittelwerte von Zufallsvariablen mit **unbekannter** Varianz.
- **Aber:** Bei der Konstruktion von Tests für $H_0 : p = p_0$ gegen $H_1 : p \neq p_0$ für ein vorgegebenes p_0 (sowie den einseitigen Varianten) spielt Verteilung der Teststatistik unter H_0, insbesondere für $p = p_0$, entscheidende Rolle.
- Da Varianz für $p = p_0$ bekannt $\rightsquigarrow$ approximativer Gauß-Test geeignet.
  Für $p = p_0$ gilt genauer $\text{Var}(Y) = \text{Var}(X_i) = p_0 \cdot (1 - p_0)$ und damit

$$\text{Var}(\widehat{p}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n \cdot \text{Var}(Y) = \frac{p_0 \cdot (1 - p_0)}{n} .$$

Als Testgröße erhält man also: $N = \dfrac{\widehat{p} - p_0}{\sqrt{p_0 \cdot (1 - p_0)}}\sqrt{n}$

Zusammenfassung: (Approx.) Gauß-Test für Anteilswert p

Anwendungs-voraussetzungen	approximativ: $Y \sim B(1, p)$ mit $p \in [0, 1]$ unbekannt $X_1, \dots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0 : p = p_0$ $H_1 : p \neq p_0$	$H_0 : p \leq p_0$ $H_1 : p > p_0$	$H_0 : p \geq p_0$ $H_1 : p < p_0$
Teststatistik	$N = \dfrac{\widehat{p} - p_0}{\sqrt{p_0 \cdot (1 - p_0)}}\sqrt{n}$		
Verteilung (H_0)	N für $p = p_0$ näherungsweise $N(0, 1)$-verteilt		
Benötigte Größen	$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$		
Kritischer Bereich zum Niveau α	$(-\infty, -N_{1-\frac{\alpha}{2}})$ $\cup(N_{1-\frac{\alpha}{2}}, \infty)$	$(N_{1-\alpha}, \infty)$	$(-\infty, -N_{1-\alpha})$
p-Wert	$2 \cdot (1 - \Phi(\lvert N\rvert))$	$1 - \Phi(N)$	$\Phi(N)$

Beispiel: Bekanntheitsgrad eines Produkts

- Untersuchungsgegenstand: Hat sich der Bekanntheitsgrad eines Produkts gegenüber bisherigem Bekanntheitsgrad von 80% reduziert, nachdem die Ausgaben für Werbemaßnahmen vor einiger Zeit drastisch gekürzt wurden?
- Annahmen: Kenntnis des Produkts wird durch $Y \sim B(1, p)$ beschrieben, wobei p als Bekanntheitsgrad des Produkts aufgefasst werden kann.
- Stichprobeninformation aus Realisation einfacher Stichprobe **(!)** zu Y: Unter $n = 500$ befragten Personen kannten 381 das Produkt $\rightsquigarrow \widehat{p} = 0.762$.
- Gewünschtes Signifikanzniveau (max. Fehlerwahrscheinlichkeit 1. Art): $\alpha = 0.05$

Geeigneter Test: **(Approx.) linksseitiger Gauß-Test für den Anteilswert** p

1. Hypothesen: $H_0 : p \geq p_0 = 0.8$ gegen $H_1 : p < p_0 = 0.8$
2. Teststatistik: $N = \frac{\widehat{p} - p_0}{\sqrt{p_0 \cdot (1 - p_0)}}\sqrt{n} \overset{\bullet}{\sim} N(0, 1)$, falls H_0 gilt ($p = p_0$)
3. Kritischer Bereich zum Niveau $\alpha = 0.05$:
   $K = (-\infty, -N_{0.95}) = (-\infty, -1.645)$
4. Realisierter Wert der Teststatistik: $N = \frac{0.762 - 0.8}{\sqrt{0.8 \cdot (1 - 0.8)}}\sqrt{500} = -2.124$
5. Entscheidung: $N \in K \rightsquigarrow H_0$ wird abgelehnt, der Bekanntheitsgrad des Produkts hat sich signifikant reduziert.

t-Test für den Mittelwert

bei unbekannter Varianz

- Konstruktion des (exakten) Gauß-Tests für den Mittelwert bei bekannter Varianz durch Verteilungsaussage

$$N := \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0, 1) ,$$

  falls $X_1, \dots, X_n$ einfache Stichprobe zu normalverteilter ZV Y.
- Analog zur Konstruktion von Konfidenzintervallen für den Mittelwert bei unbekannter Varianz: Verwendung der Verteilungsaussage

$$t := \frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n - 1) \qquad \text{mit} \qquad S = \sqrt{\frac{1}{n - 1}\sum_{i=1}^{n}(X_i - \overline{X})^2} ,$$

  falls $X_1, \dots, X_n$ einfache Stichprobe zu normalverteilter ZV Y, um geeigneten Hypothesentest für den Mittelwert μ zu entwickeln.
- Test lässt sich genauso wie Gauß-Test herleiten, lediglich
  - Verwendung von S statt σ,
  - Verwendung von $t(n - 1)$ statt $N(0, 1)$.

- Beziehung zwischen symmetrischen Konfidenzintervallen und zweiseitigen Tests bleibt wie beim Gauß-Test erhalten.
- Wegen Symmetrie der $t(n-1)$-Verteilung bleiben auch alle entsprechenden „Vereinfachungen“ bei der Bestimmung von kritischen Bereichen und p-Werten gültig.
- p-Werte können mit Hilfe der Verteilungsfunktion der $t(n-1)$-Verteilung bestimmt werden (unproblematisch mit Statistik-Software).
- Zur Berechnung der Gütefunktion: Verteilungsfunktion der „nichtzentralen“ $t(n-1)$-Verteilung benötigt (unproblematisch mit Statistik-Software).
- Zur Berechnung von p-Werten und Gütefunktionswerten für große n: Näherung der $t(n-1)$-Verteilung durch Standardnormalverteilung bzw. der nichtzentralen $t(n-1)$-Verteilung durch Normalverteilung mit Varianz 1 (vgl. Gauß-Test) möglich.
- Analog zu Konfidenzintervallen:
  Ist Y nicht normalverteilt, kann der t-Test auf den Mittelwert bei unbekannter Varianz immer noch als approximativer (näherungsweiser) Test verwendet werden.

Zusammenfassung: t-Test für den Mittelwert

bei unbekannter Varianz

Anwendungs-voraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_{++}$ unbekannt approximativ: $E(Y) = \mu \in \mathbb{R}$, $\text{Var}(Y) = \sigma^2 \in \mathbb{R}_{++}$ unbekannt $X_1, \ldots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$
Teststatistik Verteilung (H_0)	$t = \frac{\overline{X} - \mu_0}{S}\sqrt{n}$ t für $\mu = \mu_0$ (näherungsweise) $t(n-1)$-verteilt		
Benötigte Größen	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n-1;1-\frac{\alpha}{2}})$ $\cup(t_{n-1;1-\frac{\alpha}{2}}, \infty)$	$(t_{n-1;1-\alpha}, \infty)$	$(-\infty, -t_{n-1;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n-1)}(\lvert t \rvert))$	$1 - F_{t(n-1)}(t)$	$F_{t(n-1)}(t)$

Beispiel: Durchschnittliche Wohnfläche

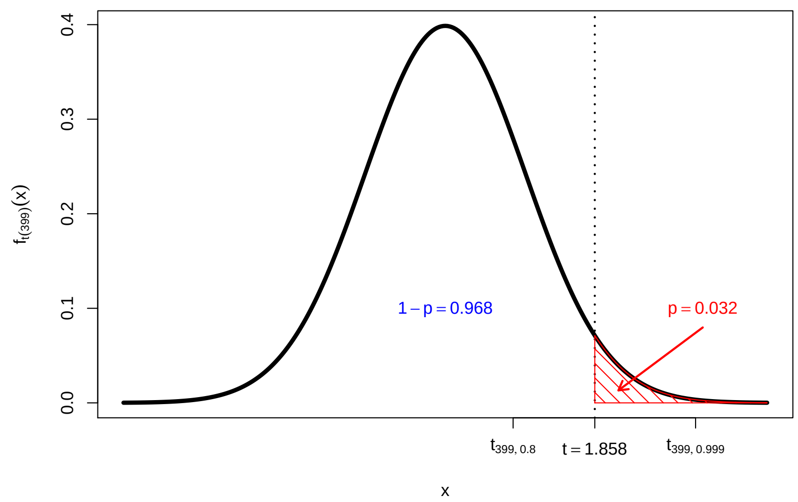
- Untersuchungsgegenstand: Hat sich die durchschnittliche Wohnfläche pro Haushalt in einer bestimmten Stadt gegenüber dem aus dem Jahr 1998 stammenden Wert von 71.2 (in $[m^2]$) **erhöht**?
- Annahmen: Verteilung der Wohnfläche Y im Jahr 2009 unbekannt.
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 400$ zu Y liefert Stichprobenmittel $\overline{x} = 73.452$ und Stichprobenstandardabweichung $s = 24.239$.
- Gewünschtes Signifikanzniveau (max. Fehlerwahrscheinlichkeit 1. Art): $\alpha = 0.05$

Geeigneter Test:
Rechtsseitiger approx. t-Test für den Mittelwert bei unbekannter Varianz

1. Hypothesen: $H_0: \mu \leq \mu_0 = 71.2$ gegen $H_1: \mu > \mu_0 = 71.2$
2. Teststatistik: $t = \frac{\overline{X} - \mu_0}{S}\sqrt{n} \overset{\bullet}{\sim} t(399)$, falls H_0 gilt ($\mu = \mu_0$)
3. Kritischer Bereich zum Niveau $\alpha = 0.05$: $K = (t_{399;0.95}, \infty) = (1.649, \infty)$
4. Realisierter Wert der Teststatistik: $t = \frac{73.452 - 71.2}{24.239}\sqrt{400} = 1.858$
5. Entscheidung: $t \in K \rightsquigarrow H_0$ wird abgelehnt; Test kommt zur Entscheidung, dass sich durchschnittliche Wohnfläche gegenüber 1998 erhöht hat.

Beispiel: p-Wert bei rechtsseitigem t-Test (Grafik)

Wohnflächenbeispiel, realisierte Teststatistik $t = 1.858$, p-Wert: 0.032



Die Familie der $\chi^2(n)$-Verteilungen

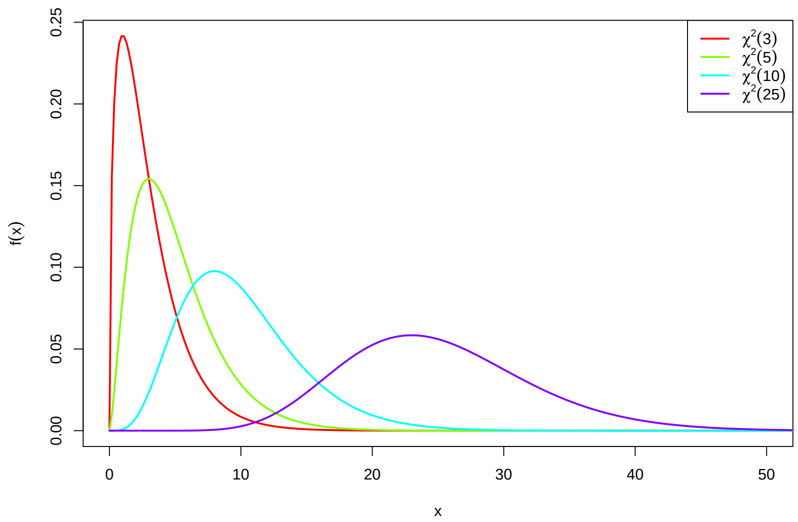
- Sind $Z_1, \ldots, Z_m$ für $m \geq 1$ unabhängig identisch standardnormalverteilte Zufallsvariablen, so genügt die Summe der *quadrierten* Zufallsvariablen

$$\chi^2 := \sum_{i=1}^{m} Z_i^2 = Z_1^2 + \ldots + Z_m^2$$

  einer sog. **Chi-Quadrat-Verteilung mit** m **Freiheitsgraden**, in Zeichen $\chi^2 \sim \chi^2(m)$.
- Offensichtlich können $\chi^2(m)$-verteilte Zufallsvariablen nur nichtnegative Werte annehmen, der Träger ist also $[0, \infty)$.
- Ist $\chi^2 \sim \chi^2(m)$, so gilt $\mathrm{E}(\chi^2) = m$ sowie $\mathrm{Var}(\chi^2) = 2m$.
- Als Abkürzung für α-Quantile der $\chi^2(m)$-Verteilung verwenden wir (wie üblich) $\chi^2_{m;\alpha}$.

Grafische Darstellung einiger $\chi^2(m)$-Verteilungen

für $m \in \{3, 5, 10, 25\}$



Tests für die Varianz

- Für Aussagen über die Varianz von Y (als mittlere quadrierte Abweichung vom Erwartungswert) auf Basis einer einfachen Stichprobe $X_1, \ldots, X_n$ zu Y naheliegend: Untersuchung der quadrierten Abweichungen

$$(X_1 - \mu)^2, \ldots, (X_n - \mu)^2$$

  bei bekanntem Erwartungswert $\mu = \mathrm{E}(Y)$ bzw. bei unbekanntem Erwartungswert der quadrierten Abweichungen vom Stichprobenmittelwert

$$(X_1 - \overline{X})^2, \ldots, (X_n - \overline{X})^2 \; .$$

- Man kann zeigen: Ist $Y \sim N(\mu, \sigma^2)$, so gilt

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(X_1 - \mu)^2}{\sigma^2} + \ldots + \frac{(X_n - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

  bzw. mit der Abkürzung $\widetilde{S}^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ für die mittlere quadratische Abweichung vom *bekannten* Erwartungswert aus der Stichprobe

$$\frac{n\widetilde{S}^2}{\sigma^2} \sim \chi^2(n) \; .$$

- Hieraus lassen sich analog zu den Tests für den Mittelwert Tests auf Abweichung der Varianz $\mathrm{Var}(Y)$ von einer vorgegebenen „Soll-Varianz“ σ_0^2 entwickeln:
  - **Über**schreitet die tatsächliche Varianz von Y die (unter H_0 angenommene) „Soll-Varianz“ σ_0^2, so verschiebt sich die Verteilung der Größe $\chi^2 := \frac{n\widetilde{S}^2}{\sigma_0^2}$ offensichtlich nach **rechts**.
  - **Unter**schreitet die tatsächliche Varianz von Y die (unter H_0 angenommene) „Soll-Varianz“ σ_0^2, so verschiebt sich die Verteilung der Größe $\chi^2 := \frac{n\widetilde{S}^2}{\sigma_0^2}$ offensichtlich nach **links**.
- Gilt $Y \sim N(\mu, \sigma^2)$ und ist der Erwartungswert μ *unbekannt*, so kann weiter gezeigt werden, dass

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} = \frac{(X_1 - \overline{X})^2}{\sigma^2} + \ldots + \frac{(X_n - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

  bzw. mit der bekannten Abkürzung $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für die Stichprobenvarianz

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

  gilt, woraus ebenfalls Tests für die Varianz abgeleitet werden können.

Bemerkungen

- Bei der Konstruktion der kritischen Bereich ist zu beachten, dass die Testgrößen
$$\chi^2 = \frac{n\widetilde{S}^2}{\sigma_0^2} \qquad \text{bzw.} \qquad \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$
nur nichtnegative Wert annehmen können.
- Durch die fehlende Symmetrie sind viele von Gauß- und t-Tests bekannte Vereinfachungen nicht mehr möglich. Insbesondere
  - darf $\chi^2_{m;\alpha}$ nicht durch $-\chi^2_{m;1-\alpha}$ ersetzt werden,
  - kann die Berechnung des p-Werts im zweiseitigen Test nicht vereinfacht werden.

Wichtig!

Die Normalverteilungsannahme $Y \sim N(\mu, \sigma^2)$ ist für den Chi-Quadrat-Test für die Varianz wesentlich. Weicht die Verteilung von Y „deutlich" von einer Normalverteilung ab, unterscheidet sich die Verteilung der Testgröße χ^2 (auch unter H_0 für $\sigma^2 = \sigma_0^2$!) wesentlich von einer $\chi^2(n)$ bzw. $\chi^2(n-1)$-Verteilung.

Quantile der χ^2-Verteilungen: $\chi^2_{n;p}$

$n\backslash p$	0.01	0.025	0.05	0.50	0.90	0.95	0.975	0.99
1	0.000	0.001	0.004	0.455	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	1.386	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	2.366	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	3.357	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	4.351	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	5.348	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	6.346	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	7.344	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	8.343	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	9.342	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	10.341	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	11.340	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	12.340	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	13.339	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	14.339	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	15.338	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	16.338	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	17.338	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	18.338	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	19.337	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	20.337	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	21.337	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	22.337	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	23.337	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	24.337	34.382	37.652	40.646	44.314

Zusammenfassung: χ^2-Test für die Varianz

einer normalverteilten Zufallsvariablen mit **bekanntem** Erwartungswert

Anwendungs-voraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ **bekannt**, $\sigma^2 \in \mathbb{R}_{++}$ unbekannt $X_1, \dots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 \geq \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$
Teststatistik Verteilung (H_0)	$\chi^2 = \frac{n \cdot \widetilde{S}^2}{\sigma_0^2}$ χ^2 (für $\sigma^2 = \sigma_0^2$) $\chi^2(n)$-verteilt		
Benötigte Größen	$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$		
Kritischer Bereich zum Niveau α	$[0, \chi^2_{n;\frac{\alpha}{2}})$ $\cup(\chi^2_{n;1-\frac{\alpha}{2}}, \infty)$	$(\chi^2_{n;1-\alpha}, \infty)$	$[0, \chi^2_{n;\alpha})$
p-Wert	$2 \cdot \min\left\{F_{\chi^2(n)}(\chi^2),\right.$ $\left.1 - F_{\chi^2(n)}(\chi^2)\right\}$	$1 - F_{\chi^2(n)}(\chi^2)$	$F_{\chi^2(n)}(\chi^2)$

Beispiel: Präzision einer Produktionsanlage

- Untersuchungsgegenstand: Bei einer Produktionsanlage für Maßbänder soll geprüft werden, ob die Herstellerangabe für die Produktionsgenauigkeit korrekt ist. Laut Hersteller ist die Länge der produzierten Maßbänder normalverteilt mit Erwartungswert 200 [mm] und Varianz $\sigma^2 = 0.1^2$. Der Betreiber der Anlage vermutet eine Abweichung der Präzision.
- Annahmen: Länge $Y \sim N(200, \sigma^2)$ mit σ^2 unbekannt.
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 16$ zu Y liefert $\widetilde{S}^2 = \frac{1}{16}\sum_{i=1}^{16}(X_i - 200)^2 = 0.019257$.
- Gewünschtes Signifikanzniveau: $\alpha = 0.10$

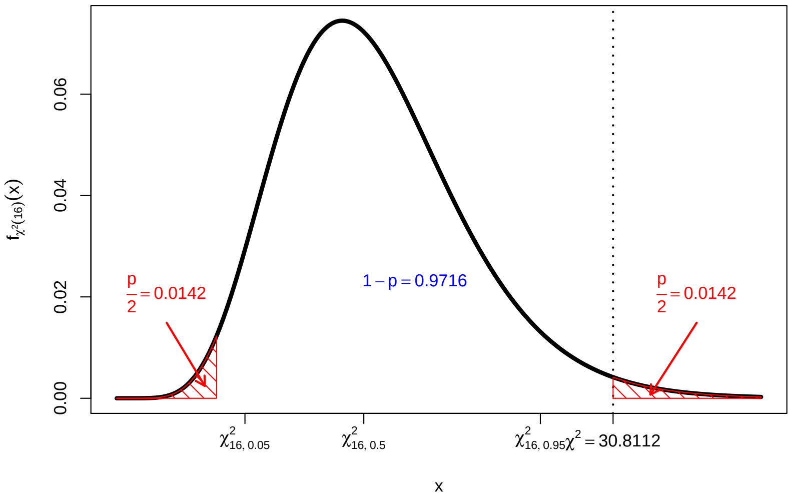
Geeigneter Test:

Zweiseitiger Chi-Quadrat-Test für Varianz bei bekanntem Erwartungswert

1. Hypothesen: $H_0: \sigma^2 = \sigma_0^2 = 0.1^2$ gegen $H_1: \sigma^2 \neq \sigma_0^2 = 0.1^2$
2. Teststatistik: $\chi^2 = \frac{n \cdot \widetilde{S}^2}{\sigma_0^2} \sim \chi^2(16)$, falls H_0 gilt ($\sigma^2 = \sigma_0^2$)
3. Kritischer Bereich zum Niveau $\alpha = 0.10$:
$K = [0, \chi^2_{16;0.05}) \cup (\chi^2_{16;0.95}, \infty) = [0, 7.962) \cup (26.296, \infty)$
4. Realisierter Wert der Teststatistik: $\chi^2 = \frac{16 \cdot 0.019257}{0.01} = 30.8112$
5. Entscheidung: $\chi^2 \in K \rightsquigarrow H_0$ wird abgelehnt; Test kommt zur Entscheidung, dass die Präzision von der Herstellerangabe abweicht.

Beispiel: p-Wert bei zweiseitigem χ^2-Test (Grafik)

Produktionsmaschinenbeispiel, realisierte Teststatistik $\chi^2 = 30.8112$, p-Wert: 0.0284



Zusammenfassung: χ^2-Test für die Varianz

einer normalverteilten Zufallsvariablen mit **unbekanntem** Erwartungswert

Anwendungs-voraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ **unbekannt**, $\sigma^2 \in \mathbb{R}_{++}$ unbekannt $X_1, \ldots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$
Teststatistik Verteilung (H_0)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ χ^2 (für $\sigma^2 = \sigma_0^2$) $\chi^2(n-1)$-verteilt		
Benötigte Größen	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$ mit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$		
Kritischer Bereich zum Niveau α	$[0, \chi^2_{n-1;\frac{\alpha}{2}})$ $\cup(\chi^2_{n-1;1-\frac{\alpha}{2}}, \infty)$	$(\chi^2_{n-1;1-\alpha}, \infty)$	$[0, \chi^2_{n-1;\alpha})$
p-Wert	$2 \cdot \min\left\{F_{\chi^2(n-1)}(\chi^2),\right.$ $\left.1 - F_{\chi^2(n-1)}(\chi^2)\right\}$	$1 - F_{\chi^2(n-1)}(\chi^2)$	$F_{\chi^2(n-1)}(\chi^2)$

Beispiel: Präzision einer neuen Abfüllmaschine

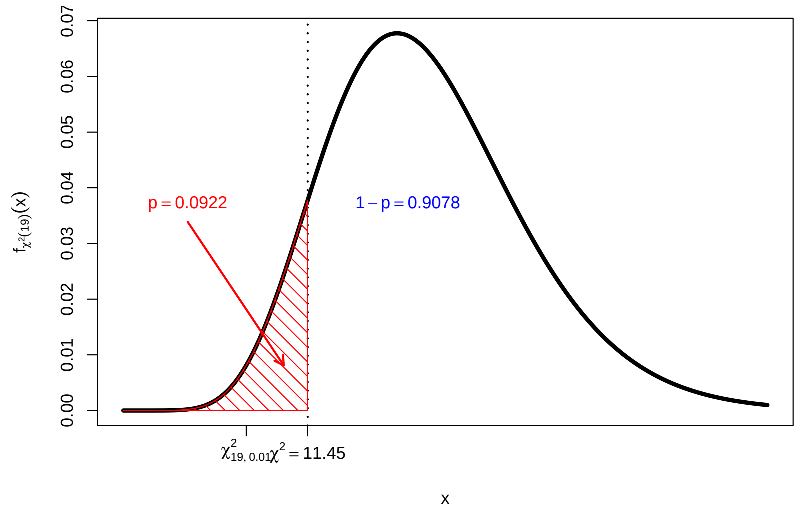
- Untersuchungsgegenstand: Für eine neue Abfüllmaschine wird geprüft, ob sie **präziser** als die alte Anlage arbeitet. Bei der alten Maschine beträgt die Standardabweichung des Füllgewichts um den eingestellten Wert 5 [g].
- Annahmen: Füllgewicht $Y \sim N(\mu, \sigma^2)$ mit μ, σ^2 unbekannt.
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 20$ zu Y liefert Stichprobenmittel $\overline{x} = 25.8097$ und mittleres Quadrat $\overline{x^2} = 680.4535$, damit also $s^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right) = 15.066$.
- Gewünschtes Signifikanzniveau: $\alpha = 0.01$

Geeigneter Test: Linksseitiger Chi-Quadrat-Test für Varianz bei unbekanntem Erwartungswert

1. Hypothesen: $H_0 : \sigma^2 \geq \sigma_0^2 = 5^2$ gegen $H_1 : \sigma^2 < \sigma_0^2 = 5^2$
2. Teststatistik: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(19)$, falls H_0 gilt ($\sigma^2 = \sigma_0^2$)
3. Kritischer Bereich zum Niveau $\alpha = 0.01$: $K = [0, \chi^2_{19;0.01}) = [0, 7.633)$
4. Realisierter Wert der Teststatistik: $\chi^2 = \frac{19 \cdot 15.066}{25} = 11.45$
5. Entscheidung: $\chi^2 \notin K \rightsquigarrow H_0$ wird **nicht** abgelehnt; Test kommt zur Entscheidung, dass es keine ausreichende statistische Evidenz für eine bessere Präzision der neueren Maschine gibt.

Beispiel: p-Wert bei linksseitigem χ^2-Test (Grafik)

Abfüllmaschinenbeispiel, realisierte Teststatistik $\chi^2 = 11.45$, p-Wert: 0.0922



Chi-Quadrat-Anpassungstest

- *Ziel:* Konstruktion eines Tests zur Überprüfung, ob Zufallsvariable Y einer bestimmten **Verteilung** (oder *später* allgemeiner: einer bestimmten Verteilungsklasse) folgt, **ohne** mögliche Verteilungen von Y bereits durch (parametrische) Verteilungsannahme eingrenzen zu müssen.
- *Eine* Möglichkeit: **Chi-Quadrat-Anpassungstest**
- *Grundlegende Idee:* Vergleich der **empirischen Häufigkeitsverteilung** aus der Stichprobenrealisation $(X_1, \dots, X_n)$ mit den **(theoretischen) Wahrscheinlichkeiten** der **hypothetischen** (d.h. unter H_0 angenommenen) Verteilung von Y.
- Hierzu nötig:
  - Erstellen der empirischen Häufigkeitsverteilung — bei diskreter hypothetischer Verteilung mit „vielen" Trägerpunkten bzw. stetiger hypothetischer Verteilung nach erfolgter Klassierung —
  - Berechnen der theoretischen Punkt- bzw. Klassenwahrscheinlichkeiten unter der hypothetischen Verteilung.
- Offensichtlich: **Große** Abweichungen der empirischen (in der Stichprobe beobachteten) Häufigkeiten von den theoretischen Wahrscheinlichkeiten sprechen eher **gegen** die hypothetische Verteilung von Y, **kleine** Abweichungen eher **dafür**.

- Noch nötig: Geeignete Testgröße zur Zusammenfassung der Abweichungen sowie Verteilungsaussage für die Testgröße bei Gültigkeit von H_0.
- $(X_1, \dots, X_n)$ sei (wie immer) eine einfache Stichprobe vom Umfang n zu Y.
- Bezeichnen
  - k die Anzahl der Ausprägungen bzw. Klassen der empirischen Häufigkeitsverteilung,
  - n_i für $i \in \{1, \dots, k\}$ die in der Stichprobe aufgetretenen (absoluten) Häufigkeiten für Ausprägung i bzw. Klasse i,
  - p_i^0 die bei Gültigkeit der hypothetischen Verteilung für Y tatsächlichen Wahrscheinlichkeiten für Ausprägung i bzw. Klasse i,

  so werden die Abweichungen $\frac{n_i}{n} - p_i^0$ (beim Vergleich relativer Häufigkeiten und Wahrscheinlichkeiten) bzw. $n_i - np_i^0$ (beim Vergleich absoluter Häufigkeiten und erwarteter Häufigkeiten) mit der Testgröße
  $$\chi^2 := n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \sum_{i=1}^{k} \frac{\left(n_i - np_i^0\right)^2}{np_i^0}$$
  zusammengefasst.
- Ist H_0 gültig, so konvergiert die Verteilung von χ^2 mit wachsendem n gegen die $\chi^2(k-1)$-Verteilung.

- Offensichtlich: Große Werte von χ^2 entstehen bei großen Abweichungen zwischen beobachteten Häufigkeiten und Wahrscheinlichkeiten bzw. erwarteten Häufigkeiten und sprechen damit gegen H_0.
- Sinnvoller kritischer Bereich zum Signifikanzniveau α also $(\chi^2_{k-1;1-\alpha}; \infty)$.
- χ^2-Anpassungstest ist immer approximativer (näherungsweiser) Test. Vernünftige Näherung der Verteilung von χ^2 (unter H_0) durch $\chi^2(k-1)$-Verteilung kann nur erwartet werden, wenn $np_i^0 \geq 5$ für alle $i \in \{1, \dots, k\}$ gilt.
- Berechnung der p_i^0 zur Durchführung des Chi-Quadrat-Anpassungstest je nach Anwendung sehr unterschiedlich:
  - Bei diskreter hypothetischer Verteilung mit endlichem Träger in der Regel (falls $np_i^0 \geq 5$ für alle $i \in \{1, \dots, k\}$) besonders einfach, da keine Klassierung erforderlich ist und sich alle p_i^0 direkt als Punktwahrscheinlichkeiten ergeben.
  - Bei diskreter hypothetischer Verteilung mit unendlichem Träger bzw. bei Verletzung der Bedingung $np_i^0 \geq 5$ für alle $i \in \{1, \dots, k\}$ Klassierung (trotz diskreter Verteilung) erforderlich, so dass Bedingung erfüllt wird.
  - Bei stetiger hypothetischer Verteilung Klassierung stets erforderlich; Durchführung so, dass Bedingung $np_i^0 \geq 5$ für alle $i \in \{1, \dots, k\}$ erfüllt ist.
- Sobald p_i^0 (ggf. nach Klassierung) bestimmt sind, identische Vorgehensweise für alle Verteilungen.

Chi-Quadrat-Anpassungstest

zur Anpassung an eine hypothetische Verteilung

- Hypothesenformulierung z.B. über Verteilungsfunktion F_0 der hypothetischen Verteilung in der Form:
  $$H_0 : F_Y = F_0 \qquad \text{gegen} \qquad H_1 : F_Y \neq F_0$$
- Allgemeine Vorgehensweise: Bilden von k Klassen durch Aufteilen der reellen Zahlen in k Intervalle
  $$K_1 = (-\infty, a_1], K_2 = (a_1, a_2], \dots, K_{k-1} = (a_{k-2}, a_{k-1}], K_k = (a_{k-1}, \infty)$$
  und Berechnen der theoretischen Klassenwahrscheinlichkeiten p_i^0 als $p_i^0 = F_0(a_k) - F_0(a_{k-1})$ mit $a_0 := -\infty$ und $a_k := \infty$, also
  $$\begin{aligned} p_1^0 &= F_0(a_1) - F_0(-\infty) = F_0(a_1), \\ p_2^0 &= F_0(a_2) - F_0(a_1), \\ &\vdots \\ p_{k-1}^0 &= F_0(a_{k-1}) - F_0(a_{k-2}), \\ p_k^0 &= F_0(\infty) - F_0(a_{k-1}) = 1 - F_0(a_{k-1}) . \end{aligned}$$

Zusammenfassung: Chi-Quadrat-Anpassungstest

zur Anpassung an **eine** vorgegebene Verteilung

Anwendungsvoraussetzungen	approximativ: Y beliebig verteilt $X_1, \ldots, X_n$ einfache Stichprobe zu Y $k-1$ Klassengrenzen $a_1 < a_2 < \ldots < a_{k-1}$ vorgegeben
Nullhypothese Gegenhypothese	$H_0 : F_Y = F_0$ $H_1 : F_Y \neq F_0$
Teststatistik Verteilung (H_0)	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} = n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \left(\frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_i^0}\right) - n$ χ^2 ist näherungsweise $\chi^2(k-1)$-verteilt, falls $F_Y = F_0$ (Näherung nur vernünftig, falls $np_i^0 \geq 5$ für $i \in \{1, \ldots, k\}$)
Benötigte Größen	$p_i^0 = F_0(a_i) - F_0(a_{i-1})$ mit $a_0 := -\infty, a_k := \infty$, $n_i = \#\{j \in \{1, \ldots, n\} \mid x_j \in (a_{i-1}, a_i]\}, i \in \{1, \ldots, k\}$
Kritischer Bereich zum Niveau α	$(\chi^2_{k-1;1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2(k-1)}(\chi^2)$

Vereinfachung bei diskreter hypothetischer Verteilung

mit k Trägerpunkten

- Einfachere „Notation“ bei Anwendung des Chi-Quadrat-Anpassungstests meist möglich, falls hypothetische Verteilung diskret mit k Trägerpunkten $a_1, \ldots, a_k$.
- Bezeichnet p_0 die Wahrscheinlichkeitsfunktion der hypothetischen Verteilungen und gilt $n \cdot p_0(a_i) \geq 5$ für alle $i \in \{1, \ldots, k\}$, so ist keine „echte“ Klassierung erforderlich (1 Trägerpunkt pro „Klasse“).
- Man erhält dann die hypothetischen Punktwahrscheinlichkeiten p_i^0 als $p_i^0 = p_0(a_i)$.
- Hypothesen meist direkt über Vektor der Punktwahrscheinlichkeiten $p := (p_1, \ldots, p_k) := (p_Y(a_1), \ldots, p_Y(a_k))$ in der Form:
$$H_0 : p = (p_1, \ldots, p_k) = (p_1^0, \ldots, p_k^0) =: p^0 \quad \text{gegen} \quad H_1 : p \neq p^0$$
- Chi-Quadrat-Anpassungstest kann so auch auf „Merkmale“ angewendet werden, deren Ausprägungen noch nicht „Zufallsvariablen-konform“ durch (reelle) Zahlenwerte ausgedrückt (kodiert) worden sind, beispielsweise bei
  - Wochentagen: a_1=„Montag“, a_2=„Dienstag“, ...
  - Produktmarken: a_1=„Automarke A“, a_2=„Automarke B“, ...
  - Monaten: a_1=„Januar“, a_2=„Februar“, ...

Beispiel: Verteilung Auftragseingänge

auf 5 Wochentage Montag–Freitag (diskrete hypothetische Verteilung)

- Untersuchungsgegenstand: Sind die Auftragseingänge in einem Unternehmen gleichmäßig auf die 5 Arbeitstage Montag–Freitag verteilt, d.h, ist der Anteil der Auftragseingänge an jedem Wochentag gleich 0.2?
[$\rightsquigarrow p^0 = (0.2, 0.2, 0.2, 0.2, 0.2)$]
- Stichprobeninformation: Einfache Stichprobe von 400 Auftragseingängen liefert folgende Verteilung auf Wochentage:

	Mo	Di	Mi	Do	Fr
n_i	96	74	92	81	57

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: Chi-Quadrat-Anpassungstest

1. **Hypothesen:**
$$H_0 : p = p^0 = (0.2, 0.2, 0.2, 0.2, 0.2) \qquad H_1 : p \neq p^0$$
2. **Teststatistik:**
$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} \text{ ist unter } H_0 \text{ approximativ } \chi^2(k-1)\text{-verteilt;}$$
Näherung vernünftig, falls $np_i^0 \geq 5$ für alle i gilt.

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
$K = (\chi^2_{k-1;1-\alpha}, +\infty) = (\chi^2_{4;0.95}, +\infty) = (9.488, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

a_i	n_i	p_i^0	np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
Mo	96	0.2	80	3.2000
Di	74	0.2	80	0.4500
Mi	92	0.2	80	1.8000
Do	81	0.2	80	0.0125
Fr	57	0.2	80	6.6125
Σ	400	1	400	$\chi^2 = 12.0750$

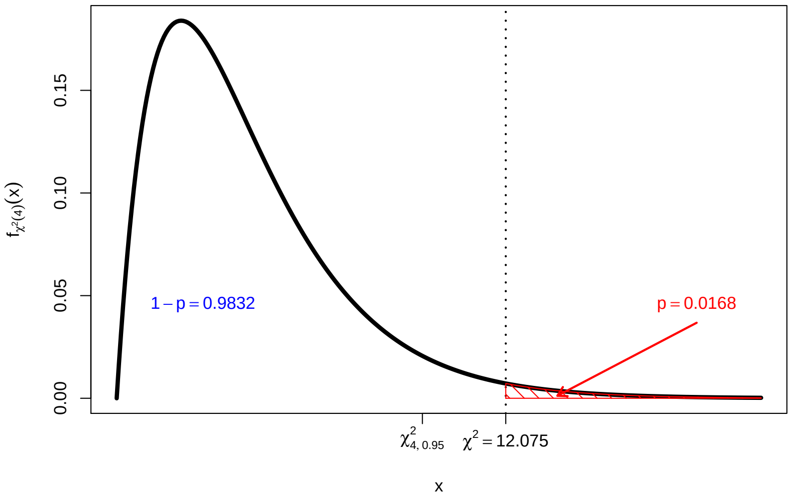
Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 5\} \rightsquigarrow$ Näherung ok.

5. **Entscheidung:**
$\chi^2 = 12.075 \in (9.488, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
(p-Wert: $1 - F_{\chi^2(4)}(\chi^2) = 1 - F_{\chi^2(4)}(12.075) = 1 - 0.9832 = 0.0168$)

Test kommt zur Entscheidung, dass die Auftragseingänge nicht gleichmäßig auf alle 5 Arbeitstage (Montag-Freitag) verteilt sind.

Beispiel: p-Wert bei Chi-Quadrat-Anpassungstest (Grafik)

Auftragseingangsbeispiel, realisierte Teststatistik $\chi^2 = 12.075$, p-Wert: 0.0168



Beispiel: Chi-Quadrat-Anpassungstest auf $H_0 : Y \sim \text{Geom}(0.25)$

- Geom(0.25)-Verteilung hat unendlichen Träger $\{0, 1, 2, \ldots\}$ und Wahrscheinlichkeitsfunktion

$$p_{\text{Geom}(0.25)} : \mathbb{N}_0 \to [0, 1]; p_{\text{Geom}(0.25)}(i) = (1 - 0.25)^i \cdot 0.25 \ ,$$

  Bedingung $np_i^0 \geq 5$ kann also mit $p_i^0 = p_{\text{Geom}(0.25)}(a_i)$ für $a_i := i - 1$ nicht für alle $i \in \mathbb{N}$ erfüllt sein.
- Klassierung hier also (trotz diskreter Verteilung) erforderlich. Wegen (für wachsendes i bzw. a_i) abnehmender p_i^0 sinnvoll: Zusammenfassung aller „großen" i in der letzten Klasse K_k so, dass Bedingung $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, k\}$ erfüllt ist.
- Wahrscheinlichkeit (unter H_0) p_k^0 für Klasse K_k über Verteilungsfunktion oder als verbleibende Wahrscheinlichkeit $p_k^0 = 1 - \sum_{i=1}^{k-1} p_i^0$.
- Je nach Verteilung F_0 und Stichprobenumfang n können aber auch komplexere Klassierungen nötig sein, um Bedingung $np_i^0 \geq 5$ zu erfüllen.

Fortsetzung Beispiel

- Stichprobeninformation: Häufigkeitsverteilung aus Klassierung einer einfachen Stichprobe vom Umfang $n = 100$ zu Y liefert:

i	1	2	3	4	5	6
a_i	0	1	2	3	4	≥ 5
n_i	32	19	16	16	6	11

- Gewünschtes Signifikanzniveau: $\alpha = 0.10$

Chi-Quadrat-Anpassungstest:

1. **Hypothesen:**
   $H_0 : F_Y = F_{\text{Geom}(0.25)} \qquad H_1 : F_Y \neq F_{\text{Geom}(0.25)}$
2. **Teststatistik:**
   $$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$$ ist unter H_0 approximativ $\chi^2(k-1)$-verteilt, falls $np_i^0 \geq 5$ für alle i gilt.

3. **Kritischer Bereich zum Niveau $\alpha = 0.10$:**
   $K = (\chi^2_{k-1;1-\alpha}, +\infty) = (\chi^2_{5;0.90}, +\infty) = (9.236, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

K_i	n_i	p_i^0	np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty, 0]$	32	$(1-0.25)^0 \cdot 0.25 = 0.25$	25.00	1.9600
$(0, 1]$	19	$(1-0.25)^1 \cdot 0.25 = 0.1875$	18.75	0.0033
$(1, 2]$	16	$(1-0.25)^2 \cdot 0.25 = 0.1406$	14.06	0.2677
$(2, 3]$	16	$(1-0.25)^3 \cdot 0.25 = 0.1055$	10.55	2.8154
$(3, 4]$	6	$(1-0.25)^4 \cdot 0.25 = 0.0791$	7.91	0.4612
$(4, +\infty)$	11	$1 - \sum_{i=1}^{5} p_i^0 = 0.2373$	23.73	6.8290
Σ	100	1	100	$\chi^2 = 12.3366$

   Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 6\}$ $\rightsquigarrow$ Näherung ok.
5. **Entscheidung:**
   $\chi^2 = 12.3366 \in (9.236, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
   (p-Wert: $1 - F_{\chi^2(5)}(\chi^2) = 1 - F_{\chi^2(5)}(12.3366) = 1 - 0.9695 = 0.0305$)

Test kommt zum Ergebnis, dass Y **nicht** einer Geom(0.25)-Verteilung genügt.

Beispiel: Chi-Quadrat-Anpassungstest (F_0 stetig)

- Klassierung bei stetigen hypothetischen Verteilungen unbedingt erforderlich.
- Hier: Klassierung soll vorgegeben sein (evtl. implizit durch bereits klassierte Stichprobeninformation statt vollständiger Urliste!)
- Bei eigener Wahl der Klassierung: Vorsicht, da Klassierung Test beeinflusst!
- Beispiel: Untersuchung, ob $Y \sim N(0,1)$.
- Stichprobeninformation (aus einfacher Stichprobe vom Umfang $n = 200$):

i	1	2	3	4	5	6
K_i	$(-\infty, -1.5]$	$(-1.5, -0.75]$	$(-0.75, 0]$	$(0, 0.75]$	$(0.75, 1.5]$	$(1.5, \infty)$
n_i	9	26	71	51	30	13

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: Chi-Quadrat-Anpassungstest

1. **Hypothesen:**
   $H_0 : F_Y = F_{N(0,1)} \qquad H_1 : F_Y \neq F_{N(0,1)}$
2. **Teststatistik:**
   $$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$$ ist unter H_0 approximativ $\chi^2(k-1)$-verteilt, falls $np_i^0 \geq 5$ für alle i gilt.

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (\chi^2_{k-1;1-\alpha}, +\infty) = (\chi^2_{5;0.95}, +\infty) = (11.070, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

$K_i = (a_{i-1}, a_i]$	n_i	$p_i^0 = F_0(a_i) - F_0(a_{i-1})$	np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty, -1.5]$	9	$0.0668 - 0 = 0.0668$	13.36	1.4229
$(-1.5, -0.75]$	26	$0.2266 - 0.0668 = 0.1598$	31.96	1.1114
$(-0.75, 0]$	71	$0.5 - 0.2266 = 0.2734$	54.68	4.8709
$(0, 0.75]$	51	$0.7734 - 0.5 = 0.2734$	54.68	0.2477
$(0.75, 1.5]$	30	$0.9332 - 0.7734 = 0.1598$	31.96	0.1202
$(1.5, +\infty)$	13	$1 - 0.9332 = 0.0668$	13.36	0.0097
Σ	200	1	200	7.7828

   Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 6\}$ $\rightsquigarrow$ Näherung ok.
5. **Entscheidung:**
   $\chi^2 = 7.7828 \notin (11.070, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
   (p-Wert: $1 - F_{\chi^2(5)}(\chi^2) = 1 - F_{\chi^2(5)}(7.7828) = 1 - 0.8314 = 0.1686$)

Test kann Hypothese, dass Y standardnormalverteilt ist, nicht verwerfen.

Chi-Quadrat-Anpassungstest auf parametrisches Verteilungsmodell

- Chi-Quadrat-Anpassungstest kann auch durchgeführt werden, wenn statt (einzelner) hypothetischer Verteilung eine parametrische Klasse von Verteilungen als hypothetische Verteilungsklasse fungiert.
- Durchführung des Chi-Quadrat-Anpassungstests dann in zwei Schritten:
  1. Schätzung der Verteilungsparameter innerhalb der hypothetischen Verteilungsklasse mit der ML-Methode.
  2. Durchführung des (regulären) Chi-Quadrat-Anpassungstest mit der hypothetischen Verteilung zu den geschätzen Parametern.
- Zu beachten:
  - **Verteilung der Testgröße χ^2 ändert sich!** Bei ML-Schätzung auf Basis der für die Durchführung des Chi-Quadrat-Anpassungstest maßgeblichen Klassierung der Stichprobe gilt unter H_0 näherungsweise $\chi^2 \sim \chi^2(k-r-1)$, wobei r die Anzahl der per ML-Methode geschätzten Parameter ist.
  - Werden die Verteilungsparameter nicht aus den klassierten Daten, sondern aus den ursprünglichen Daten mit ML-Methode geschätzt, gilt diese Verteilungsaussage so nicht mehr (Abweichung allerdings moderat).

Zusammenfassung: Chi-Quadrat-Anpassungstest

zur Anpassung an parametrische Verteilungs**familie**

Anwendungs-voraussetzungen	approx.: Y beliebig verteilt, $X_1, \ldots, X_n$ einf. Stichprobe zu Y Familie von Verteilungsfunktionen F_θ für $\theta \in \Theta$ vorgegeben $k-1$ Klassengrenzen $a_1 < a_2 < \ldots < a_{k-1}$ vorgegeben
Nullhypothese Gegenhypothese	$H_0 : F_Y = F_\theta$ für ein $\theta \in \Theta$ $H_1 : F_Y \neq F_\theta$ (für alle $\theta \in \Theta$)
Teststatistik	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} = n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \left(\frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_i^0}\right) - n$
Verteilung (H_0)	χ^2 ist unter H_0 näherungsweise $\chi^2(k-r-1)$-verteilt, wenn $\widehat{\theta}$ ML-Schätzer des r-dim. Verteilungsparameters θ auf Basis klassierter Daten ist (Verwendung von $\widehat{\theta}$ siehe unten). (Näherung nur vernünftig, falls $np_i^0 \geq 5$ für $i \in \{1, \ldots, k\}$)
Benötigte Größen	$p_i^0 = F_{\widehat{\theta}}(a_i) - F_{\widehat{\theta}}(a_{i-1})$ mit $a_0 := -\infty, a_k := \infty$, $n_i = \#\{j \in \{1, \ldots, n\} \mid x_j \in (a_{i-1}, a_i]\}, i \in \{1, \ldots, k\}$
Kritischer Bereich zum Niveau α	$(\chi^2_{k-r-1;1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2(k-r-1)}(\chi^2)$

Beispiel: Chi-Quadrat-Anpassungstest auf $H_0 : Y \sim \text{Geom}(p)$ für $p \in (0, 1)$

- Stichprobeninformation: Häufigkeitsverteilung aus vorangegangenem Beispiel:

i	1	2	3	4	5	6
a_i	0	1	2	3	4	≥ 5
n_i	32	19	16	16	6	11

- **Erster Schritt:**
  ML-Schätzung von p mit Hilfe der klassierten Stichprobeninformation:
  - Man kann zeigen, dass der ML-Schätzer auf Basis der klassierten Stichprobe durch
    $$\widehat{p} = \frac{n - n_k}{n - n_k + \sum_{i=1}^{k}(i-1) \cdot n_i}$$
    gegeben ist.
  - Hier erhält man also die Realisation
    $$\widehat{p} = \frac{100 - 11}{100 - 11 + 0 \cdot 32 + 1 \cdot 19 + 2 \cdot 16 + 3 \cdot 16 + 4 \cdot 6 + 5 \cdot 11} = \frac{89}{267} = 0.3333$$

- **Zweiter Schritt:**
  Durchführung des Chi-Quadrat-Anpassungstest für $H_0 : F_Y = F_{0.3333}$ (mit $F_p := F_{\text{Geom}(p)}$) gegen $H_1 : F_Y \neq F_{0.3333}$ **unter Berücksichtigung der ML-Schätzung von p durch geänderte Verteilung von χ^2 unter H_0!**

Insgesamt: Chi-Quadrat-Anpassungstest für Verteilungsfamilie:

1. **Hypothesen:**
   $H_0 : F_Y = F_p$ für ein $p \in (0, 1)$ (mit $F_p := F_{\text{Geom}(p)}$)   gegen $H_1 : F_Y \neq F_p$
2. **Teststatistik:**
   $$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$$ ist unter H_0 approximativ $\chi^2(k-1-r)$-verteilt, falls $np_i^0 \geq 5$ für alle i gilt und r-dimensionaler Verteilungsparameter per ML-Methode aus den klassierten Daten geschätzt wurde.
3. **Kritischer Bereich zum Niveau $\alpha = 0.10$:**
   $K = (\chi^2_{k-1-r;1-\alpha}, +\infty) = (\chi^2_{4;0.90}, +\infty) = (7.779, +\infty)$

4. **Berechnung der realisierten Teststatistik:**
   Eine ML-Schätzung aus den klassierten Daten liefert den Schätzwert $\widehat{p} = 0.3333$ für den unbekannten Verteilungsparameter p.

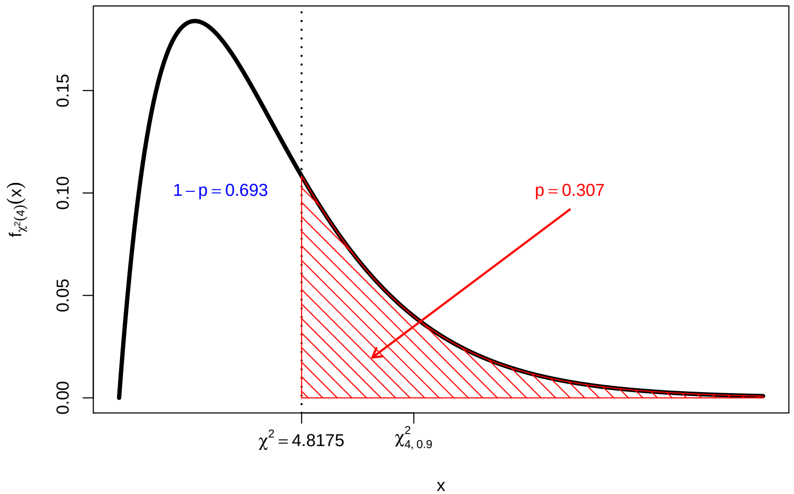
K_i	n_i	p_i^0	np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty, 0]$	32	$(1 - 0.3333)^0 \cdot 0.3333 = 0.3333$	33.33	0.0531
$(0, 1]$	19	$(1 - 0.3333)^1 \cdot 0.3333 = 0.2223$	22.23	0.4693
$(1, 2]$	16	$(1 - 0.3333)^2 \cdot 0.3333 = 0.1481$	14.81	0.0956
$(2, 3]$	16	$(1 - 0.3333)^3 \cdot 0.3333 = 0.0988$	9.88	3.7909
$(3, 4]$	6	$(1 - 0.3333)^4 \cdot 0.3333 = 0.0658$	6.58	0.0511
$(4, +\infty)$	11	$1 - \sum_{i=1}^{5} p_i^0 = 0.1317$	13.17	0.3575
Σ	100	1	100	$\chi^2 = 4.8175$

   Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 6\}$ $\rightsquigarrow$ Näherung ok.
5. **Entscheidung:**
   $\chi^2 = 4.8175 \notin (7.779, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
   (p-Wert: $1 - F_{\chi^2(4)}(\chi^2) = 1 - F_{\chi^2(4)}(4.8175) = 1 - 0.6935 = 0.3065$)

Test kommt zum Ergebnis, dass $Y \sim \text{Geom}(p)$ nicht verworfen werden kann.
(ML-Schätzung von p: $\widehat{p} = 0.3333$)

Beispiel: p-Wert bei Chi-Quadrat-Anpassungstest (Grafik)

Test auf geometrische Verteilung, realisierte Teststatistik $\chi^2 = 4.8175$, p-Wert: 0.307



Chi-Quadrat-Unabhängigkeitstest (Kontingenztest)

- *Bisher:* Einfache Stichprobe $X_1, \ldots, X_n$ zu **einer** Zufallsvariablen Y.
- *Im Folgenden:* Betrachtung von einfachen Stichproben zu mehrdimensionalen Zufallsvariablen bzw. (später) mehreren (unabhängigen) einfachen Stichproben zu mehreren Zufallsvariablen.
- Erste Problemstellung: **Untersuchung** von zwei Zufallsvariablen Y^A, Y^B **auf stochastische Unabhängigkeit**.
- Erforderliche Stichprobeninformation: Einfache Stichprobe

$$(X_1^A, X_1^B), (X_2^A, X_2^B), \ldots, (X_n^A, X_n^B)$$

vom Umfang n zu zweidimensionaler Zufallsvariable (Y^A, Y^B).
- *Testidee:* den bei Unabhängigkeit von Y^A, Y^B bestehenden Zusammenhang zwischen Randverteilungen von Y^A und Y^B sowie gemeinsamer Verteilung von (Y^A, Y^B) ausnutzen:
  - Gemeinsame Wahrscheinlichkeiten stimmen bei Unabhängigkeit mit Produkt der Randwahrscheinlichkeiten überein (falls (Y^A, Y^B) diskret).
  - Daher sprechen geringe Abweichungen zwischen gemeinsamen (relativen) Häufigkeiten und Produkt der (relativen) Randhäufigkeiten für Unabhängigkeit, große Abweichungen dagegen.

- Betrachtete Anwendungssituationen:
  1. Sowohl Y^A als auch Y^B sind diskret mit „wenigen" Ausprägungen, in der Stichprobe treten die Ausprägungen $a_1, \ldots, a_k$ von Y^A bzw. $b_1, \ldots, b_l$ von Y^B auf.
  2. Y^A und Y^B sind diskret mit „vielen" Ausprägungen oder stetig, die Stichprobeninformation wird dann mit Hilfe von Klassierungen $A_1 = (-\infty, a_1], A_2 = (a_1, a_2], \ldots, A_k = (a_{k-1}, \infty)$ von Y^A bzw. $B_1 = (-\infty, b_1], B_2 = (b_1, b_2], \ldots, B_l = (b_{l-1}, \infty)$ von Y^B zusammengefasst.
  3. Mischformen von 1 und 2.
- Der Vergleich zwischen (in der Stichprobe) **beobachteten** gemeinsamen absoluten Häufigkeiten n_{ij} und **bei Unabhängigkeit** (auf Basis der Randhäufigkeiten) **zu erwartenden** gemeinsamen absoluten Häufigkeiten $\widetilde{n}_{ij}$ erfolgt durch die Größe

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} ,$$

wobei n_{ij} die beobachteten gemeinsamen Häufigkeiten für (a_i, b_j) bzw. (A_i, B_j) aus der Stichprobenrealisation und $\widetilde{n}_{ij} = n \cdot \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ die erwarteten gemeinsamen Häufigkeiten aus den Randhäufigkeiten $n_{i\cdot}$ von a_i bzw. A_i und $n_{\cdot j}$ von b_j bzw. B_j sind $(i \in \{1, \ldots, k\}, j \in \{1, \ldots, l\})$.

- Für wachsenden Stichprobenumfang n konvergiert die Verteilung der Testgröße χ^2 bei Gültigkeit von

$$H_0 : Y^A, Y^B \text{ sind stochastisch unabhängig}$$

gegen die $\chi^2((k-1) \cdot (l-1))$-Verteilung.
- Die Näherung der Verteilung von χ^2 unter H_0 ist für endlichen Stichprobenumfang n vernünftig, falls gilt:

$$\widetilde{n}_{ij} \geq 5 \text{ für alle } i \in \{1, \ldots, k\}, j \in \{1, \ldots, l\}$$

- Wie beim Chi-Quadrat-Anpassungstest sprechen **große** Werte der Teststatistik χ^2 **gegen** die Nullhypothese „Y^A und Y^B sind stochastisch unabhängig", während kleine Werte für H_0 sprechen.
- Als kritischer Bereich zum Signifikanzniveau α ergibt sich also entsprechend:

$$K = (\chi^2_{(k-1)\cdot(l-1);1-\alpha}, \infty)$$

- Die Testgröße χ^2 ist eng verwandt mit der bei der Berechnung des korrigierten Pearsonschen Kontingenzkoeffizienten benötigten Größe χ^2.
- Analog zum Chi-Quadrat-Anpassungstest kann der Chi-Quadrat-Unabhängigkeitstest ebenfalls auf „Merkmale" Y^A bzw. Y^B angewendet werden, deren Ausprägungen $a_1, \ldots, a_k$ bzw. $b_1, \ldots, b_l$ noch nicht „Zufallsvariablen-konform" als reelle Zahlen „kodiert" wurden.

- Darstellung der Stichprobeninformation üblicherweise in Kontingenztabelle der Form

$Y^A \setminus Y^B$	b_1	b_2	$\cdots$	b_l
a_1	n_{11}	n_{12}	$\cdots$	n_{1l}
a_2	n_{21}	n_{22}	$\cdots$	n_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}

bzw.

$Y^A \setminus Y^B$	B_1	B_2	$\cdots$	B_l
A_1	n_{11}	n_{12}	$\cdots$	n_{1l}
A_2	n_{21}	n_{22}	$\cdots$	n_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}

.

- Benötigte Größen $\widetilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ können dann — nach Ergänzung der Kontingenztabelle um ihre Randhäufigkeiten $n_{i\cdot} = \sum_{j=1}^{l} n_{ij}$ und $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$ — in weiterer Tabelle mit analogem Aufbau

$Y^A \setminus Y^B$	B_1	B_2	$\cdots$	B_l	$n_{i\cdot}$
A_1	$\widetilde{n}_{11} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{12} = \frac{n_{1\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{1l} = \frac{n_{1\cdot} \cdot n_{\cdot l}}{n}$	$n_{1\cdot}$
A_2	$\widetilde{n}_{21} = \frac{n_{2\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{22} = \frac{n_{2\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{2l} = \frac{n_{2\cdot} \cdot n_{\cdot l}}{n}$	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	$\widetilde{n}_{k1} = \frac{n_{k\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{k2} = \frac{n_{k\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{kl} = \frac{n_{k\cdot} \cdot n_{\cdot l}}{n}$	$n_{k\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot l}$	n

(hier für 2. Variante) oder (falls genügend Raum vorhanden) direkt in der Kontingenztabelle berechnet werden.

Zusammenfassung: Chi-Quadrat-Unabhängigkeitstest

Anwendungs-voraussetzungen	approximativ: (Y^A, Y^B) beliebig verteilt $(X_1^A, X_1^B), \dots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B) Ausprägungen $\{a_1, \dots, a_k\}$ von Y^A, $\{b_1, \dots, b_l\}$ von Y^B *oder* Klassengrenzen $a_1 < \dots < a_{k-1}$ zu Y^A, $b_1 < \dots < b_{l-1}$ zu Y^B
Nullhypothese Gegenhypothese	$H_0 : Y^A, Y^B$ stochastisch unabhängig $H_1 : Y^A, Y^B$ nicht stochastisch unabhängig
Teststatistik	$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} = \left(\sum_{i=1}^{k} \sum_{j=1}^{l} \frac{n_{ij}^2}{\widetilde{n}_{ij}} \right) - n$
Verteilung (H_0)	χ^2 ist näherungsweise $\chi^2((k-1)\cdot(l-1))$-verteilt, falls H_0 gilt (Näherung nur vernünftig, falls $\widetilde{n}_{ij} \geq 5$ für alle i, j)
Benötigte Größen	$n_{ij} = \#\{m \in \{1, \dots, n\} \mid (x_m, y_m) \in A_i \times B_j\}$ für alle i, j mit $A_i = \{a_i\}$, $B_j = \{b_j\}$ bzw. Klassen A_i, B_j nach vorg. Grenzen, $\widetilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ mit $n_{i\cdot} = \sum_{j=1}^{l} n_{ij}$, $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$,
Kritischer Bereich zum Niveau α	$(\chi^2_{(k-1)\cdot(l-1);1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2((k-1)\cdot(l-1))}(\chi^2)$

Beispiel: Zusammenhang Geschlecht/tägl. Fahrzeit (PKW)

- Untersuchungsgegenstand: Sind die beiden Zufallsvariablen „Geschlecht" (Y^A) und „täglich mit PKW zurückgelegte Strecke" (Y^B) stochastisch unabhängig?
- Stichprobeninformation: (Kontingenz-)Tabelle mit gemeinsamen (in der Stichprobe vom Umfang $n = 2000$ beobachteten) Häufigkeiten, wobei für Y^B eine Klassierung in die Klassen „kurz", „mittel" und „lang" durchgeführt wurde:

	Fahrzeit (Y^B)		
Geschlecht (Y^A)	kurz	mittel	lang
Männlich	524	455	221
Weiblich	413	263	124

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: **Chi-Quadrat-Unabhängigkeitstest**

1. **Hypothesen:**

   $H_0 : Y^A, Y^B$ stochastisch unabhängig gegen $H_1 : Y^A, Y^B$ stoch. abhängig
2. **Teststatistik:**

   $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$ ist unter H_0 approximativ

   $\chi^2((k-1)\cdot(l-1))$-verteilt, falls $\widetilde{n}_{ij} \geq 5$ für alle $1 \leq i \leq k$ und $1 \leq j \leq l$.

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**

   $K = (\chi^2_{(k-1)\cdot(l-1);1-\alpha}, +\infty) = (\chi^2_{2;0.95}, +\infty) = (5.991, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

   Um Randhäufigkeiten $n_{i\cdot}$ und $n_{\cdot j}$ ergänzte Tabelle der gemeinsamen Häufigkeiten:

$Y^A \setminus Y^B$	kurz	mittel	lang	$n_{i\cdot}$
Männlich	524	455	221	1200
Weiblich	413	263	124	800
$n_{\cdot j}$	937	718	345	2000

Tabelle der $\widetilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$:

$Y^A \setminus Y^B$	kurz	mittel	lang	$n_{i\cdot}$
Männlich	562.2	430.8	207.0	1200
Weiblich	374.8	287.2	138.0	800
$n_{\cdot j}$	937	718	345	2000

Es gilt $\widetilde{n}_{ij} \geq 5$ für alle $1 \leq i \leq 2$ und $1 \leq j \leq 3$ $\rightsquigarrow$ Näherung ok.

4. *(Fortsetzung: Berechnung der realisierten Teststatistik)*

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} \\
&= \frac{(524 - 562.2)^2}{562.2} + \frac{(455 - 430.8)^2}{430.8} + \frac{(221 - 207)^2}{207} \\
&\quad + \frac{(413 - 374.8)^2}{374.8} + \frac{(263 - 287.2)^2}{287.2} + \frac{(124 - 138)^2}{138} \\
&= 2.5956 + 1.3594 + 0.9469 \\
&\quad + 3.8934 + 2.0391 + 1.4203 \\
&= 12.2547
\end{aligned}$$

5. **Entscheidung:**

   $\chi^2 = 12.2547 \in (5.991, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

   (p-Wert: $1 - F_{\chi^2(2)}(\chi^2) = 1 - F_{\chi^2(2)}(12.2547) = 1 - 0.9978 = 0.0022$)

Der Test kommt also zum Ergebnis, dass die beiden Zufallsvariablen „Geschlecht" und „tägliche Fahrzeit (PKW)" stochastisch **abhängig** sind.

Mittelwertvergleiche

- *Nächste Anwendung:* Vergleich der Mittelwerte zweier *normalverteilter* Zufallsvariablen Y^A und Y^B
  1. auf **derselben** Grundgesamtheit durch Beobachtung von Realisationen $(x_1^A, x_1^B), \ldots, (x_n^A, x_n^B)$ einer (gemeinsamen) einfachen Stichprobe $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ zur **zweidimensionalen** Zufallsvariablen (Y^A, Y^B), insbesondere von Realisationen von Y^A und Y^B für **dieselben** Elemente der Grundgesamtheit („verbundene Stichprobe"),
  2. auf **derselben oder unterschiedlichen** Grundgesamtheit(en) durch Beobachtung von Realisationen $x_1^A, \ldots, x_{n_A}^A$ und $x_1^B, \ldots, x_{n_B}^B$ zu zwei **unabhängigen** einfachen Stichproben $X_1^A, \ldots, X_{n_A}^A$ und $X_1^B, \ldots, X_{n_B}^B$ (möglicherweise mit $n_A \neq n_B$) zu den beiden Zufallsvariablen Y^A und Y^B.
- Anwendungsbeispiele für beide Fragestellungen:
  1. Vergleich der Montagezeiten zweier unterschiedlicher Montageverfahren auf Grundlage von Zeitmessungen beider Verfahren *für dieselbe (Stichproben-)Auswahl von Arbeitern*.
  2. Vergleich der in Eignungstests erreichten Punktzahlen von männlichen und weiblichen Bewerbern (auf Basis zweier unabhängiger einfacher Stichproben).

t-Differenzentest bei verbundener Stichprobe

- Idee für Mittelwertvergleich bei verbundenen Stichproben:
  - Ein Vergleich der Mittelwerte von Y^A und Y^B kann anhand des Mittelwerts $\mu := \mathrm{E}(Y)$ der Differenz $Y := Y^A - Y^B$ erfolgen, denn mit $\mu_A := \mathrm{E}(Y^A)$ und $\mu_B := \mathrm{E}(Y^B)$ gilt offensichtlich $\mu = \mu_A - \mu_B$ und damit:
    $$\mu < 0 \iff \mu_A < \mu_B \qquad \mu = 0 \iff \mu_A = \mu_B \qquad \mu > 0 \iff \mu_A > \mu_B$$
  - Mit $x_1 := x_1^A - x_1^B, \ldots, x_n := x_n^A - x_n^B$ liegt eine Realisation einer einfachen Stichprobe $X_1 := X_1^A - X_1^B, \ldots, X_n := X_n^A - X_n^B$ vom Umfang n zu $Y = Y^A - Y^B$ vor.
  - Darüberhinaus gilt: Ist (Y^A, Y^B) gemeinsam (zweidimensional) normalverteilt, so ist auch die Differenz $Y = Y^A - Y^B$ normalverteilt.
- Es liegt also nahe, die gemeinsame Stichprobe zu (Y^A, Y^B) zu „einer" Stichprobe zu $Y = Y^A - Y^B$ zusammenzufassen und den bekannten t-Test für den Mittelwert einer (normalverteilten) Zufallsvariablen bei unbekannter Varianz auf der Grundlage der einfachen Stichprobe $X_1, \ldots, X_n$ zu Y durchzuführen.
- Prinzipiell wäre bei bekannter Varianz von $Y = Y^A - Y^B$ auch ein entsprechender Gauß-Test durchführbar; Anwendungen hierfür sind aber selten.

Zusammenfassung: t-Differenzentest

Anwendungs-voraussetzungen	exakt: (Y^A, Y^B) gemeinsam (zweidimensional) normalverteilt, $\mathrm{E}(Y^A) = \mu_A, \mathrm{E}(Y^B) = \mu_B$ sowie Varianzen/Kovarianz unbekannt approx.: $\mathrm{E}(Y^A) = \mu_A, \mathrm{E}(Y^B) = \mu_B, \mathrm{Var}(Y^A), \mathrm{Var}(Y^B)$ unbek. $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B)		
Nullhypothese Gegenhypothese	$H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$	$H_0: \mu_A \leq \mu_B$ $H_1: \mu_A > \mu_B$	$H_0: \mu_A \geq \mu_B$ $H_1: \mu_A < \mu_B$
Teststatistik	$t = \frac{\overline{X}}{S}\sqrt{n}$		
Verteilung (H_0)	t für $\mu_A = \mu_B$ (näherungsweise) $t(n-1)$-verteilt		
Benötigte Größen	$X_i = X_i^A - X_i^B$ für $i \in \{1, \ldots, n\}$, $\quad \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right)}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n-1;1-\frac{\alpha}{2}}) \cup (t_{n-1;1-\frac{\alpha}{2}}, \infty)$	$(t_{n-1;1-\alpha}, \infty)$	$(-\infty, -t_{n-1;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n-1)}(\lvert t \rvert))$	$1 - F_{t(n-1)}(t)$	$F_{t(n-1)}(t)$

Beispiel: Montagezeiten von zwei Verfahren

- Untersuchungsgegenstand: Ist ein neu vorgeschlagenes Montageverfahren besser (im Sinne einer im Mittel kürzeren Bearbeitungsdauer Y^B) als das zur Zeit eingesetzte Montageverfahren (mit Bearbeitungsdauer Y^A)?
- Stichprobeninformation: Zeitmessungen der Montagedauern x_i^A für Verfahren A und x_i^B für Verfahren B bei **denselben** $n = 7$ Arbeitern:

Arbeiter i	1	2	3	4	5	6	7
x_i^A	64	71	68	66	73	62	70
x_i^B	60	66	66	69	63	57	62

- Annahme: (Y^A, Y^B) gemeinsam normalverteilt, $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B).
- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: Exakter t-**Differenzentest** für verbundene Stichproben

1. **Hypothesen:**
   $$H_0: \mu_A \leq \mu_B \qquad \text{gegen} \qquad H_1: \mu_A > \mu_B$$
2. **Teststatistik:**
   $$t = \frac{\overline{X}}{S}\sqrt{n} \text{ ist unter } H_0 \ t(n-1)\text{-verteilt (für } \mu_A = \mu_B).$$

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (t_{n-1;1-\alpha}, +\infty) = (t_{6;0.95}, +\infty) = (1.943, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

Arbeiter i	1	2	3	4	5	6	7
x_i^A	64	71	68	66	73	62	70
x_i^B	60	66	66	69	63	57	62
$x_i = x_i^A - x_i^B$	4	5	2	-3	10	5	8

   Mit $\overline{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 4.4286$ und $s = \sqrt{\frac{1}{7-1}\sum_{i=1}^{7}(x_i - \overline{x})^2} = 4.1975$:

   $$t = \frac{\overline{x}}{s}\sqrt{n} = \frac{4.4286}{4.1975}\sqrt{7} = 2.7914$$
5. **Entscheidung:**
   $t = 2.7914 \in (1.943, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
   (p-Wert: $1 - F_{t(6)}(t) = 1 - F_{t(6)}(2.7914) = 1 - 0.9842 = 0.0158$)

Der Test kommt also zur Entscheidung, dass das neue Montageverfahren eine im Mittel signifikant kürzere Montagedauer aufweist.

Mittelwertvergleiche bei zwei unabhängigen Stichproben

- Liegen zwei unabhängige Stichproben $X_1^A, \dots, X_{n_A}^A$ und $X_1^B, \dots, X_{n_B}^B$ zu jeweils normalverteilten Zufallsvariablen Y^A und Y^B vor, kann eine „Aggregation" zu einer einzigen Stichprobe wie beim Vorliegen verbundener Stichproben so nicht durchgeführt werden.
- Verglichen werden nun nicht mehr Beobachtungspaare, sondern die (getrennt) berechneten Mittelwerte $\overline{X^A}$ und $\overline{X^B}$ der beiden Stichprobenrealisationen zu Y^A bzw. Y^B.
- Wir setzen zunächst die *Normalverteilungsannahme für Y^A und Y^B* voraus!
- Die Differenz $\overline{X^A} - \overline{X^B}$ ist wegen der Unabhängigkeit der Stichproben dann offensichtlich normalverteilt mit Erwartungswert $\mu_A - \mu_B$ (für $\mu_A = \mu_B$ gilt also gerade $\mathrm{E}(\overline{X^A} - \overline{X^B}) = 0$) und Varianz

$$\mathrm{Var}(\overline{X^A} - \overline{X^B}) = \mathrm{Var}(\overline{X^A}) + \mathrm{Var}(\overline{X^B}) = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\ .$$

- Sind die beteiligten Varianzen bekannt, kann zum Vergleich von μ_A und μ_B somit unmittelbar ein exakter Gauß-Test konstruiert werden.

Zusammenfassung: 2-Stichproben-Gauß-Test

bei bekannten Varianzen

Anwendungs-voraussetzungen	exakt: $Y^A \sim N(\mu_A, \sigma_A^2)$, $Y^B \sim N(\mu_B, \sigma_B^2)$, σ_A^2, σ_B^2 bekannt; $X_1^A, \dots, X_{n_A}^A$ einfache Stichprobe zu Y^A, unabhängig von einfacher Stichprobe $X_1^B, \dots, X_{n_B}^B$ zu Y^B.		
Nullhypothese Gegenhypothese	$H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$	$H_0: \mu_A \leq \mu_B$ $H_1: \mu_A > \mu_B$	$H_0: \mu_A \geq \mu_B$ $H_1: \mu_A < \mu_B$
Teststatistik	$N = \frac{\overline{X^A} - \overline{X^B}}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$		
Verteilung (H_0)	N für $\mu_A = \mu_B$ $N(0,1)$-verteilt		
Benötigte Größen	$\overline{X^A} = \frac{1}{n_A}\sum_{i=1}^{n_A} X_i^A$, $\overline{X^B} = \frac{1}{n_B}\sum_{i=1}^{n_B} X_i^B$		
Kritischer Bereich zum Niveau α	$(-\infty, -N_{1-\frac{\alpha}{2}}) \cup (N_{1-\frac{\alpha}{2}}, \infty)$	$(N_{1-\alpha}, \infty)$	$(-\infty, -N_{1-\alpha})$
p-Wert	$2 \cdot (1 - \Phi(\lvert N \rvert))$	$1 - \Phi(N)$	$\Phi(N)$

- Sind die Varianzen σ_A^2 und σ_B^2 unbekannt, so ist zu unterscheiden, ob man wenigstens $\sigma_A^2 = \sigma_B^2$ annehmen kann oder nicht.
- Im Fall übereinstimmender Varianzen $\sigma_A^2 = \sigma_B^2$ wird diese mit Hilfe eines gewichteten Mittelwerts S^2 der Stichprobenvarianzen

$$S_{Y^A}^2 = \frac{1}{n_A - 1}\sum_{i=1}^{n_A}(X_i^A - \overline{X^A})^2 \qquad \text{und} \qquad S_{Y^B}^2 = \frac{1}{n_B - 1}\sum_{j=1}^{n_B}(X_j^B - \overline{X^B})^2$$

in der Form

$$S^2 = \frac{(n_A - 1)S_{Y^A}^2 + (n_B - 1)S_{Y^B}^2}{n_A + n_B - 2} = \frac{\sum_{i=1}^{n_A}(X_i^A - \overline{X^A})^2 + \sum_{j=1}^{n_B}(X_j^B - \overline{X^B})^2}{n_A + n_B - 2}$$

geschätzt, ein exakter t-Test ist damit konstruierbar.

- Für $n_A = n_B$ erhält man die einfachere Darstellung $S^2 = \frac{S_{Y^A}^2 + S_{Y^B}^2}{2}$.

Zusammenfassung: 2-Stichproben-t-Test

bei unbekannten, aber übereinstimmenden Varianzen

Anwendungsvoraussetzungen	exakt: $Y^A \sim N(\mu_A, \sigma_A^2)$, $Y^B \sim N(\mu_B, \sigma_B^2)$, $\mu_A, \mu_B, \sigma_A^2 = \sigma_B^2$ unbek. approx.: $\mathrm{E}(Y^A) = \mu_A, \mathrm{E}(Y^B) = \mu_B, \mathrm{Var}(Y^A) = \mathrm{Var}(Y^B)$ unbekannt $X_1^A, \ldots, X_{n_A}^A$ einfache Stichprobe zu Y^A, unabhängig von einfacher Stichprobe $X_1^B, \ldots, X_{n_B}^B$ zu Y^B.		
Nullhypothese Gegenhypothese	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$	$H_0 : \mu_A \leq \mu_B$ $H_1 : \mu_A > \mu_B$	$H_0 : \mu_A \geq \mu_B$ $H_1 : \mu_A < \mu_B$
Teststatistik	$t = \dfrac{\overline{X^A} - \overline{X^B}}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} = \dfrac{\overline{X^A} - \overline{X^B}}{S}\sqrt{\dfrac{n_A \cdot n_B}{n_A + n_B}}$		
Verteilung (H_0)	t für $\mu_A = \mu_B$ (näherungsweise) $t(n_A + n_B - 2)$-verteilt		
Benötigte Größen	$\overline{X^A} = \frac{1}{n_A}\sum_{i=1}^{n_A} X_i^A$, $\overline{X^B} = \frac{1}{n_B}\sum_{i=1}^{n_B} X_i^B$, $S = \sqrt{\frac{(n_A-1)S_{Y^A}^2 + (n_B-1)S_{Y^B}^2}{n_A+n_B-2}} = \sqrt{\frac{\sum_{i=1}^{n_A}(X_i^A - \overline{X^A})^2 + \sum_{i=1}^{n_B}(X_i^B - \overline{X^B})^2}{n_A+n_B-2}}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n_A+n_B-2;1-\frac{\alpha}{2}}) \cup (t_{n_A+n_B-2;1-\frac{\alpha}{2}}, \infty)$	$(t_{n_A+n_B-2;1-\alpha}, \infty)$	$(-\infty, -t_{n_A+n_B-2;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n_A+n_B-2)}(\lvert t \rvert))$	$1 - F_{t(n_A+n_B-2)}(t)$	$F_{t(n_A+n_B-2)}(t)$

Beispiel: Absatzwirkung einer Werbeaktion

- Untersuchungsgegenstand: Hat eine spezielle Sonderwerbeaktion positiven Einfluss auf den mittleren Absatz?
- Stichprobeninformation: Messung der prozentualen Absatzänderungen $x_1^A, \ldots, x_{10}^A$ in $n_A = 10$ Supermärkten **ohne** Sonderwerbeaktion und $x_1^B, \ldots, x_5^B$ in $n_B = 5$ Supermärkten **mit** Sonderwerbeaktion.
- Annahme: Für prozentuale Absatzänderungen Y^A ohne bzw. Y^B mit Sonderwerbeaktion gilt $Y^A \sim N(\mu_A, \sigma_A^2)$, $Y^B \sim N(\mu_B, \sigma_B^2)$, $\mu_A, \mu_B, \sigma_A^2 = \sigma_B^2$ unbekannt, $X_1^A, \ldots, X_{10}^A$ einfache Stichprobe zu Y^A, unabhängig von einfacher Stichprobe $X_1^B, \ldots, X_5^B$ zu Y^B.
- (Zwischen-)Ergebnisse aus Stichprobenrealisation:

$$\overline{x^A} = 6.5, \quad \overline{x^B} = 8, \quad s_{Y^A}^2 = 20.25, \quad s_{Y^B}^2 = 23.04$$

$$\Rightarrow s = \sqrt{\frac{(n_A - 1)s_{Y^A}^2 + (n_B - 1)s_{Y^B}^2}{n_A + n_B - 2}} = \sqrt{\frac{9 \cdot 20.25 + 4 \cdot 23.04}{13}} = 4.5944$$

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test:

2-Stichproben-t-Test bei übereinstimmenden, aber unbekannten Varianzen

1. **Hypothesen:**
   $H_0 : \mu_A \geq \mu_B \quad$ gegen $\quad H_1 : \mu_A < \mu_B$
2. **Teststatistik:**
   $$t = \frac{\overline{X^A} - \overline{X^B}}{S}\sqrt{\frac{n_A \cdot n_B}{n_A + n_B}}$$ ist unter H_0 $t(n_A + n_B - 2)$-verteilt (für $\mu_A = \mu_B$).
3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (-\infty, -t_{n_A+n_B-2;1-\alpha}) = (-\infty, -t_{13;0.95}) = (-\infty, -1.771)$
4. **Berechnung der realisierten Teststatistik:**
   $$t = \frac{\overline{x^A} - \overline{x^B}}{s}\sqrt{\frac{n_A \cdot n_B}{n_A + n_B}} = \frac{6.5 - 8}{4.5944}\sqrt{\frac{10 \cdot 5}{10 + 5}} = -0.5961$$
5. **Entscheidung:**
   $t = -0.5961 \notin (-\infty, -1.771) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
   (p-Wert: $F_{t(13)}(t) = F_{t(13)}(-0.5961) = 0.2807$)

Der Test kommt also zur Entscheidung, dass eine positive Auswirkung der Sonderwerbeaktion auf die mittlere prozentuale Absatzänderung nicht bestätigt werden kann.

Sonderfall: Vergleich von Anteilswerten

- Ein Sonderfall des (approximativen) 2-Stichproben-t-Test bei unbekannten, aber übereinstimmenden Varianzen liegt vor, wenn zwei Anteilswerte miteinander verglichen werden sollen.
- Es gelte also speziell $Y^A \sim B(1, p_A)$ und $Y^B \sim B(1, p_B)$ für $p_A \in (0, 1)$ und $p_B \in (0, 1)$, außerdem seien $X_1^A, \ldots, X_{n_A}^A$ sowie $X_1^B, \ldots, X_{n_B}^B$ unabhängige einfache Stichproben vom Umfang n_A zu Y^A bzw. vom Umfang n_B zu Y^B.
- Zur Überprüfung stehen die Hypothesenpaare:

	$H_0 : p_A = p_B$	$H_0 : p_A \leq p_B$	$H_0 : p_A \geq p_B$
gegen	$H_1 : p_A \neq p_B$	$H_1 : p_A > p_B$	$H_1 : p_A < p_B$

- Für die Varianzen von Y^A und Y^B gilt bekanntlich $\mathrm{Var}(Y^A) = p_A \cdot (1 - p_A)$ bzw. $\mathrm{Var}(Y^B) = p_B \cdot (1 - p_B)$, d.h. die Varianzen sind zwar unbekannt, unter H_0 — genauer für $p_A = p_B$ — jedoch gleich.
- Mit den üblichen Schreibweisen $\widehat{p}_A := \frac{1}{n_A}\sum_{i=1}^{n_A} X_i^A$ bzw. $\widehat{p}_B := \frac{1}{n_B}\sum_{i=1}^{n_B} X_i^B$ erhält man für S^2 in Abhängigkeit von $\widehat{p}_A$ und $\widehat{p}_B$ die Darstellung:

$$S^2 = \frac{n_A \cdot \widehat{p}_A \cdot (1 - \widehat{p}_A) + n_B \cdot \widehat{p}_B \cdot (1 - \widehat{p}_B)}{n_A + n_B - 2}$$

- Approximation vernünftig, falls $5 \leq n_A\widehat{p}_A \leq n_A - 5$ und $5 \leq n_B\widehat{p}_B \leq n_B - 5$.

Zusammenfassung: 2-Stichproben-t-Test für Anteilswerte

Anwendungs-voraussetzungen	approx.: $Y^A \sim B(1, p_A)$, $Y^B \sim B(1, p_B)$, p_A, p_B unbekannt $X_1^A, \ldots, X_{n_A}^A$ einfache Stichprobe zu Y^A, unabhängig von einfacher Stichprobe $X_1^B, \ldots, X_{n_B}^B$ zu Y^B.		
Nullhypothese Gegenhypothese	$H_0 : p_A = p_B$ $H_1 : p_A \neq p_B$	$H_0 : p_A \leq p_B$ $H_1 : p_A > p_B$	$H_0 : p_A \geq p_B$ $H_1 : p_A < p_B$
Teststatistik	$t = \dfrac{\widehat{p}_A - \widehat{p}_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} = \dfrac{\widehat{p}_A - \widehat{p}_B}{S}\sqrt{\dfrac{n_A \cdot n_B}{n_A + n_B}}$		
Verteilung (H_0)	t für $p_A = p_B$ näherungsweise $t(n_A + n_B - 2)$-verteilt (Näherung ok, falls $5 \leq n_A \widehat{p}_A \leq n_A - 5$ und $5 \leq n_B \widehat{p}_B \leq n_B - 5$)		
Benötigte Größen	$\widehat{p}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} X_i^A$, $\widehat{p}_B = \frac{1}{n_B}\sum_{i=1}^{n_B} X_i^B$, $S = \sqrt{\frac{n_A \cdot \widehat{p}_A \cdot (1-\widehat{p}_A) + n_B \cdot \widehat{p}_B \cdot (1-\widehat{p}_B)}{n_A + n_B - 2}}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n_A+n_B-2;1-\frac{\alpha}{2}}) \cup (t_{n_A+n_B-2;1-\frac{\alpha}{2}}, \infty)$	$(t_{n_A+n_B-2;1-\alpha}, \infty)$	$(-\infty, -t_{n_A+n_B-2;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n_A+n_B-2)}(\lvert t\rvert))$	$1 - F_{t(n_A+n_B-2)}(t)$	$F_{t(n_A+n_B-2)}(t)$

Beispiel: Vergleich von zwei Fehlerquoten

mit approximativem 2-Stichproben-t-Test für Anteilswerte

- Untersuchungsgegenstand: Vergleich von Fehlerquoten zweier Sortiermaschinen
- Für einen automatisierten Sortiervorgang werden eine günstige (A) sowie eine hochpreisige Maschine (B) angeboten. Es soll anhand von 2 (unabhängigen) Testläufen mit jeweils $n_A = n_B = 1000$ Sortiervorgängen überprüft werden, ob die Fehlerquote p_A bei der günstigen Maschine A höher ist als die Fehlerquote p_B der hochpreisigen Maschine B.
- Resultat der Testläufe soll jeweils als Realisation einer einfachen Stichprobe aufgefasst werden können.
- Stichprobeninformation: Bei Maschine A traten 29 Fehler auf, bei Maschine B 21 Fehler.
- (Zwischen-) Ergebnisse aus Stichprobenrealisation: $\widehat{p}_A = \frac{29}{1000} = 0.029$, $\widehat{p}_B = \frac{21}{1000} = 0.021$, $s = \sqrt{\frac{1000 \cdot 0.029 \cdot (1-0.029) + 1000 \cdot 0.021 \cdot (1-0.021)}{1000+1000-2}} = 0.156$
- Gewünschtes Signifikanzniveau $\alpha = 0.05$.

1. **Hypothesen:**
   $H_0 : p_A \leq p_B$ gegen $H_1 : p_A > p_B$
2. **Teststatistik:**
   $t = \dfrac{\widehat{p}_A - \widehat{p}_B}{S}\sqrt{\dfrac{n_A \cdot n_B}{n_A + n_B}}$ ist unter H_0 näherungsweise $t(n_A + n_B - 2)$-verteilt (für $p_A = p_B$). Näherung ok, da $5 \leq 29 \leq 995$ und $5 \leq 21 \leq 995$.
3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (t_{n_A+n_B-2;1-\alpha}, +\infty) = (t_{1998;0.95}, +\infty) = (1.646, +\infty)$
4. **Berechnung der realisierten Teststatistik:**
   $$t = \frac{\widehat{p}_A - \widehat{p}_B}{s}\sqrt{\frac{n_A \cdot n_B}{n_A + n_B}} = \frac{0.029 - 0.021}{0.1562}\sqrt{\frac{1000 \cdot 1000}{1000 + 1000}} = 1.1452$$
5. **Entscheidung:**
   $t = 1.1452 \notin (1.646, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!
   (p-Wert: $1 - F_{t(1998)}(t) = 1 - F_{t(1998)}(1.1452) = 1 - 0.8739 = 0.1261$)

Der Test kommt also zum Ergebnis, dass eine höhere Fehlerquote der günstigen Maschine nicht bestätigt werden kann.

Approximativer 2-Stichproben-Gauß-Test

für Mittelwertvergleiche, wenn Gleichheit der Varianzen ungewiss

- Kann in der Situation des exakten 2-Stichproben-t-Test (Y^A und Y^B sind normalverteilt mit unbekannten Varianzen) auch unter H_0 keine Gleichheit der Varianzen vorausgesetzt werden, müssen andere Testverfahren verwendet werden, z.B. der **Welch-Test** (hier nicht besprochen).
- Als approximativer Test lässt sich (zumindest bei hinreichend großen Stichprobenumfängen, „Daumenregel“ $n_A > 30$ und $n_B > 30$) auch eine leichte Modifikation des 2-Stichproben-Gauß-Tests aus Folie 187 verwenden.
- Anstelle der (dort als bekannt vorausgesetzten) Varianzen σ_A^2 und σ_B^2 sind die erwartungstreuen Schätzfunktionen $S_{Y^A}^2$ und $S_{Y^B}^2$ einzusetzen und der Test als approximativer Test durchzuführen.
- Die Teststatistik nimmt damit die Gestalt
  $$N = \frac{\overline{X^A} - \overline{X^B}}{\sqrt{\frac{S_{Y^A}^2}{n_A} + \frac{S_{Y^B}^2}{n_B}}}$$
  an und ist unter H_0 näherungsweise standardnormalverteilt.

Varianzvergleiche bei normalverteilten Zufallsvariablen

- *Nächste Anwendung:* Vergleich der Varianzen σ_A^2 und σ_B^2 zweier *normalverteilter* Zufallsvariablen $Y^A \sim N(\mu_A, \sigma_A^2)$ und $Y^B \sim N(\mu_B, \sigma_B^2)$ auf Grundlage zweier unabhängiger einfacher Stichproben $X_1^A, \ldots, X_{n_A}^A$ vom Umfang n_A zu Y^A und $X_1^B, \ldots, X_{n_B}^B$ vom Umfang n_B zu Y^B.
- *Idee:* Vergleich auf Grundlage der erwartungstreuen Schätzfunktionen

$$S_{Y^A}^2 = \frac{1}{n_A - 1}\sum_{i=1}^{n_A}(X_i^A - \overline{X^A})^2 = \frac{1}{n_A - 1}\left(\left(\sum_{i=1}^{n_A}(X_i^A)^2\right) - n_A\overline{X^A}^2\right)$$

$$\text{bzw. } S_{Y^B}^2 = \frac{1}{n_B - 1}\sum_{i=1}^{n_B}(X_i^B - \overline{X^B})^2 = \frac{1}{n_B - 1}\left(\left(\sum_{i=1}^{n_B}(X_i^B)^2\right) - n_B\overline{X^B}^2\right)$$

für die Varianz von Y^A bzw. die Varianz von Y^B.
- Es gilt $\frac{(n_A-1)\cdot S_{Y^A}^2}{\sigma_A^2} \sim \chi^2(n_A - 1)$ unabhängig von $\frac{(n_B-1)\cdot S_{Y^B}^2}{\sigma_B^2} \sim \chi^2(n_B - 1)$.
- Geeignete Testgröße lässt sich aus (standardisiertem) Verhältnis von $\frac{(n_A-1)\cdot S_{Y^A}^2}{\sigma_A^2}$ und $\frac{(n_B-1)\cdot S_{Y^B}^2}{\sigma_B^2}$ herleiten.

Die Familie der $F(m, n)$-Verteilungen

- Sind χ_m^2 und χ_n^2 stochastisch unabhängige, mit m bzw. n Freiheitsgraden χ^2-verteilte Zufallsvariablen, so heißt die Verteilung der Zufallsvariablen

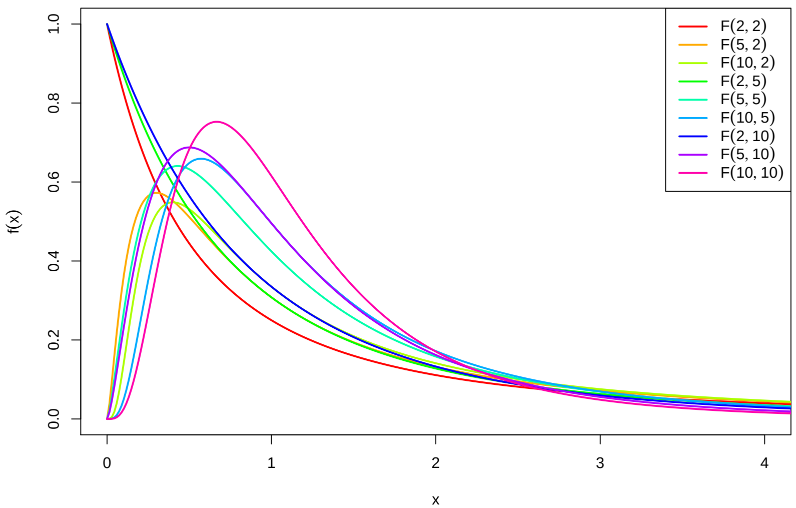
$$F_n^m := \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}} = \frac{\chi_m^2}{\chi_n^2} \cdot \frac{n}{m}$$

F-**Verteilung mit** m **Zähler- und** n **Nennerfreiheitsgraden**, in Zeichen $F_n^m \sim F(m, n)$.
- Offensichtlich können $F(m, n)$-verteilte Zufallsvariablen nur nichtnegative Werte annehmen, der Träger ist also $[0, \infty)$.
- Für $n > 2$ gilt $\mathrm{E}(F_n^m) = \frac{n}{n-2}$.
- Als Abkürzung für α-Quantile der $F(m, n)$-Verteilung verwenden wir (wie üblich) $F_{m,n;\alpha}$.
- Für die Quantile der $F(m, n)$-Verteilungen gilt der folgende Zusammenhang:

$$F_{m,n;\alpha} = \frac{1}{F_{n,m;1-\alpha}}$$

Grafische Darstellung einiger $F(m, n)$-Verteilungen

für $m, n \in \{2, 5, 10\}$



Varianzvergleiche (Fortsetzung)

- Eine $F(n_A - 1, n_B - 1)$-verteilte Zufallsvariable erhält man also in der Anwendungssituation der Varianzvergleiche durch das Verhältnis

$$\frac{\frac{(n_A-1)\cdot S_{Y^A}^2}{\sigma_A^2}}{\frac{(n_B-1)\cdot S_{Y^B}^2}{\sigma_B^2}} \cdot \frac{n_B - 1}{n_A - 1} = \frac{\frac{S_{Y^A}^2}{\sigma_A^2}}{\frac{S_{Y^B}^2}{\sigma_B^2}} ,$$

das allerdings von den (unbekannten!) Varianzen σ_A^2 und σ_B^2 abhängt.
- Gilt jedoch $\sigma_A^2 = \sigma_B^2$, so hat auch das Verhältnis

$$F := \frac{S_{Y^A}^2}{S_{Y^B}^2}$$

eine $F(n_A - 1, n_B - 1)$-Verteilung und ist somit als Testgröße geeignet, wenn unter H_0 (eventuell im Grenzfall) $\sigma_A^2 = \sigma_B^2$ angenommen wird.
- Offensichtlich sprechen große Werte von F eher für $\sigma_A^2 > \sigma_B^2$, kleine eher für $\sigma_A^2 < \sigma_B^2$, Verhältnisse in der Nähe von 1 für $\sigma_A^2 = \sigma_B^2$.

- Da die Klasse der F-Verteilungen von 2 Verteilungsparametern abhängt, ist es nicht mehr möglich, α-Quantile für verschiedene Freiheitsgradkombinationen und verschiedene α darzustellen.
- In Formelsammlung: Tabellen (nur) mit 0.95-Quantilen für verschiedene Kombinationen von m und n für $F(m,n)$-Verteilungen verfügbar.
- Bei linksseitigen Tests (zum Niveau $\alpha = 0.05$) und zweiseitigen Tests (zum Niveau $\alpha = 0.10$) muss also regelmäßig die „Symmetrieeigenschaft"

$$F_{m,n;\alpha} = \frac{1}{F_{n,m;1-\alpha}}$$

verwendet werden, um auch 0.05-Quantile bestimmen zu können.
- Der resultierende Test ist insbesondere zur Überprüfung der Anwendungsvoraussetzungen für den 2-Stichproben-t-Test hilfreich.

Wichtig!

Die Normalverteilungsannahme für Y^A und Y^B ist wesentlich. Ist diese (deutlich) verletzt, ist auch eine näherungsweise Verwendung des Tests nicht mehr angebracht.

0.95-Quantile der $F(m,n)$-Verteilungen $F_{m,n;0.95}$

$n\backslash m$	1	2	3	4	5	6	7	8
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032
150	3.904	3.056	2.665	2.432	2.274	2.160	2.071	2.001

Zusammenfassung: F-Test zum Vergleich der Varianzen

zweier normalverteilter Zufallsvariablen

Anwendungsvoraussetzungen	exakt: $Y^A \sim N(\mu_A, \sigma_A^2)$, $Y^B \sim N(\mu_B, \sigma_B^2)$, $\mu_A, \mu_B, \sigma_A^2, \sigma_B^2$ unbek. $X_1^A, \dots, X_{n_A}^A$ einfache Stichprobe zu Y^A, unabhängig von einfacher Stichprobe $X_1^B, \dots, X_{n_B}^B$ zu Y^B.		
Nullhypothese Gegenhypothese	$H_0: \sigma_A^2 = \sigma_B^2$ $H_1: \sigma_A^2 \neq \sigma_B^2$	$H_0: \sigma_A^2 \leq \sigma_B^2$ $H_1: \sigma_A^2 > \sigma_B^2$	$H_0: \sigma_A^2 \geq \sigma_B^2$ $H_1: \sigma_A^2 < \sigma_B^2$
Teststatistik	$F = \frac{S_{Y^A}^2}{S_{Y^B}^2}$		
Verteilung (H_0)	F unter H_0 für $\sigma_A^2 = \sigma_B^2$ $F(n_A - 1, n_B - 1)$-verteilt		
Benötigte Größen	$\overline{X^A} = \frac{1}{n_A}\sum_{i=1}^{n_A} X_i^A$, $\overline{X^B} = \frac{1}{n_B}\sum_{i=1}^{n_B} X_i^B$, $S_{Y^A}^2 = \frac{1}{n_A-1}\sum_{i=1}^{n_A}(X_i^A - \overline{X^A})^2 = \frac{1}{n_A-1}\left(\left(\sum_{i=1}^{n_A}(X_i^A)^2\right) - n_A\overline{X^A}^2\right)$ $S_{Y^B}^2 = \frac{1}{n_B-1}\sum_{i=1}^{n_B}(X_i^B - \overline{X^B})^2 = \frac{1}{n_B-1}\left(\left(\sum_{i=1}^{n_B}(X_i^B)^2\right) - n_B\overline{X^B}^2\right)$		
Kritischer Bereich zum Niveau α	$[0, F_{n_A-1,n_B-1;\frac{\alpha}{2}})$ $\cup (F_{n_A-1,n_B-1;1-\frac{\alpha}{2}}, \infty)$	$(F_{n_A-1,n_B-1;1-\alpha}, \infty)$	$[0, F_{n_A-1,n_B-1;\alpha})$
p-Wert	$2 \cdot \min\left\{F_{F(n_A-1,n_B-1)}(F), 1 - F_{F(n_A-1,n_B-1)}(F)\right\}$	$1 - F_{F(n_A-1,n_B-1)}(F)$	$F_{F(n_A-1,n_B-1)}(F)$

Beispiel: Präzision von 2 Abfüllanlagen

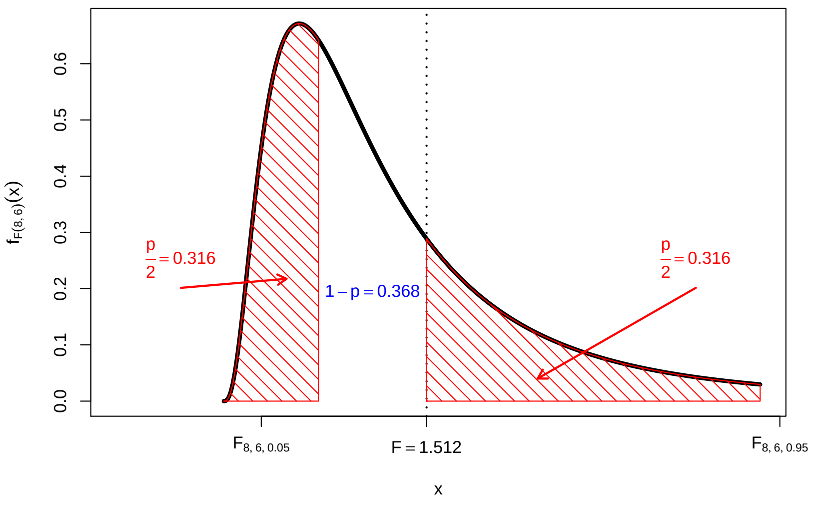
- Untersuchungsgegenstand: Entscheidung, ob Varianz der Abfüllmenge von zwei Abfüllanlagen übereinstimmt oder nicht.
- Annahmen: Abfüllmengen Y^A und Y^B jeweils normalverteilt.
- Unabhängige einfache Stichproben vom Umfang $n_A = 9$ zu Y^A und vom Umfang $n_B = 7$ zu Y^B liefern realisierte Varianzschätzungen $s_{Y^A}^2 = 16.22$ sowie $s_{Y^B}^2 = 10.724$.
- Gewünschtes Signifikanzniveau $\alpha = 0.10$.

Geeigneter Test: F-**Test für die Varianzen normalverteilter Zufallsvariablen**

1. **Hypothesen:** $H_0: \sigma_A^2 = \sigma_B^2$ gegen $H_1: \sigma_A^2 \neq \sigma_B^2$
2. **Teststatistik:** $F = \frac{S_{Y^A}^2}{S_{Y^B}^2}$ ist unter H_0 $F(n_A - 1, n_B - 1)$-verteilt.
3. **Kritischer Bereich zum Niveau** $\alpha = 0.10$**:** Mit $F_{8,6;0.05} = 1/F_{6,8;0.95} = 1/3.581 = 0.279$:
$K = [0, F_{n_A-1,n_B-1;\frac{\alpha}{2}}) \cup (F_{n_A-1,n_B-1;1-\frac{\alpha}{2}}, +\infty) =$
$[0, F_{8,6;0.05}) \cup (F_{8,6;0.95}, +\infty) = [0, 0.279) \cup (4.147, +\infty)$
4. **Berechnung der realisierten Teststatistik:** $F = \frac{s_{Y^A}^2}{s_{Y^B}^2} = \frac{16.22}{10.724} = 1.512$
5. **Entscheidung:** $F \notin K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!

Beispiel: p-Wert bei F-Test für Varianzen (Grafik)

Abfüllanlagenbeispiel, realisierte Teststatistik $F = 1.512$, p-Wert: 0.632



Mittelwertvergleiche bei $k > 2$ unabhängigen Stichproben

- *Nächste Anwendung:* Vergleich der Mittelwerte von $k > 2$ normalverteilten Zufallsvariablen $Y_1 \sim N(\mu_1, \sigma^2), \ldots, Y_k \sim N(\mu_k, \sigma^2)$ mit *übereinstimmender* Varianz σ^2.
- Es soll eine Entscheidung getroffen werden zwischen

  $$H_0 : \mu_1 = \mu_j \text{ für alle } j \qquad \text{und} \qquad H_1 : \mu_1 \neq \mu_j \text{ für (mindestens) ein } j$$

  auf Basis von k unabhängigen einfachen Stichproben

  $$X_{1,1}, \ldots, X_{1,n_1}, \qquad \ldots, \qquad X_{k,1}, \ldots, X_{k,n_k}$$

  mit Stichprobenumfängen $n_1, \ldots, n_k$ (Gesamtumfang: $n := \sum_{j=1}^{k} n_j$).
- Häufiger Anwendungsfall: Untersuchung des Einflusses *einer* nominalskalierten Variablen (mit mehr als 2 Ausprägungen) auf eine (kardinalskalierte) Zufallsvariable, z.B.
  - Einfluss verschiedener Düngemittel auf Ernteertrag,
  - Einfluss verschiedener Behandlungsmethoden auf Behandlungserfolg,
  - Einfluss der Zugehörigkeit zu bestimmten Gruppen (z.B. Schulklassen).
- Beteiligte nominalskalierte Einflussvariable wird dann meist **Faktor** genannt, die einzelnen Ausprägungen **Faktorstufen**.
- Geeignetes statistisches Untersuchungswerkzeug: **Einfache Varianzanalyse**

Einfache Varianzanalyse

- Idee der einfachen („einfaktoriellen") Varianzanalyse:
  Vergleich der Streuung der **Stufenmittel** (auch „Gruppenmittel")

  $$\overline{X}_1 := \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1,i}, \qquad \ldots, \qquad \overline{X}_k := \frac{1}{n_k}\sum_{i=1}^{n_k} X_{k,i}$$

  um das Gesamtmittel

  $$\overline{X} := \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{j,i} = \frac{1}{n}\sum_{j=1}^{k} n_j \cdot \overline{X}_j$$

  mit den Streuungen der Beobachtungswerte $X_{j,i}$ um die jeweiligen Stufenmittel $\overline{X}_j$ innerhalb der j-ten Stufe.
- Sind die Erwartungswerte in allen Stufen gleich (gilt also H_0), so ist die Streuung der Stufenmittel vom Gesamtmittel im Vergleich zur Streuung der Beobachtungswerte um die jeweiligen Stufenmittel *tendenziell* nicht so groß wie es bei Abweichungen der Erwartungswerte für die einzelnen Faktorstufen der Fall wäre.

- Messung der Streuung der Stufenmittel vom Gesamtmittel durch Größe SB (**„Squares Between"**) als (gew.) Summe der quadrierten Abweichungen:

  $$SB = \sum_{j=1}^{k} n_j \cdot (\overline{X}_j - \overline{X})^2 = n_1 \cdot (\overline{X}_1 - \overline{X})^2 + \ldots + n_k \cdot (\overline{X}_k - \overline{X})^2$$

- Messung der (Summe der) Streuung(en) der Beobachtungswerte um die Stufenmittel durch Größe SW (**„Squares Within"**) als (Summe der) Summe der quadrierten Abweichungen:

  $$SW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{j,i} - \overline{X}_j)^2 = \sum_{i=1}^{n_1}(X_{1,i} - \overline{X}_1)^2 + \ldots + \sum_{i=1}^{n_k}(X_{k,i} - \overline{X}_k)^2$$

- Man kann zeigen:
  - Für die Gesamtsumme SS (**„Sum of Squares"**) der quadrierten Abweichungen der Beobachtungswerte vom Gesamtmittelwert mit

    $$SS = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{j,i} - \overline{X})^2 = \sum_{i=1}^{n_1}(X_{1,i} - \overline{X})^2 + \ldots + \sum_{i=1}^{n_k}(X_{k,i} - \overline{X})^2$$

    gilt die **Streuungszerlegung** $SS = SB + SW$.
  - Mit den getroffenen Annahmen sind $\frac{SB}{\sigma^2}$ bzw. $\frac{SW}{\sigma^2}$ unter H_0 unabhängig $\chi^2(k-1)$- bzw. $\chi^2(n-k)$-verteilt $\rightsquigarrow$ Konstruktion geeigneter Teststatistik.

- Da $\frac{SB}{\sigma^2}$ bzw. $\frac{SW}{\sigma^2}$ unter H_0 unabhängig $\chi^2(k-1)$- bzw. $\chi^2(n-k)$-verteilt sind, ist der Quotient
$$F := \frac{\frac{SB}{\sigma^2}}{\frac{SW}{\sigma^2}} \cdot \frac{n-k}{k-1} = \frac{SB}{SW} \cdot \frac{n-k}{k-1} = \frac{\frac{SB}{k-1}}{\frac{SW}{n-k}} = \frac{SB/(k-1)}{SW/(n-k)}$$
unter H_0 also $F(k-1, n-k)$-verteilt.
- Zur Konstruktion des kritischen Bereichs ist zu beachten, dass **große** Quotienten F **gegen** die Nullhypothese sprechen, da in diesem Fall die Abweichung der Stufenmittel vom Gesamtmittel SB verhältnismäßig groß ist.
- Als kritischer Bereich zum Signifikanzniveau α ergibt sich $K = (F_{k-1,n-k;1-\alpha}, \infty)$
- Die Bezeichnung „Varianzanalyse" erklärt sich dadurch, dass (zur Entscheidungsfindung über die Gleichheit der Erwartungswerte!) die Stichprobenvarianzen $SB/(k-1)$ und $SW/(n-k)$ untersucht werden.
- Die Varianzanalyse kann als näherungsweiser Test auch angewendet werden, wenn die Normalverteilungsannahme verletzt ist.
- Das Vorliegen gleicher Varianzen in allen Faktorstufen („Varianzhomogenität") muss jedoch (auch für vernünftige näherungsweise Verwendung) gewährleistet sein! Überprüfung z.B. mit „Levene-Test" oder „Bartlett-Test" (hier nicht besprochen).

Zusammenfassung: Einfache Varianzanalyse

Anwendungs-voraussetzungen	exakt: $Y_j \sim N(\mu_j, \sigma^2)$ für $j \in \{1, \ldots, k\}$ approximativ: Y_j beliebig verteilt mit $E(Y_j) = \mu_j$, $\text{Var}(Y_j) = \sigma^2$ k unabhängige einfache Stichproben $X_{j,1}, \ldots, X_{j,n_j}$ vom Umfang n_j zu Y_j für $j \in \{1, \ldots, k\}$, $n = \sum_{j=1}^k n_j$
Nullhypothese Gegenhypothese	$H_0: \mu_1 = \mu_j$ für alle $j \in \{2, \ldots, k\}$ $H_1: \mu_1 \neq \mu_j$ für (mindestens) ein $j \in \{2, \ldots, k\}$
Teststatistik Verteilung (H_0)	$F = \frac{SB/(k-1)}{SW/(n-k)}$ F ist (approx.) $F(k-1, n-k)$-verteilt, falls $\mu_1 = \ldots = \mu_k$
Benötigte Größen	$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i}$ für $j \in \{1, \ldots, k\}$, $\overline{x} = \frac{1}{n}\sum_{j=1}^{k} n_j \cdot \overline{x}_j$, $SB = \sum_{j=1}^{k} n_j \cdot (\overline{x}_j - \overline{x})^2, SW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{j,i} - \overline{x}_j)^2$
Kritischer Bereich zum Niveau α	$(F_{k-1,n-k;1-\alpha}, \infty)$
p-Wert	$1 - F_{F(k-1,n-k)}(F)$

- Alternative Berechnungsmöglichkeiten mit „Verschiebungssatz"
  - für Realisation von SB:
$$SB = \sum_{j=1}^{k} n_j \cdot (\overline{x}_j - \overline{x})^2 = \left(\sum_{j=1}^{k} n_j \overline{x}_j^2\right) - n\overline{x}^2$$
  - für Realisation von SW:
$$SW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{j,i} - \overline{x}_j)^2 = \sum_{j=1}^{k}\left(\left(\sum_{i=1}^{n_j} x_{j,i}^2\right) - n_j\overline{x}_j^2\right)$$
- Liegen für $j \in \{1, \ldots, k\}$ die Stichprobenvarianzen
$$S_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(X_{j,i} - \overline{X}_j)^2$$
bzw. deren Realisationen s_j^2 für die k (Einzel-)Stichproben
$$X_{1,1}, \ldots, X_{1,n_1}, \qquad \ldots, \qquad X_{k,1}, \ldots, X_{k,n_k}$$
vor, so erhält man die Realisation von SW offensichtlich auch durch
$$SW = \sum_{j=1}^{k}(n_j - 1) \cdot s_j^2 \ .$$

Beispiel: Bedienungszeiten an $k = 3$ Servicepunkten

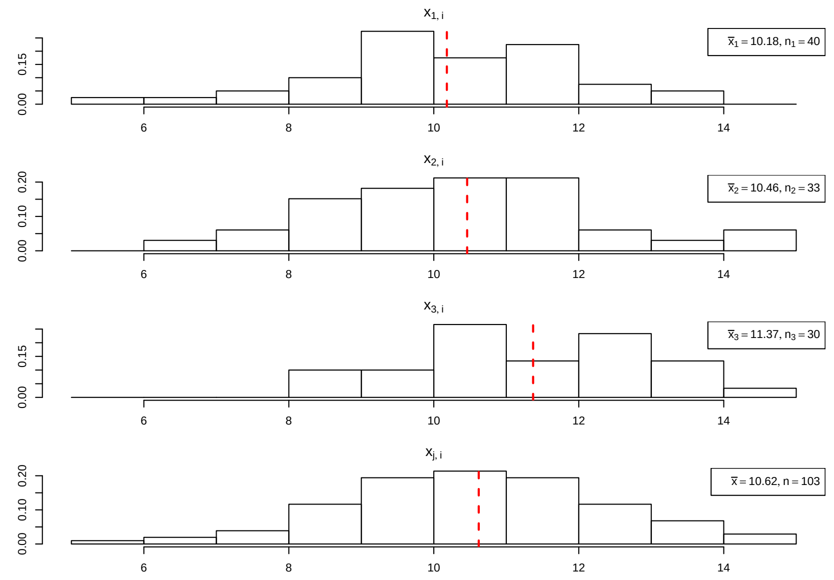
- Untersuchungsgegenstand: Stimmen die mittleren Bedienungszeiten μ_1, μ_2, μ_3 an 3 verschiedenen Servicepunkten überein oder nicht?
- Annahme: Bedienungszeiten Y_1, Y_2, Y_3 an den 3 Servicestationen sind jeweils normalverteilt mit $E(Y_j) = \mu_j$ und **identischer** (unbekannter) Varianz $\text{Var}(Y_j) = \sigma^2$.
- Es liegen Realisationen von 3 unabhängigen einfache Stichproben zu den Zufallsvariablen Y_1, Y_2, Y_3 mit den Stichprobenumfängen $n_1 = 40, n_2 = 33, n_3 = 30$ wie folgt vor:

j (Servicepunkt)	n_j	$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i}$	$\sum_{i=1}^{n_j} x_{j,i}^2$
1	40	10.18	4271.59
2	33	10.46	3730.53
3	30	11.37	3959.03

(Daten simuliert mit $\mu_1 = 10, \mu_2 = 10, \mu_3 = 11.5, \sigma^2 = 2^2$)
- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeignetes Verfahren: **Varianzanalyse**

Grafische Darstellung der Stichprobeninformation



1. **Hypothesen:**
   $H_0 : \mu_1 = \mu_2 = \mu_3 \qquad H_1 : \mu_1 \neq \mu_j$ für mindestens ein j
2. **Teststatistik:**
   $F = \frac{SB/(k-1)}{SW/(n-k)}$ ist unter H_0 $F(k-1, n-k)$-verteilt.
3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (F_{k-1;n-k;1-\alpha}, +\infty) = (F_{2;100;0.95}, +\infty) = (3.087, +\infty)$
4. **Berechnung der realisierten Teststatistik:**
   Mit $\overline{x}_1 = 10.18, \overline{x}_2 = 10.46, \overline{x}_3 = 11.37$ erhält man

$$\overline{x} = \frac{1}{103} \sum_{j=1}^{3} n_j \cdot \overline{x}_j = \frac{1}{103} (40 \cdot 10.18 + 33 \cdot 10.46 + 30 \cdot 11.37) = 10.62$$

und damit

$$\begin{aligned} SB &= \sum_{j=1}^{3} n_j \left(\overline{x}_j - \overline{x}\right)^2 = n_1 \left(\overline{x}_1 - \overline{x}\right)^2 + n_2 \left(\overline{x}_2 - \overline{x}\right)^2 + n_3 \left(\overline{x}_3 - \overline{x}\right)^2 \\ &= 40(10.18 - 10.62)^2 + 33(10.46 - 10.62)^2 + 30(11.37 - 10.62)^2 \\ &= 25.46 \ . \end{aligned}$$

4. (Fortsetzung)
   Außerdem errechnet man

$$\begin{aligned} SW &= \sum_{j=1}^{3} \sum_{i=1}^{n_j} (x_{j,i} - \overline{x}_j)^2 = \sum_{j=1}^{3} \left(\left(\sum_{i=1}^{n_j} x_{j,i}^2 \right) - n_j \cdot \overline{x}_j^2 \right) \\ &= \left(\sum_{i=1}^{n_1} x_{j,i}^2 \right) - n_1 \cdot \overline{x}_1^2 + \left(\sum_{i=1}^{n_2} x_{j,i}^2 \right) - n_2 \cdot \overline{x}_2^2 + \left(\sum_{i=1}^{n_3} x_{j,i}^2 \right) - n_3 \cdot \overline{x}_3^2 \\ &= 4271.59 - 40 \cdot 10.18^2 + 3730.53 - 33 \cdot 10.46^2 + 3959.03 - 30 \cdot 11.37^2 \\ &= 326.96 \ . \end{aligned}$$

Insgesamt erhält man

$$F = \frac{SB/(k-1)}{SW/(n-k)} = \frac{25.46/(3-1)}{326.96/(103-3)} = \frac{12.73}{3.27} = 3.89 \ .$$

5. **Entscheidung:**
   $F = 3.89 \in (3.087, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
   (p-Wert: $1 - F_{F(2,100)}(F) = 1 - F_{F(2,100)}(3.89) = 1 - 0.98 = 0.02$)

ANOVA-Tabelle

- Zusammenfassung der (Zwischen-)Ergebnisse einer Varianzanalyse oft in Form einer sog. ANOVA(ANalysis Of VAriance) - Tabelle wie folgt:

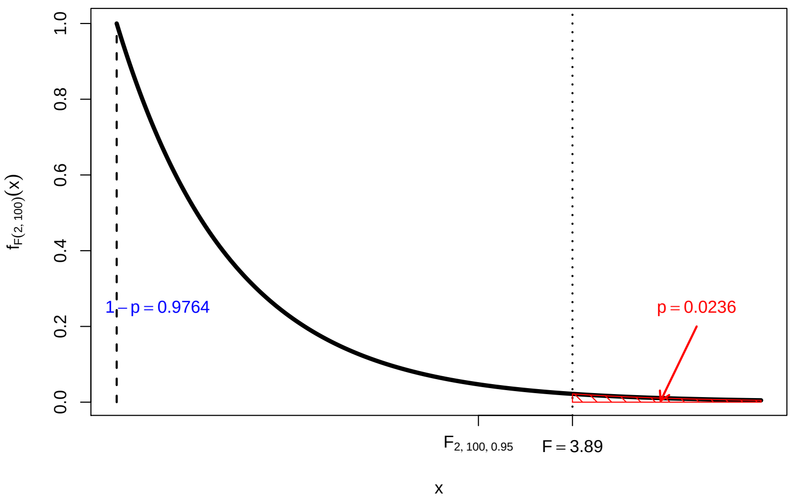
Streuungsursache	Freiheitsgrade	Quadratsumme	Mittleres Quadrat
Faktor	$k-1$	SB	$\frac{SB}{k-1}$
Zufallsfehler	$n-k$	SW	$\frac{SW}{n-k}$
Summe	$n-1$	SS	

- Im Bedienungszeiten-Beispiel erhält man so:

Streuungsursache	Freiheitsgrade	Quadratsumme	Mittleres Quadrat
Faktor	2	25.46	12.73
Zufallsfehler	100	326.96	3.27
Summe	102	352.42	

Beispiel: p-Wert bei Varianzanalyse (Grafik)

Bedienungszeiten-Beispiel, realisierte Teststatistik $F = 3.89$, p-Wert: 0.0236



Varianzanalyse und 2-Stichproben-t-Test

- Varianzanalyse zwar für $k > 2$ unabhängige Stichproben eingeführt, Anwendung aber auch für $k = 2$ möglich.
- Nach Zuordnung der beteiligten Größen in den unterschiedlichen Notationen ($\mu_A \equiv \mu_1$, $\mu_B \equiv \mu_2$, $X_i^A \equiv X_{1,i}$, $X_i^B \equiv X_{2,i}$, $n_A \equiv n_1$, $n_B \equiv n_2$, $n = n_A + n_B$) enger Zusammenhang zum 2-Stichproben-t-Test erkennbar:
  - Fragestellungen (Hypothesenpaare) und Anwendungsvoraussetzungen identisch mit denen des zweiseitigen 2-Stichproben-t-Tests für den Mittelwertvergleich bei unbekannten, aber übereinstimmenden Varianzen.
  - Man kann zeigen: Für Teststatistik F der Varianzanalyse im Fall $k = 2$ und Teststatistik t des 2-Stichproben-t-Tests gilt $F = t^2$.
  - Es gilt außerdem zwischen Quantilen der $F(1, n)$ und der $t(n)$-Verteilung der Zusammenhang $F_{1,n;1-\alpha} = t^2_{n;1-\frac{\alpha}{2}}$. Damit:

$$x \in (-\infty, -t_{n;1-\frac{\alpha}{2}}) \cup (t_{n;1-\frac{\alpha}{2}}, \infty) \iff x^2 \in (F_{1,n;1-\alpha}, \infty)$$

- Insgesamt sind damit die Varianzanalyse mit $k = 2$ Faktorstufen und der zweiseitige 2-Stichproben-t-Test für den Mittelwertvergleich bei unbekannten, aber übereinstimmenden Varianzen also äquivalent in dem Sinn, dass Sie stets übereinstimmende Testentscheidungen liefern!

Deskriptive Beschreibung linearer Zusammenhänge

- Aus deskriptiver Statistik bekannt: Pearsonscher Korrelationskoeffizient als Maß der Stärke des *linearen* Zusammenhangs zwischen zwei (kardinalskalierten) Merkmalen X und Y.
- *Nun:* Ausführlichere Betrachtung linearer Zusammenhänge zwischen Merkmalen (zunächst rein deskriptiv!):
  Liegt ein linearer Zusammenhang zwischen zwei Merkmalen X und Y nahe, ist nicht nur die Stärke dieses Zusammenhangs interessant, sondern auch die genauere „Form" des Zusammenhangs.
- „Form" linearer Zusammenhänge kann durch Geraden(gleichungen) spezifiziert werden.
- *Problemstellung:* Wie kann zu einer Urliste $(x_1, y_1), \ldots, (x_n, y_n)$ der Länge n zu (X, Y) eine sog. **Regressiongerade** (auch: Ausgleichsgerade) gefunden werden, die den linearen Zusammenhang zwischen X und Y „möglichst gut" widerspiegelt?
- *Wichtig:* Was soll „möglichst gut" überhaupt bedeuten?
  Hier: Summe der quadrierten Abstände von der Geraden zu den Datenpunkten (x_i, y_i) in **vertikaler** Richtung soll möglichst gering sein.
  (Begründung für Verwendung dieses „Qualitätskriteriums" wird nachgeliefert!)

- Geraden (eindeutig) bestimmt (zum Beispiel) durch Absolutglied a und Steigung b in der bekannten Darstellung

$$y = f_{a,b}(x) := a + b \cdot x \,.$$

- Für den i-ten Datenpunkt (x_i, y_i) erhält man damit den vertikalen Abstand

$$u_i(a, b) := y_i - f_{a,b}(x_i) = y_i - (a + b \cdot x_i)$$

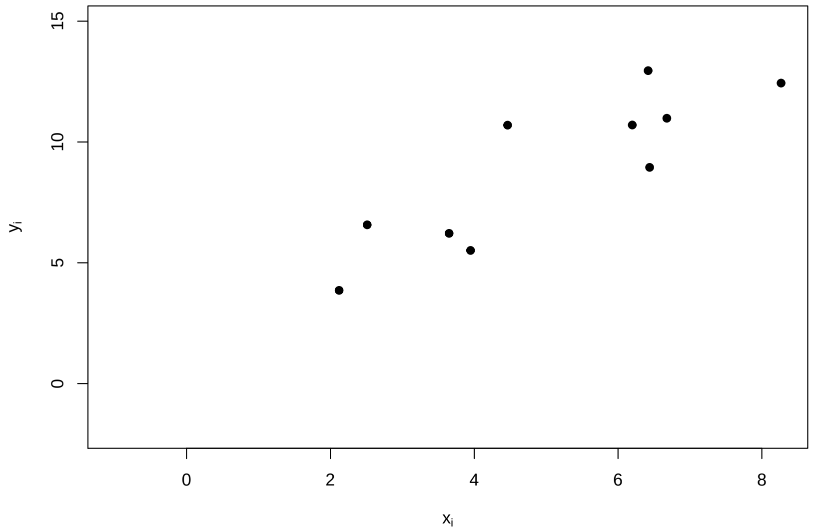
  von der Geraden mit Absolutglied a und Steigung b.
- Gesucht werden a und b so, dass die Summe der quadrierten vertikalen Abstände der „Punktwolke" (x_i, y_i) von der durch a und b festgelegten Geraden,

$$\sum_{i=1}^{n} (u_i(a, b))^2 = \sum_{i=1}^{n} (y_i - f_{a,b}(x_i))^2 = \sum_{i=1}^{n} (y_i - (a + b \cdot x_i))^2 \,,$$

  möglichst klein wird.
- Verwendung dieses Kriteriums heißt auch **Methode der kleinsten Quadrate (KQ-Methode)** oder **Least-Squares-Methode (LS-Methode)**.

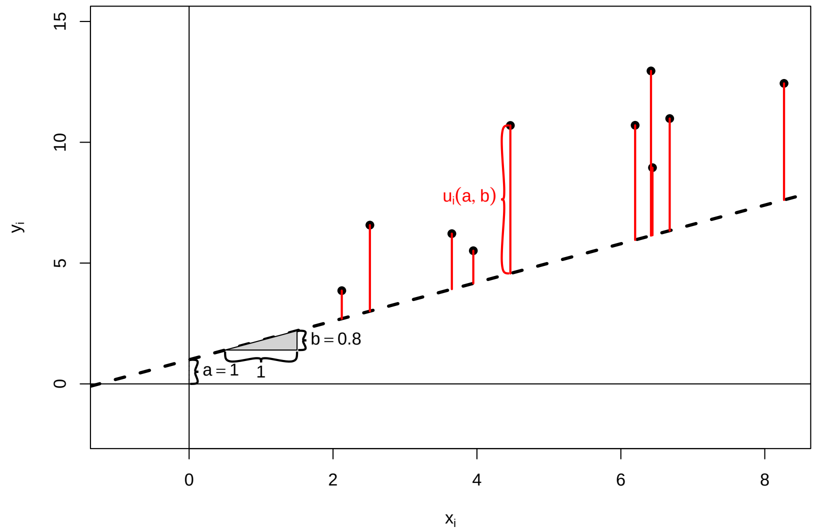
Beispiel: „Punktwolke“

aus $n = 10$ Paaren (x_i, y_i)



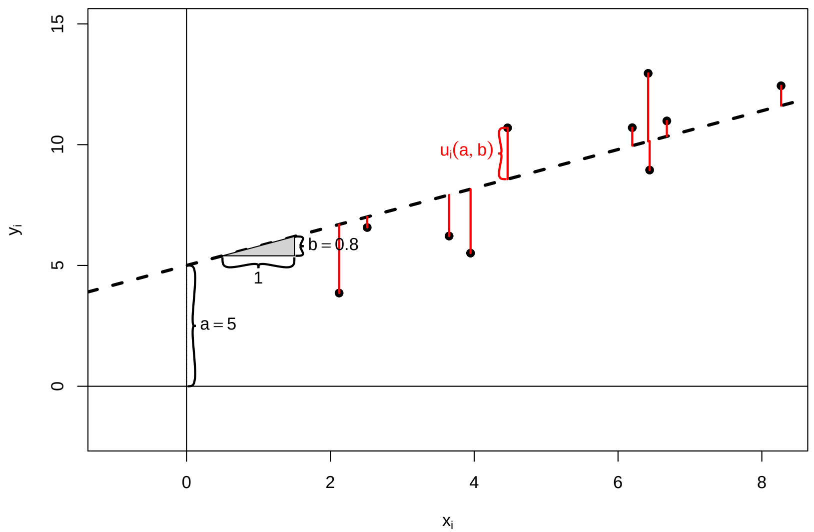
Beispiel: „Punktwolke“ und verschiedene Geraden (I)

$a = 1,\ b = 0.8,\ \sum_{i=1}^{n}(u_i(a,b))^2 = 180.32$



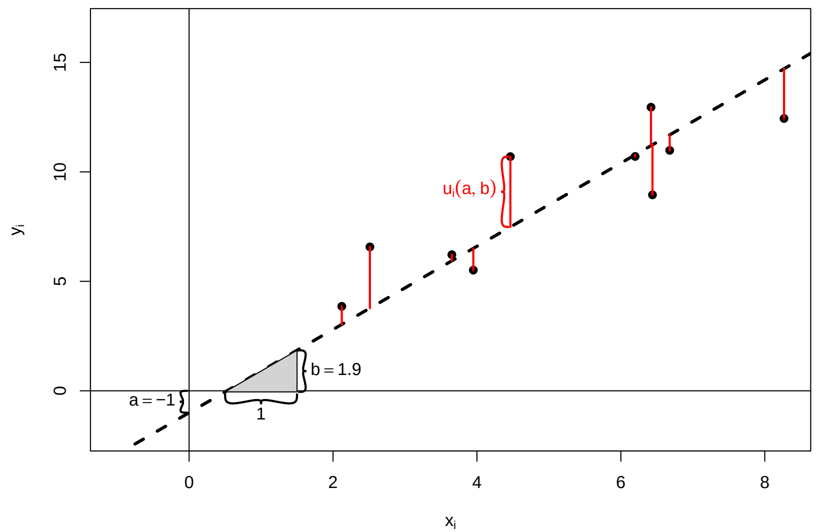
Beispiel: „Punktwolke“ und verschiedene Geraden (II)

$a = 5,\ b = 0.8,\ \sum_{i=1}^{n}(u_i(a,b))^2 = 33.71$



Beispiel: „Punktwolke“ und verschiedene Geraden (III)

$a = -1,\ b = 1.9,\ \sum_{i=1}^{n}(u_i(a,b))^2 = 33.89$



Rechnerische Bestimmung der Regressionsgeraden (I)

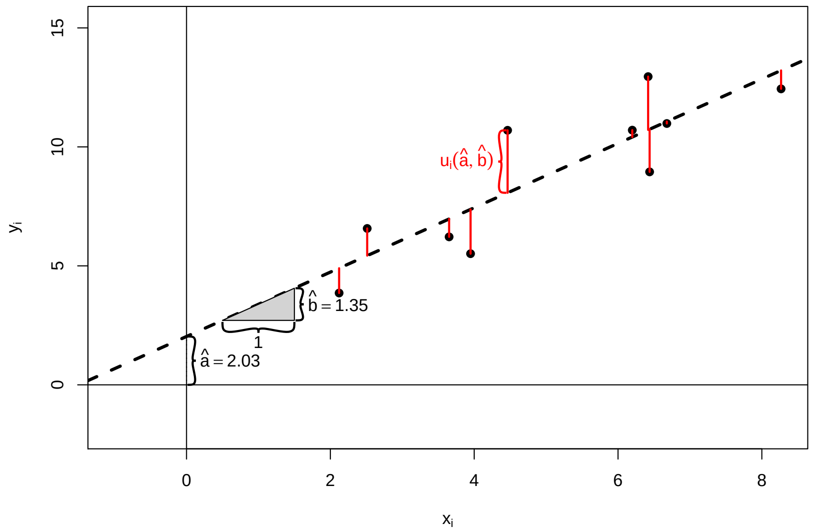
- Gesucht sind also $\widehat{a}, \widehat{b} \in \mathbb{R}$ mit
$$\sum_{i=1}^{n}(y_i - (\widehat{a} + \widehat{b}x_i))^2 = \min_{a,b\in\mathbb{R}} \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$
- Lösung dieses Optimierungsproblems durch Nullsetzen des Gradienten, also
$$\frac{\partial \sum_{i=1}^{n}(y_i - (a + bx_i))^2}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) \overset{!}{=} 0$$
$$\frac{\partial \sum_{i=1}^{n}(y_i - (a + bx_i))^2}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i \overset{!}{=} 0 \;,$$
führt zu sogenannten **Normalgleichungen**:
$$na + \left(\sum_{i=1}^{n} x_i\right) b \overset{!}{=} \sum_{i=1}^{n} y_i$$
$$\left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b \overset{!}{=} \sum_{i=1}^{n} x_i y_i$$

Rechnerische Bestimmung der Regressionsgeraden (II)

- Aufgelöst nach a und b erhält man die Lösungen
$$\widehat{b} = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\cdot\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$
$$\widehat{a} = \tfrac{1}{n}\left(\sum_{i=1}^{n} y_i\right) - \tfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\cdot \widehat{b}$$
oder kürzer mit den aus der deskr. Statistik bekannten Bezeichnungen
$$\overline{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{x^2} = \tfrac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \overline{y} = \tfrac{1}{n}\sum_{i=1}^{n} y_i \quad \text{und} \quad \overline{xy} = \tfrac{1}{n}\sum_{i=1}^{n} x_i y_i$$
bzw. den empirischen Momenten $s_{X,Y} = \overline{xy} - \overline{x}\cdot\overline{y}$ und $s_X^2 = \overline{x^2} - \overline{x}^2$:
$$\widehat{b} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{s_{X,Y}}{s_X^2}$$
$$\widehat{a} = \overline{y} - \overline{x}\widehat{b}$$
- Die erhaltenen Werte $\widehat{a}$ und $\widehat{b}$ minimieren tatsächlich die Summe der quadrierten vertikalen Abstände, da die Hesse-Matrix positiv definit ist.

Beispiel: „Punktwolke“ und Regressionsgerade

$\widehat{a} = 2.03$, $\widehat{b} = 1.35$, $\sum_{i=1}^{n}(u_i(\widehat{a},\widehat{b}))^2 = 22.25$

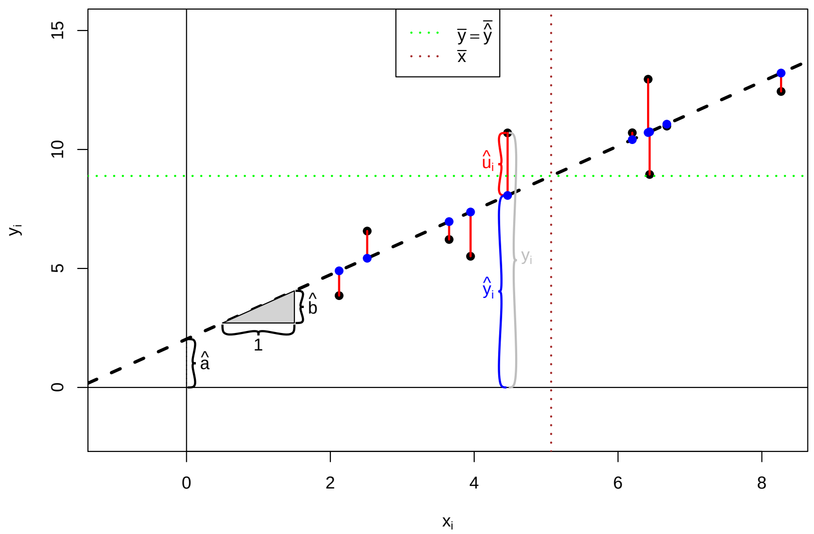


- Zu $\widehat{a}$ und $\widehat{b}$ kann man offensichtlich die folgende, *durch die Regressionsgerade erzeugte* Zerlegung der Merkmalswerte y_i betrachten:
$$y_i = \underbrace{\widehat{a} + \widehat{b}\cdot x_i}_{=:\widehat{y}_i} + \underbrace{y_i - (\widehat{a} + \widehat{b}\cdot x_i)}_{=u_i(\widehat{a},\widehat{b})=:\widehat{u}_i}$$
- Aus den Normalgleichungen lassen sich leicht einige wichtige Eigenschaften für die so definierten $\widehat{u}_i$ und $\widehat{y}_i$ herleiten, insbesondere:
  - $\sum_{i=1}^{n} \widehat{u}_i = 0$ und damit $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \widehat{y}_i$ bzw. $\overline{y} = \overline{\widehat{y}} := \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i$.
  - $\sum_{i=1}^{n} x_i\widehat{u}_i = 0$.
  - Mit $\sum_{i=1}^{n} \widehat{u}_i = 0$ und $\sum_{i=1}^{n} x_i\widehat{u}_i = 0$ folgt auch$\sum_{i=1}^{n} \widehat{y}_i\widehat{u}_i = 0$.

  Mit diesen Eigenschaften erhält man die folgende Varianzzerlegung:
$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\text{Gesamtvarianz der } y_i} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - \overline{\widehat{y}})^2}_{\text{erklärte Varianz}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^2}_{\text{unerklärte Varianz}}$$
- Die als Anteil der erklärten Varianz an der Gesamtvarianz gemessene Stärke des linearen Zusammenhangs steht in engem Zusammenhang mit $r_{X,Y}$; es gilt:
$$r_{X,Y}^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - \overline{\widehat{y}})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Beispiel: Regressionsgerade mit Zerlegung $y_i = \widehat{y}_i + \widehat{u}_i$

$\widehat{a} = 2.03$, $\widehat{b} = 1.35$, $\sum_{i=1}^{n}(\widehat{u}_i)^2 = 22.25$



Beispiel: Berechnung von $\widehat{a}$ und $\widehat{b}$

- Daten im Beispiel:

i	1	2	3	4	5	6	7	8	9	10
x_i	2.51	8.27	4.46	3.95	6.42	6.44	2.12	3.65	6.2	6.68
y_i	6.57	12.44	10.7	5.51	12.95	8.95	3.86	6.22	10.7	10.98

- Berechnete (deskriptive/empirische) Größen:

$$\overline{x} = 5.0703 \qquad \overline{y} = 8.8889 \qquad \overline{x^2} = 29.3729 \qquad \overline{y^2} = 87.9398$$
$$s_X^2 = 3.665 \qquad s_Y^2 = 8.927 \qquad s_{X,Y} = 4.956 \qquad r_{X,Y} = 0.866$$

- Damit erhält man Absolutglied $\widehat{a}$ und Steigung $\widehat{b}$ als

$$\widehat{b} = \frac{s_{X,Y}}{s_X^2} = \frac{4.956}{3.665} = 1.352$$
$$\widehat{a} = \overline{y} - \widehat{b} \cdot \overline{x} = 8.8889 - 1.352 \cdot 5.0703 = 2.03$$

und damit die Regressionsgerade

$$y = f(x) = 2.03 + 1.352 \cdot x \,.$$

- *Bisher: rein deskriptive Betrachtung linearer Zusammenhänge*
- Bereits erläutert/bekannt: Korrelation $\neq$ Kausalität:
  Aus einem beobachteten (linearen) Zusammenhang zwischen zwei Merkmalen lässt sich **nicht** schließen, dass der Wert eines Merkmals den des anderen beeinflusst.
- Bereits durch die Symmetrieeigenschaft $r_{X,Y} = r_{Y,X}$ bei der Berechnung von Pearsonschen Korrelationskoeffizienten wird klar, dass diese Kennzahl alleine auch keine Wirkungsrichtung erkennen lassen **kann**.
- *Nun: statistische Modelle für lineare Zusammenhänge*
- **Keine** symmetrische Behandlung von X und Y mehr, sondern:
  - Interpretation von X („Regressor“) als **erklärende** *deterministische* Variable.
  - Interpretation von Y („Regressand“) als **abhängige, zu erklärende** (Zufalls-)Variable.
- Es wird angenommen, dass Y in linearer Form von X abhängt, diese Abhängigkeit jedoch nicht „perfekt“ ist, sondern durch zufällige Einflüsse „gestört“ wird.
- Anwendung in Experimenten: Festlegung von X durch Versuchsplaner, Untersuchung des Effekts auf Y
- Damit auch Kausalitätsanalysen möglich!

Das einfache lineare Regressionsmodell

- Es wird genauer angenommen, dass für $i \in \{1, \ldots, n\}$ die Beziehung

$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i$$

gilt, wobei
  - $u_1, \ldots, u_n$ (Realisationen von) Zufallsvariablen mit $\mathrm{E}(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma^2$ (unbekannt) und $\mathrm{Cov}(u_i, u_j) = 0$ für $i \neq j$ sind, die zufällige Störungen der linearen Beziehung (**„Störgrößen“**) beschreiben,
  - $x_1, \ldots, x_n$ deterministisch sind mit $s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0$ (d.h. nicht alle x_i sind gleich),
  - β_1, β_2 feste, **unbekannte** reelle Parameter sind.
- Man nimmt an, dass man neben $x_1, \ldots, x_n$ auch $y_1, \ldots, y_n$ beobachtet, die wegen der Abhängigkeit von den Zufallsvariablen $u_1, \ldots, u_n$ ebenfalls (Realisationen von) Zufallsvariablen sind. Dies bedeutet **nicht**, dass man auch (Realisationen von) $u_1, \ldots, u_n$ beobachten kann (β_1 und β_2 unbekannt!).
- Für die Erwartungswerte von y_i gilt

$$E(y_i) = \beta_1 + \beta_2 \cdot x_i \text{ für } i \in \{1, \ldots, n\} \,.$$

- Das durch obige Annahmen beschriebene Modell heißt auch **einfaches lineares Regressionsmodell**.

- Im einfachen linearen Regressionsmodell sind also (neben σ^2) insbesondere β_1 und β_2 Parameter, deren Schätzung für die Quantifizierung des linearen Zusammenhangs zwischen x_i und y_i nötig ist.
- Die Schätzung dieser beiden Parameter führt wieder zum Problem der Suche nach Absolutglied und Steigung einer geeigneten Geradengleichung

$$y = f_{\beta_1,\beta_2}(x) = \beta_1 + \beta_2 \cdot x \ .$$

Satz 10.1 (Satz von Gauß-Markov)

Unter den getroffenen Annahmen liefert die aus dem deskriptiven Ansatz bekannte Verwendung der **KQ-Methode**, *also die Minimierung der Summe der quadrierten vertikalen Abstände zur durch β_1 und β_2 bestimmten Geraden, in Zeichen*

$$\sum_{i=1}^{n}(y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i))^2 \stackrel{!}{=} \min_{\beta_1,\beta_2\in\mathbb{R}} \sum_{i=1}^{n}(y_i - (\beta_1 + \beta_2 \cdot x_i))^2 \ ,$$

die **beste** *(varianzminimale)* **lineare** *(in y_i)* **erwartungstreue** *Schätzfunktion $\widehat{\beta}_1$ für β_1 bzw. $\widehat{\beta}_2$ für β_2.*

- Dies rechtfertigt letztendlich die Verwendung des Optimalitätskriteriums „Minimierung der quadrierten vertikalen Abstände“.

- Man erhält also — ganz analog zum deskriptiven Ansatz — die folgenden Parameterschätzer:

Parameterschätzer im einfachen linearen Regressionsmodell

$$\widehat{\beta}_2 = \frac{n\left(\sum_{i=1}^n x_i y_i\right) - \left(\sum_{i=1}^n x_i\right)\cdot\left(\sum_{i=1}^n y_i\right)}{n\left(\sum_{i=1}^n x_i^2\right) - \left(\sum_{i=1}^n x_i\right)^2} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{s_{X,Y}}{s_X^2} = r_{X,Y}\cdot\frac{s_Y}{s_X} \ ,$$

$$\widehat{\beta}_1 = \frac{1}{n}\left(\sum_{i=1}^n y_i\right) - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)\cdot\widehat{\beta}_2 = \overline{y} - \overline{x}\widehat{\beta}_2 \ .$$

- Wegen der Abhängigkeit von y_i handelt es sich bei $\widehat{\beta}_1$ und $\widehat{\beta}_2$ (wie in der schließenden Statistik gewohnt) um (Realisationen von) *Zufallsvariablen*.
- Die resultierenden vertikalen Abweichungen $\widehat{u}_i := y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i) = y_i - \widehat{y}_i$ der y_i von den auf der Regressionsgeraden liegenden Werten $\widehat{y}_i := \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i$ nennt man **Residuen**.
- Wie im deskriptiven Ansatz gelten die Beziehungen

$$\sum_{i=1}^n \widehat{u}_i = 0, \quad \sum_{i=1}^n y_i = \sum_{i=1}^n \widehat{y}_i, \quad \sum_{i=1}^n x_i\widehat{u}_i = 0, \quad \sum_{i=1}^n \widehat{y}_i\widehat{u}_i = 0$$

sowie die Varianzzerlegung

$$\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 = \frac{1}{n}\sum_{i=1}^n (\widehat{y}_i - \overline{\widehat{y}})^2 + \frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 \ .$$

Das (multiple) Bestimmtheitsmaß R^2

- Auch im linearen Regressionsmodell wird die Stärke des linearen Zusammenhangs mit dem Anteil der erklärten Varianz an der Gesamtvarianz gemessen und mit

$$R^2 = \frac{\sum_{i=1}^n (\widehat{y}_i - \overline{\widehat{y}})^2}{\sum_{i=1}^n (y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^n \widehat{u}_i^2}{\sum_{i=1}^n (y_i - \overline{y})^2}$$

bezeichnet. R^2 wird auch **(multiples) Bestimmtheitsmaß** genannt.
- Es gilt $0 \leq R^2 \leq 1$ sowie der (bekannte) Zusammenhang $R^2 = r_{X,Y}^2 = \frac{s_{X,Y}^2}{s_X^2 \cdot s_Y^2}$.
- Größere Werte von R^2 (in der Nähe von 1) sprechen für eine hohe Modellgüte, niedrige Werte (in der Nähe von 0) für eine geringe Modellgüte.

Vorsicht!

s_X^2, s_Y^2 sowie $s_{X,Y}$ bezeichnen in diesem Kapitel die **empirischen** Größen

$$s_X^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2, \qquad s_Y^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 = \overline{y^2} - \overline{y}^2$$

$$\text{und } s_{X,Y} = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})\cdot(y_i - \overline{y}) = \overline{xy} - \overline{x}\cdot\overline{y} \ .$$

Beispiel: Ausgaben in Abhängigkeit vom Einkommen (I)

- Es wird angenommen, dass die Ausgaben eines Haushalts für Nahrungs- und Genussmittel y_i linear vom jeweiligen Haushaltseinkommen x_i (jeweils in 100 €) in der Form

$$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \qquad u_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \qquad i \in \{1, \ldots, n\}$$

abhängen. Für $n = 7$ Haushalte beobachte man nun neben dem Einkommen x_i auch die (Realisation der) Ausgaben für Nahrungs- und Genussmittel y_i wie folgt:

Haushalt i	1	2	3	4	5	6	7
Einkommen x_i	35	49	21	39	15	28	25
NuG-Ausgaben y_i	9	15	7	11	5	8	9

- Mit Hilfe dieser Stichprobeninformation sollen nun die Parameter β_1 und β_2 der linearen Modellbeziehung geschätzt sowie die Werte $\widehat{y}_i$, die Residuen $\widehat{u}_i$ und das Bestimmtheitsmaß R^2 bestimmt werden.

- Berechnete (deskriptive/empirische) Größen:

$$\overline{x} = 30.28571 \quad \overline{y} = 9.14286 \quad \overline{x^2} = 1031.71429 \quad \overline{y^2} = 92.28571$$
$$s_X^2 = 114.4901 \quad s_Y^2 = 8.6938 \quad s_{X,Y} = 30.2449 \quad r_{X,Y} = 0.9587$$

- Damit erhält man die Parameterschätzer $\widehat{\beta}_1$ und $\widehat{\beta}_2$ als

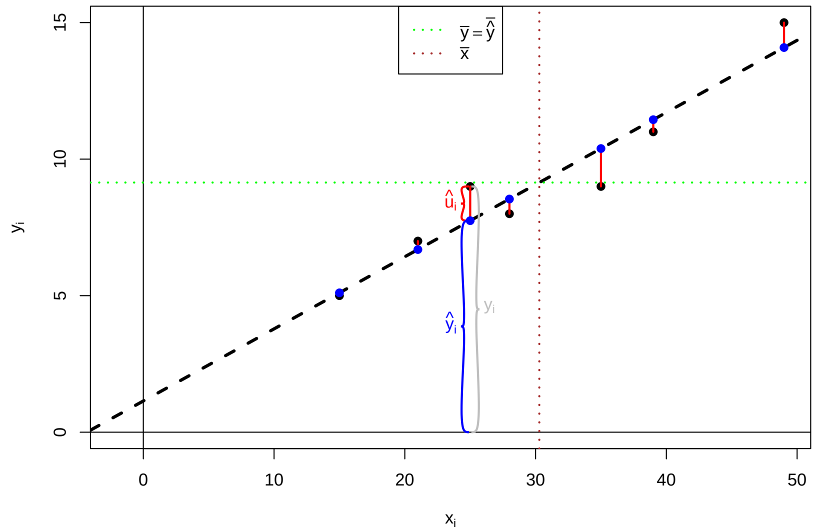
$$\widehat{\beta}_2 = \frac{s_{X,Y}}{s_X^2} = \frac{30.2449}{114.4901} = 0.26417$$
$$\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2 \cdot \overline{x} = 9.14286 - 0.26417 \cdot 30.28571 = 1.14228\ .$$

- Als Bestimmtheitsmaß erhält man $R^2 = r_{X,Y}^2 = 0.9587^2 = 0.9191$.
- Für $\widehat{y}_i$ und $\widehat{u}_i$ erhält man durch Einsetzen ($\widehat{y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_i$, $\widehat{u}_i = y_i - \widehat{y}_i$):

i	1	2	3	4	5	6	7
x_i	35	49	21	39	15	28	25
y_i	9	15	7	11	5	8	9
$\widehat{y}_i$	10.39	14.09	6.69	11.44	5.1	8.54	7.75
$\widehat{u}_i$	−1.39	0.91	0.31	−0.44	−0.1	−0.54	1.25

Grafik: Ausgaben in Abhängigkeit vom Einkommen

$\widehat{\beta}_1 = 1.14228$, $\widehat{\beta}_2 = 0.26417$, $R^2 = 0.9191$



Eigenschaften der Schätzfunktionen $\widehat{\beta}_1$ und $\widehat{\beta}_2$

- $\widehat{\beta}_1$ und $\widehat{\beta}_2$ sind **linear in** y_i, man kann genauer zeigen:

$$\widehat{\beta}_1 = \sum_{i=1}^{n} \frac{\overline{x^2} - \overline{x} \cdot x_i}{n(\overline{x^2} - \overline{x}^2)} \cdot y_i \qquad \text{und} \qquad \widehat{\beta}_2 = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{n(\overline{x^2} - \overline{x}^2)} \cdot y_i$$

- $\widehat{\beta}_1$ und $\widehat{\beta}_2$ sind **erwartungstreu für** β_1 **und** β_2, denn wegen $\mathrm{E}(u_i) = 0$ gilt
  - $\mathrm{E}(y_i) = \beta_1 + \beta_2 \cdot x_i + \mathrm{E}(u_i) = \beta_1 + \beta_2 \cdot x_i$,
  - $\mathrm{E}(\overline{y}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(y_i) = \frac{1}{n}\sum_{i=1}^n(\beta_1 + \beta_2 \cdot x_i) = \beta_1 + \beta_2 \cdot \overline{x}$,
  - $\mathrm{E}(\overline{xy}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right) = \frac{1}{n}\sum_{i=1}^n x_i(\beta_1 + \beta_2 \cdot x_i) = \beta_1 \cdot \overline{x} + \beta_2 \cdot \overline{x^2}$

  und damit

$$\mathrm{E}(\widehat{\beta}_2) = \mathrm{E}\left(\frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}\right) = \frac{\mathrm{E}(\overline{xy}) - \overline{x} \cdot \mathrm{E}(\overline{y})}{\overline{x^2} - \overline{x}^2}$$
$$= \frac{\beta_1 \cdot \overline{x} + \beta_2 \cdot \overline{x^2} - \overline{x} \cdot (\beta_1 + \beta_2 \cdot \overline{x})}{\overline{x^2} - \overline{x}^2} = \frac{\beta_2 \cdot (\overline{x^2} - \overline{x}^2)}{\overline{x^2} - \overline{x}^2} = \beta_2$$

  sowie

$$\mathrm{E}(\widehat{\beta}_1) = \mathrm{E}(\overline{y} - \overline{x}\widehat{\beta}_2) = \mathrm{E}(\overline{y}) - \overline{x}\,\mathrm{E}(\widehat{\beta}_2) = \beta_1 + \beta_2 \cdot \overline{x} - \overline{x} \cdot \beta_2 = \beta_1\ .$$

(Diese Eigenschaften folgen bereits mit dem Satz von Gauß-Markov.)

- Für die Varianzen der Schätzfunktionen erhält man:

$$\mathrm{Var}(\widehat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sigma^2}{n \cdot (\overline{x^2} - \overline{x}^2)} = \frac{\sigma^2}{n \cdot s_X^2}$$
$$\mathrm{Var}(\widehat{\beta}_1) = \frac{\sigma^2}{n} \cdot \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sigma^2 \cdot \overline{x^2}}{n \cdot (\overline{x^2} - \overline{x}^2)} = \frac{\sigma^2 \cdot \overline{x^2}}{n \cdot s_X^2}$$

  Diese hängen von der unbekannten Varianz σ^2 der u_i ab.
- Eine erwartungstreue Schätzfunktion für σ^2 ist gegeben durch

$$\widehat{\sigma^2} := \widehat{\mathrm{Var}(u_i)} = \frac{1}{n-2} \sum_{i=1}^{n} \widehat{u}_i^2$$
$$= \frac{n}{n-2} \cdot s_Y^2 \cdot (1 - R^2) = \frac{n}{n-2} \cdot (s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y})$$

- Die positive Wurzel $\widehat{\sigma} = +\sqrt{\widehat{\sigma^2}}$ dieser Schätzfunktion heißt auch **S**tandard **E**rror of the **R**egression (SER) oder **residual standard error**.

- Einsetzen des Schätzers $\widehat{\sigma^2}$ für σ^2 liefert die geschätzten Varianzen der Parameterschätzer

$$\widehat{\sigma^2}_{\widehat{\beta}_2} := \widehat{\mathrm{Var}(\widehat{\beta}_2)} = \frac{\widehat{\sigma^2}}{n \cdot (\overline{x^2} - \overline{x}^2)} = \frac{\widehat{\sigma^2}}{n \cdot s_X^2} = \frac{s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y}}{(n-2) \cdot s_X^2}$$

und

$$\widehat{\sigma^2}_{\widehat{\beta}_1} := \widehat{\mathrm{Var}(\widehat{\beta}_1)} = \frac{\widehat{\sigma^2} \cdot \overline{x^2}}{n \cdot (\overline{x^2} - \overline{x}^2)} = \frac{\widehat{\sigma^2} \cdot \overline{x^2}}{n \cdot s_X^2} = \frac{(s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y}) \cdot \overline{x^2}}{(n-2) \cdot s_X^2} \; .$$

- Die positiven Wurzeln $\widehat{\sigma}_{\widehat{\beta}_1} = \sqrt{\widehat{\sigma^2}_{\widehat{\beta}_1}}$ und $\widehat{\sigma}_{\widehat{\beta}_2} = \sqrt{\widehat{\sigma^2}_{\widehat{\beta}_2}}$ dieser geschätzten Varianzen werden wie üblich als (geschätzte) **Standardfehler** von $\widehat{\beta}_1$ und $\widehat{\beta}_2$ bezeichnet.
- Trifft man eine weitergehende Verteilungannahme für u_i und damit für y_i, so lassen sich auch die Verteilungen von $\widehat{\beta}_1$ und $\widehat{\beta}_2$ weiter untersuchen und zur Konstruktion von Tests, Konfidenzintervallen und *Prognoseintervallen* verwenden.

Konfidenzintervalle und Tests

unter Normalverteilungsannahme für u_i

- Häufig nimmt man für die Störgrößen an, dass speziell

$$u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

gilt, d.h. dass alle u_i (für $i \in \{1, \ldots, n\}$) unabhängig identisch normalverteilt sind mit Erwartungswert 0 und (unbekannter) Varianz σ^2.
- In diesem Fall sind offensichtlich auch $y_1, \ldots, y_n$ stochastisch unabhängig und jeweils normalverteilt mit Erwartungswert $\mathrm{E}(y_i) = \beta_1 + \beta_2 \cdot x_i$ und Varianz $\mathrm{Var}(y_i) = \sigma^2$.
- Da $\widehat{\beta}_1$ und $\widehat{\beta}_2$ linear in y_i sind, folgt insgesamt mit den bereits berechneten Momenten von $\widehat{\beta}_1$ und $\widehat{\beta}_2$:

$$\widehat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2 \cdot \overline{x^2}}{n \cdot s_X^2}\right) \qquad \text{und} \qquad \widehat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{n \cdot s_X^2}\right)$$

Konfidenzintervalle

unter Normalverteilungsannahme für u_i

- Da σ^2 unbekannt ist, ist für Anwendungen wesentlich relevanter, dass im Falle unabhängig identisch normalverteilter Störgrößen u_i mit den Schätzfunktionen $\widehat{\sigma^2}_{\widehat{\beta}_1}$ für $\mathrm{Var}(\widehat{\beta}_1)$ und $\widehat{\sigma^2}_{\widehat{\beta}_2}$ für $\mathrm{Var}(\widehat{\beta}_2)$ gilt:

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \sim t(n-2) \qquad \text{und} \qquad \frac{\widehat{\beta}_2 - \beta_2}{\widehat{\sigma}_{\widehat{\beta}_2}} \sim t(n-2)$$

- Hieraus erhält man unmittelbar die „Formeln"

$$\left[\widehat{\beta}_1 - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\widehat{\beta}_1}, \widehat{\beta}_1 + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\widehat{\beta}_1}\right]$$

für (symmetrische) Konfidenzintervalle zur Vertrauenswahrscheinlichkeit $1 - \alpha$ für β_1 bzw.

$$\left[\widehat{\beta}_2 - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\widehat{\beta}_2}, \widehat{\beta}_2 + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\widehat{\beta}_2}\right]$$

für (symmetrische) Konfidenzintervalle zur Vertrauenswahrscheinlichkeit $1 - \alpha$ für β_2.

Beispiel: Ausgaben in Abhängigkeit vom Einkommen (II)

- Im bereits erläuterten Beispiel erhält man als Schätzwert für σ^2:

$$\widehat{\sigma^2} = \frac{n \cdot (s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y})}{n-2} = \frac{7 \cdot (8.6938 - 0.26417 \cdot 30.2449)}{7-2} = 0.9856$$

- Die (geschätzten) Standardfehler für $\widehat{\beta}_1$ und $\widehat{\beta}_2$ sind damit

$$\widehat{\sigma}_{\widehat{\beta}_1} = \sqrt{\frac{\widehat{\sigma^2} \cdot \overline{x^2}}{n \cdot s_X^2}} = \sqrt{\frac{0.9856 \cdot 1031.71429}{7 \cdot 114.4901}} = 1.1264 \; ,$$

$$\widehat{\sigma}_{\widehat{\beta}_2} = \sqrt{\frac{\widehat{\sigma^2}}{n \cdot s_X^2}} = \sqrt{\frac{0.9856}{7 \cdot 114.4901}} = 0.0351 \; .$$

- Für $\alpha = 0.05$ erhält man mit $t_{n-2;1-\frac{\alpha}{2}} = t_{5;0.975} = 2.571$ für β_1 also

$$[1.14228 - 2.571 \cdot 1.1264, 1.14228 + 2.571 \cdot 1.1264] = [-1.7537, 4.0383]$$

als Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ bzw.

$$[0.26417 - 2.571 \cdot 0.0351, 0.26417 + 2.571 \cdot 0.0351] = [0.1739, 0.3544]$$

als Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_2.

Hypothesentests

unter Normalverteilungsannahme für u_i

- Genauso lassen sich unter der Normalverteilungsannahme (exakte) t-Tests für die Parameter β_1 und β_2 konstruieren.
- Trotz unterschiedlicher Problemstellung weisen die Tests Ähnlichkeiten zum t-Test für den Mittelwert einer normalverteilten Zufallsvariablen bei unbekannter Varianz auf.
- Untersucht werden können die Hypothesenpaare

$H_0 : \beta_1 = \beta_1^0$ gegen $H_1 : \beta_1 \neq \beta_1^0$ $\qquad$ $H_0 : \beta_1 \leq \beta_1^0$ gegen $H_1 : \beta_1 > \beta_1^0$ $\qquad$ $H_0 : \beta_1 \geq \beta_1^0$ gegen $H_1 : \beta_1 < \beta_1^0$

bzw.

$H_0 : \beta_2 = \beta_2^0$ gegen $H_1 : \beta_2 \neq \beta_2^0$ $\qquad$ $H_0 : \beta_2 \leq \beta_2^0$ gegen $H_1 : \beta_2 > \beta_2^0$ $\qquad$ $H_0 : \beta_2 \geq \beta_2^0$ gegen $H_1 : \beta_2 < \beta_2^0$

- Besonders anwendungsrelevant sind Tests auf die „Signifikanz" der Parameter (insbesondere β_2), die den zweiseitigen Tests mit $\beta_1^0 = 0$ bzw. $\beta_2^0 = 0$ entsprechen.

Zusammenfassung: t-Test für den Parameter β_1

im einfachen linearen Regressionsmodell mit Normalverteilungsannahme

Anwendungs-voraussetzungen	exakt: $y_i = \beta_1 + \beta_2 \cdot x_i + u_i$ mit $u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ für $i \in \{1, \ldots, n\}$, σ^2 unbekannt, $x_1, \ldots, x_n$ deterministisch und bekannt, Realisation $y_1, \ldots, y_n$ beobachtet		
Nullhypothese Gegenhypothese	$H_0 : \beta_1 = \beta_1^0$ $H_1 : \beta_1 \neq \beta_1^0$	$H_0 : \beta_1 \leq \beta_1^0$ $H_1 : \beta_1 > \beta_1^0$	$H_0 : \beta_1 \geq \beta_1^0$ $H_1 : \beta_1 < \beta_1^0$
Teststatistik	$t = \frac{\widehat{\beta}_1 - \beta_1^0}{\widehat{\sigma}_{\widehat{\beta}_1}}$		
Verteilung (H_0)	t für $\beta_1 = \beta_1^0$ $t(n-2)$-verteilt		
Benötigte Größen	$\widehat{\beta}_2 = \frac{s_{X,Y}}{s_X^2}, \widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2 \cdot \overline{x}, \widehat{\sigma}_{\widehat{\beta}_1} = \sqrt{\frac{(s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y}) \cdot \overline{x^2}}{(n-2) \cdot s_X^2}}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n-2;1-\frac{\alpha}{2}}) \cup (t_{n-2;1-\frac{\alpha}{2}}, \infty)$	$(t_{n-2;1-\alpha}, \infty)$	$(-\infty, -t_{n-2;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n-2)}(\lvert t \rvert))$	$1 - F_{t(n-2)}(t)$	$F_{t(n-2)}(t)$

Zusammenfassung: t-Test für den Parameter β_2

im einfachen linearen Regressionsmodell mit Normalverteilungsannahme

Anwendungs-voraussetzungen	exakt: $y_i = \beta_1 + \beta_2 \cdot x_i + u_i$ mit $u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ für $i \in \{1, \ldots, n\}$, σ^2 unbekannt, $x_1, \ldots, x_n$ deterministisch und bekannt, Realisation $y_1, \ldots, y_n$ beobachtet		
Nullhypothese Gegenhypothese	$H_0 : \beta_2 = \beta_2^0$ $H_1 : \beta_2 \neq \beta_2^0$	$H_0 : \beta_2 \leq \beta_2^0$ $H_1 : \beta_2 > \beta_2^0$	$H_0 : \beta_2 \geq \beta_2^0$ $H_1 : \beta_2 < \beta_2^0$
Teststatistik	$t = \frac{\widehat{\beta}_2 - \beta_2^0}{\widehat{\sigma}_{\widehat{\beta}_2}}$		
Verteilung (H_0)	t für $\beta_2 = \beta_2^0$ $t(n-2)$-verteilt		
Benötigte Größen	$\widehat{\beta}_2 = \frac{s_{X,Y}}{s_X^2}, \widehat{\sigma}_{\widehat{\beta}_2} = \sqrt{\frac{s_Y^2 - \widehat{\beta}_2 \cdot s_{X,Y}}{(n-2) \cdot s_X^2}}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n-2;1-\frac{\alpha}{2}}) \cup (t_{n-2;1-\frac{\alpha}{2}}, \infty)$	$(t_{n-2;1-\alpha}, \infty)$	$(-\infty, -t_{n-2;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n-2)}(\lvert t \rvert))$	$1 - F_{t(n-2)}(t)$	$F_{t(n-2)}(t)$

Beispiel: Ausgaben in Abhängigkeit vom Einkommen (III)

- Im bereits erläuterten Beispiel soll zum Signifikanzniveau $\alpha = 0.05$ getestet werden, ob β_1 signifikant von Null verschieden ist. Geeigneter Test: **t-Test für den Regressionsparameter β_1**

1. **Hypothesen:**
   $H_0 : \beta_1 = 0$ gegen $H_1 : \beta_1 \neq 0$
2. **Teststatistik:**
   $t = \frac{\widehat{\beta}_1 - 0}{\widehat{\sigma}_{\widehat{\beta}_1}}$ ist unter H_0 (für $\beta_1 = 0$) $t(n-2)$-verteilt.
3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**
   $K = (-\infty, -t_{n-2;1-\frac{\alpha}{2}}) \cup (t_{n-2;1-\frac{\alpha}{2}}, +\infty) = (-\infty, -t_{5;0.975}) \cup (t_{5;0.975}, +\infty)$
   $= (-\infty, -2.571) \cup (2.571, +\infty)$
4. **Berechnung der realisierten Teststatistik:**
   $t = \frac{\widehat{\beta}_1 - 0}{\widehat{\sigma}_{\widehat{\beta}_1}} = \frac{1.14228 - 0}{1.1264} = 1.014$
5. **Entscheidung:**
   $t = 1.014 \notin (-\infty, -2.571) \cup (2.571, +\infty) = K \Rightarrow H_0$ wird nicht abgelehnt!
   (p-Wert: $2 - 2 \cdot F_{t(5)}(\lvert t \rvert) = 2 - 2 \cdot F_{t(5)}(\lvert 1.014 \rvert) = 2 - 2 \cdot 0.8215 = 0.357$)

Der Test kann für β_1 keine signifikante Abweichung von Null feststellen.

Beispiel: Ausgaben in Abhängigkeit vom Einkommen (IV)

- Nun soll zum Signifikanzniveau $\alpha = 0.01$ getestet werden, ob β_2 **positiv** ist.
  Geeigneter Test:
  t-Test für den Regressionsparameter β_2
  1. **Hypothesen:**
     $H_0 : \beta_2 \leq 0 \qquad \text{gegen} \qquad H_1 : \beta_2 > 0$
  2. **Teststatistik:**
     $t = \frac{\widehat{\beta}_2 - 0}{\widehat{\sigma}_{\widehat{\beta}_2}}$ ist unter H_0 (für $\beta_2 = 0$) $t(n-2)$-verteilt.
  3. **Kritischer Bereich zum Niveau $\alpha = 0.01$:**
     $K = (t_{n-2;1-\alpha}, +\infty) = (t_{5;0.99}, +\infty) = (3.365, +\infty)$
  4. **Berechnung der realisierten Teststatistik:**
     $t = \frac{\widehat{\beta}_2 - 0}{\widehat{\sigma}_{\widehat{\beta}_2}} = \frac{0.26417 - 0}{0.0351} = 7.5262$
  5. **Entscheidung:**
     $t = 7.5262 \in (3.365, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
     (p-Wert: $1 - F_{t(5)}(t) = 1 - F_{t(5)}(7.5262) = 1 - 0.9997 = 0.0003$)

  Der Test stellt fest, dass β_2 signifikant positiv ist.

Punkt- und Intervallprognosen

im einfachen linearen Regressionsmodell mit Normalverteilungsannahme

- Neben Konfidenzintervallen und Tests für die Parameter β_1 und β_2 in linearen Regressionsmodellen vor allem **Prognosen** wichtige Anwendung.
- Zur Erstellung von Prognosen: Erweiterung der Modellannahme
  $$y_i = \beta_1 + \beta_2 \cdot x_i + u_i, \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad i \in \{1, \dots, n\}$$
  auf (zumindest) einen weiteren, hier mit (x_0, y_0) bezeichneten Datenpunkt, bei dem jedoch y_0 **nicht** beobachtet wird, sondern lediglich der Wert des Regressors x_0 bekannt ist.
- *Ziel:* „Schätzung" (Prognose) von $y_0 = \beta_1 + \beta_2 \cdot x_0 + u_0$ bzw. $\mathrm{E}(y_0) = \beta_1 + \beta_2 \cdot x_0$ auf Grundlage von x_0.
- Wegen $\mathrm{E}(u_0) = 0$ und der Erwartungstreue von $\widehat{\beta}_1$ für β_1 bzw. $\widehat{\beta}_2$ für β_2 ist
  $$\widehat{y}_0 := \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0 =: \widehat{\mathrm{E}(y_0)}$$
  offensichtlich erwartungstreu für y_0 bzw. $\mathrm{E}(y_0)$ gegeben x_0.
- $\widehat{y}_0$ bzw. $\widehat{\mathrm{E}(y_0)}$ wird auch **(bedingte) Punktprognose für y_0 bzw. $\mathrm{E}(y_0)$ gegeben x_0** genannt.

Prognosefehler

- Zur Beurteilung der Genauigkeit der Prognosen:
  Untersuchung der sogenannten Prognosefehler
  $$\widehat{y}_0 - y_0 \qquad \text{bzw.} \qquad \widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)\ .$$
- Qualitativer Unterschied:
  - Prognosefehler
    $$\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0) = \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0 - (\beta_1 + \beta_2 \cdot x_0) = (\widehat{\beta}_1 - \beta_1) + (\widehat{\beta}_2 - \beta_2) \cdot x_0$$
    resultiert **nur** aus Fehler bei der Schätzung von β_1 bzw. β_2 durch $\widehat{\beta}_1$ bzw. $\widehat{\beta}_2$.
  - Prognosefehler
    $$\widehat{y}_0 - y_0 = \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0 - (\beta_1 + \beta_2 \cdot x_0 + u_0) = (\widehat{\beta}_1 - \beta_1) + (\widehat{\beta}_2 - \beta_2) \cdot x_0 - u_0$$
    ist Kombination von Schätzfehlern (für β_1 und β_2) sowie zufälliger Schwankung von $u_0 \sim N(0, \sigma^2)$.
- Zunächst: Untersuchung von $e_E := \widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)$

- Wegen der Erwartungstreue stimmen mittlerer quadratischer (Prognose-) Fehler und Varianz von $e_E = \widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)$ überein und man erhält
  $$\begin{aligned}\mathrm{Var}(\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)) = \mathrm{Var}(\widehat{\mathrm{E}(y_0)}) &= \mathrm{Var}(\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0) \\ &= \mathrm{Var}(\widehat{\beta}_1) + x_0^2\,\mathrm{Var}(\widehat{\beta}_2) + 2 \cdot x_0 \cdot \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2).\end{aligned}$$
- Es kann gezeigt werden, dass für die Kovarianz von $\widehat{\beta}_1$ und $\widehat{\beta}_2$ gilt:
  $$\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) = -\sigma^2 \cdot \frac{\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = -\sigma^2 \cdot \frac{\overline{x}}{n \cdot s_X^2}$$
- Insgesamt berechnet man so die Varianz des Prognosefehlers
  $$\begin{aligned}\sigma_{e_E}^2 := \mathrm{Var}(e_E) &= \frac{\sigma^2 \cdot \overline{x^2}}{n \cdot s_X^2} + x_0^2 \cdot \frac{\sigma^2}{n \cdot s_X^2} - 2 \cdot x_0 \cdot \frac{\sigma^2 \cdot \overline{x}}{n \cdot s_X^2} \\ &= \sigma^2 \cdot \frac{\overline{x^2} + x_0^2 - 2 \cdot x_0 \cdot \overline{x}}{n \cdot s_X^2} \\ &= \sigma^2 \cdot \frac{(\overline{x^2} - \overline{x}^2) + (\overline{x}^2 + x_0^2 - 2 \cdot x_0 \cdot \overline{x})}{n \cdot s_X^2} \\ &= \sigma^2 \cdot \frac{s_X^2 + (x_0 - \overline{x})^2}{n \cdot s_X^2} = \sigma^2 \cdot \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right)\ .\end{aligned}$$

- Die Linearität von $\widehat{\beta}_1$ und $\widehat{\beta}_2$ (in y_i) überträgt sich (natürlich) auch auf $\widehat{\mathrm{E}(y_0)}$, damit gilt offensichtlich

$$e_E = \widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0) \sim N\left(0, \sigma_{e_E}^2\right) \qquad \text{bzw.} \qquad \frac{\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)}{\sigma_{e_E}} \sim N(0,1) \; .$$

- Da σ^2 unbekannt ist, erhält man durch Ersetzen von σ^2 durch die erwartungstreue Schätzfunktion $\widehat{\sigma^2}$ die geschätzte Varianz

$$\widehat{\sigma^2}_{e_E} := \widehat{\mathrm{Var}}(e_E) = \widehat{\sigma^2} \cdot \frac{s_X^2 + (x_0 - \overline{x})^2}{n \cdot s_X^2} = \widehat{\sigma^2} \cdot \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right)$$

von $\widehat{\mathrm{E}(y_0)}$ und damit die praktisch wesentlich relevantere Verteilungsaussage

$$\frac{e_E}{\widehat{\sigma}_{e_E}} = \frac{\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)}{\widehat{\sigma}_{e_E}} \sim t(n-2) \; ,$$

aus der sich in bekannter Weise (symmetrische) Konfidenzintervalle (und Tests) konstruieren lassen.

Prognoseintervalle für $\mathrm{E}(y_0)$ gegeben x_0

- Intervallprognosen zur Vertrauenswahrscheinlichkeit $1-\alpha$ erhält man also als Konfidenzintervalle zum Konfidenzniveau $1-\alpha$ für $\mathrm{E}(y_0)$ in der Form

$$\left[\widehat{\mathrm{E}(y_0)} - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_E} \, , \, \widehat{\mathrm{E}(y_0)} + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_E}\right]$$
$$= \left[(\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0) - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_E} \, , \, (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0) + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_E}\right] \; .$$

- Im Beispiel (Ausgaben in Abhängigkeit vom Einkommen) erhält man zu gegebenem $x_0 = 50$ (in 100 €)

$$\widehat{\sigma^2}_{e_E} = \widehat{\sigma^2} \cdot \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right) = 0.9856 \cdot \left(\frac{1}{7} + \frac{(50 - 30.28571)^2}{7 \cdot 114.4901}\right) = 0.6188$$

die Punktprognose $\widehat{\mathrm{E}(y_0)} = \widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0 = 1.14228 + 0.26417 \cdot 50 = 14.3508$ (in 100 €) sowie die Intervallprognose zur Vertrauenswahrscheinlichkeit 0.95

$$\left[14.3508 - 2.571 \cdot \sqrt{0.6188} \, , \, 14.3508 + 2.571 \cdot \sqrt{0.6188}\right]$$
$$= [12.3284 \, , \, 16.3732] \;\; (\text{in } 100 \text{ €}) \; .$$

Prognosefehler $e_0 := \widehat{y}_0 - y_0$

- *Nun:* Untersuchung des Prognosefehlers $e_0 := \widehat{y}_0 - y_0$
- Offensichtlich gilt für $e_0 = \widehat{y}_0 - y_0$ die Zerlegung

$$\begin{aligned}\widehat{y}_0 - y_0 &= \underbrace{(\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0)}_{=\widehat{\mathrm{E}(y_0)}} - (\underbrace{\beta_1 + \beta_2 \cdot x_0}_{=\mathrm{E}(y_0)} + u_0) \\ &= \underbrace{\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)}_{\substack{\text{Fehler aus Schätzung von} \\ \beta_1 \text{ und } \beta_2}} \quad - \quad \underbrace{u_0}_{\substack{\text{zufällige Schwankung} \\ \text{der Störgröße}}} \; .\end{aligned}$$

- $\widehat{\mathrm{E}(y_0)}$ hängt nur von $u_1, \ldots, u_n$ ab (über $y_1, \ldots, y_n$ bzw. $\widehat{\beta}_1$ und $\widehat{\beta}_2$) und ist wegen der Annahme $u_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ **unabhängig** von u_0.
- Damit sind die beiden Bestandteile des Prognosefehlers insbesondere auch unkorreliert und man erhält:

$$\begin{aligned}\sigma_{e_0}^2 &:= \mathrm{Var}(\widehat{y}_0 - y_0) = \mathrm{Var}(\widehat{\mathrm{E}(y_0)} - \mathrm{E}(y_0)) + \mathrm{Var}(u_0) \\ &= \sigma^2 \cdot \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right) + \sigma^2 = \sigma^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right)\end{aligned}$$

- Aus der Unkorreliertheit der beiden Komponenten des Prognosefehlers folgt auch sofort die Normalverteilungseigenschaft des Prognosefehlers $e_0 = y_0 - \widehat{y}_0$, genauer gilt:

$$e_0 = \widehat{y}_0 - y_0 \sim N\left(0, \sigma_{e_0}^2\right) \qquad \text{bzw.} \qquad \frac{\widehat{y}_0 - y_0}{\sigma_{e_0}} \sim N(0,1) \; .$$

- Wieder muss σ^2 durch $\widehat{\sigma^2}$ ersetzt werden, um mit Hilfe der geschätzen Varianz

$$\widehat{\sigma^2}_{e_0} := \widehat{\mathrm{Var}}(\widehat{y}_0 - y_0) = \widehat{\sigma^2} \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right)$$

des Prognosefehlers die für die Praxis relevante Verteilungsaussage

$$\frac{e_0}{\widehat{\sigma}_{e_0}} = \frac{\widehat{y}_0 - y_0}{\widehat{\sigma}_{e_0}} \sim t(n-2) \; ,$$

zu erhalten, aus der sich dann wieder Prognoseintervalle konstruieren lassen.

Prognoseintervalle für y_0 gegeben x_0

- Intervallprognosen für y_0 zur Vertrauenswahrscheinlichkeit $1-\alpha$ erhält man also analog zu den Intervallprognosen für $\mathrm{E}(y_0)$ in der Form

$$\left[\widehat{y_0} - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_0}\,,\, \widehat{y_0} + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_0}\right]$$
$$= \left[(\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0) - t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_0}\,,\, (\widehat{\beta}_1 + \widehat{\beta}_2 \cdot x_0) + t_{n-2;1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{e_0}\right] .$$

- Im Beispiel (Ausgaben in Abhängigkeit vom Einkommen) erhält man zu gegebenem $x_0 = 50$ (in 100 €)

$$\widehat{\sigma^2}_{e_0} = \widehat{\sigma^2} \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n \cdot s_X^2}\right) = 0.9856 \cdot \left(1 + \frac{1}{7} + \frac{(50 - 30.28571)^2}{7 \cdot 114.4901}\right) = 1.6044$$

mit der bereits berechneten Punktprognose $\widehat{y_0} = \widehat{\mathrm{E}(y_0)} = 14.3508$ (in 100 €) die zugehörige Intervallprognose für y_0 zur Vertrauenswahrscheinlichkeit 0.95

$$\left[14.3508 - 2.571 \cdot \sqrt{1.6044}\,,\, 14.3508 + 2.571 \cdot \sqrt{1.6044}\right]$$
$$= [11.0942\,,\, 17.6074] \ \text{(in 100 €)} .$$

Lineare Modelle mit Statistik-Software **R**

Beispiel (Ausgaben in Abhängigkeit vom Einkommen)

- Modellschätzung mit aussagekräftiger Zusammenfassung in nur einer Zeile:

```
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
      1       2       3       4       5       6       7
-1.3882  0.9134  0.3102 -0.4449 -0.1048 -0.5390  1.2535

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14225    1.12645   1.014 0.357100
x            0.26417    0.03507   7.533 0.000653 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9928 on 5 degrees of freedom
Multiple R-squared:  0.919, Adjusted R-squared:  0.9028
F-statistic: 56.74 on 1 and 5 DF,  p-value: 0.0006529
```


Interpretation des Outputs (I)

Residuen, $\widehat{\sigma^2}$ und R^2

```
Residuals:
      1       2       3       4       5       6       7
-1.3882  0.9134  0.3102 -0.4449 -0.1048 -0.5390  1.2535

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14225    1.12645   1.014 0.357100
x            0.26417    0.03507   7.533 0.000653 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9928 on 5 degrees of freedom
Multiple R-squared:  0.919, Adjusted R-squared:  0.9028
F-statistic: 56.74 on 1 and 5 DF,  p-value: 0.0006529
```

- Auflistung bzw. Zusammenfassung der Residuen $\widehat{u}_i$
- Geschätzte Standardabweichung $\widehat{\sigma} = \sqrt{\widehat{\sigma^2}}$, hier: $\widehat{\sigma} = 0.9928 \Rightarrow \widehat{\sigma^2} = 0.9857$
- Anzahl Freiheitsgrade $n-2$, hier: $n-2=5 \Rightarrow n=7$
- (Multiples) Bestimmtheitsmaß R^2, hier: $R^2 = 0.919$

Interpretation des Outputs (II)

Ergebnisse zur Schätzung von β_1 und β_2

```
Residuals:
      1       2       3       4       5       6       7
-1.3882  0.9134  0.3102 -0.4449 -0.1048 -0.5390  1.2535

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14225    1.12645   1.014 0.357100
x            0.26417    0.03507   7.533 0.000653 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9928 on 5 degrees of freedom
Multiple R-squared:  0.919, Adjusted R-squared:  0.9028
F-statistic: 56.74 on 1 and 5 DF,  p-value: 0.0006529
```

- Realisationen von $\widehat{\beta}_1$, $\widehat{\beta}_2$, hier: $\widehat{\beta}_1 = 1.14225$, $\widehat{\beta}_2 = 0.26417$
- Standardfehler von $\widehat{\beta}_1$, $\widehat{\beta}_2$, hier: $\widehat{\sigma}_{\widehat{\beta}_1} = 1.12645$, $\widehat{\sigma}_{\widehat{\beta}_2} = 0.03507$
- t-Statistiken zu Tests auf Signifikanz, hier: zu $\beta_1 : t = 1.014$, zu $\beta_2 : t = 7.533$
- p-Werte zu Tests auf Signifikanz, hier: zu $\beta_1 : p = 0.3571$, zu $\beta_2 : p = 0.000653$

Zusammenhang zwischen p-Werten

zu zweiseitigen und einseitigen Tests bei unter H_0 (um Null) symmetrisch verteilter Teststatistik

- Erinnerung: $t(n)$- sowie $N(0,1)$-Verteilung sind symmetrisch um Null, für die zugehörigen Verteilungsfunktionen F gilt also $F(x) = 1 - F(-x)$ für alle $x \in \mathbb{R}$ und $F(0) = 0.5$, $F(x) < 0.5$ für $x < 0$ sowie $F(x) > 0.5$ für $x > 0$.
- Für die p-Werte p_z der zweiseitigen Tests auf den Mittelwert bei bekannter (Gauß-Test) sowie unbekannter (t-Test) Varianz gilt daher bekanntlich

$$p_z = 2 \cdot \min\{F(x), 1 - F(x)\} = \begin{cases} 2 \cdot F(x) & \text{falls } x < 0 \\ 2 \cdot (1 - F(x)) & \text{falls } x \geq 0 \end{cases} ,$$

wobei x den realisierten Wert der Teststatistik sowie F die Verteilungsfunktion der Teststatistik unter H_0 bezeichne.
- Für die p-Werte $p_l = F(x)$ zum linksseitigen sowie $p_r = 1 - F(x)$ zum rechtsseitigen Test bei realisierter Teststatistik x gelten demnach die folgenden Zusammenhänge:

$$p_l = \begin{cases} \frac{p_z}{2} & \text{falls } x < 0 \\ 1 - \frac{p_z}{2} & \text{falls } x \geq 0 \end{cases} \quad \text{sowie} \quad p_r = \begin{cases} 1 - \frac{p_z}{2} & \text{falls } x < 0 \\ \frac{p_z}{2} & \text{falls } x \geq 0 \end{cases} .$$

- Somit auch p-Werte zu einseitigen Tests aus **R**-Output bestimmbar!

Verallgemeinerungen des einfachen linearen Modells

- Zahlreiche Verallgemeinerungen des einfachen linearen Modells möglich.
- Statt einem Regressor mehrere Regressoren $\rightsquigarrow$ multiples Regressionsmodell.
- Statt unabhängiger identisch verteilter Störgrößen (z.B.)
  - unabhängige Störgrößen mit unterschiedlichen Varianzen,
  - abhängige (korrelierte) Störgrößen.
- Statt deterministischer Regressoren stochastische Regressoren.
- Statt nur einer Gleichung für einen Regressanden (simultane) Betrachtung mehrerer Regressanden $\rightsquigarrow$ Mehrgleichungsmodelle.
- Über Betrachtung linearer Abhängigkeiten hinaus auch nichtlineare Regressionsmodelle möglich.
- Verallgemeinerungen werden in weiterführenden Vorlesungen diskutiert, insbesondere „Ökonometrie“ (Bachelorstudiengang) und „Econometric Methods and Applications“ (Masterstudiengang Economics, Finance, and Philosophy).