Leuphana Universität Lüneburg

Statistik II

Wahrscheinlichkeitsrechnung und induktive Statistik

Fakultät Wirtschaft
Professur 'Statistik und Freie Berufe'
Univ.-Prof. Dr. Joachim Merz

Skriptum zur Vorlesung

Elfte verbesserte Auflage 2015
Impressum: Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik,
herausgegeben von der Leuphana Universität Lüneburg,
Fakultät Wirtschaft
Univ.-Prof. Dr. Joachim Merz,
Forschungsinstitut Freie Berufe, Professur 'Statistik und Freie Berufe'
www.leuphana.de/ffb,
Gedruckt auf 100% Altpapier und chlorfrei gebleichtem Papier.

Vorwort

Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik ist Thema dieses Skriptums zu meiner gleichlautenden Vorlesung an der Universität Lüneburg.

Das Skriptum soll vorlesungsbegleitend helfen, den Blick auf das Wesentliche, auf das Verständnis der Methoden und ihrer Anwendungen zu erleichtern. Im Rahmen einer anwendungsorientierten Statistik stehen bei der Auswahl des Stoffes Beispiele und Bezüge aus den Wirtschafts- und Sozialwissenschaften im Vordergrund.

Ich empfehle, den Stoff mit der angegebenen Literatur zu vertiefen; manchmal hilft ein anderer Blickwinkel, die Dinge besser zu begreifen. Das Verstehen, das verständige Umgehen mit der Statistik als ein wesentlicher Baustein, Theorie mit der Empirie zu verbinden, ist mir ein wichtiges Anliegen.

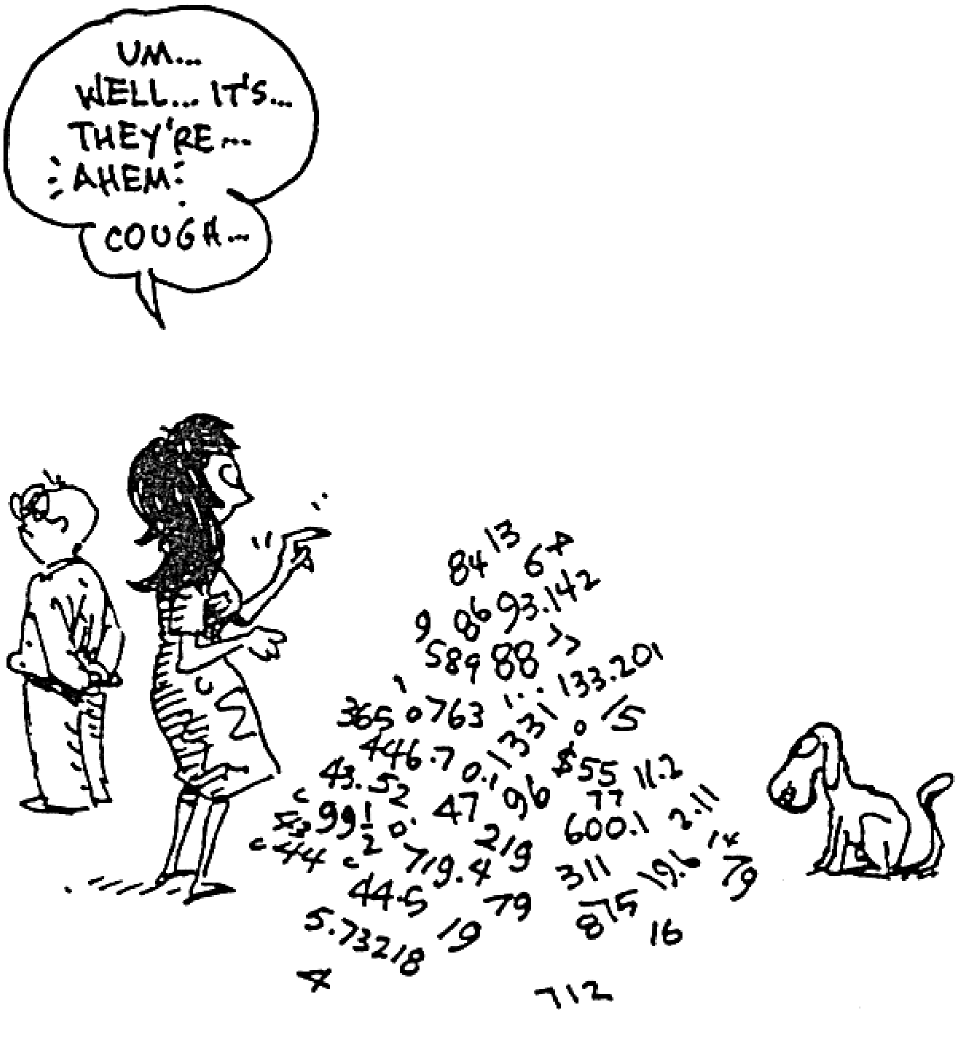
Für den problemorientierten Einstieg und den Umgang mit dem Computer werden verschiedene Programmpakete wie ET (Econometrics Toolkit), LIMDEP, Stata, SPSS und andere Programmpakete verwendet.

Die elfte neue Auflage entspricht bis auf kleinere Änderungen in Kapitel 8 der bisherigen Auflage.

Nicht zu vergessen: Studium und späterer Beruf sollen auch Spaß machen. Die Cartoons im Skriptum sind entsprechende Lockerungsübungen.

Viel Spaß und Erfolg!

Lüneburg, im Februar 2015 Univ.-Prof. Dr. Joachim Merz



Gonick, Smith 1993

Prof. Dr. Joachim Merz

STATISTIK II – Wahrscheinlichkeitsrechnung und induktive Statistik

THEMENBEREICHE

Prof. Dr. Joachim Merz

STATISTIK II – Wahrscheinlichkeitsrechnung und induktive Statistik

GLIEDERUNG

Vorbemerkungen zu Statistik I und II: Deskriptive Statistik, Wahrscheinlichkeitsrechnung und induktive Statistik

Statistik I – Deskription:

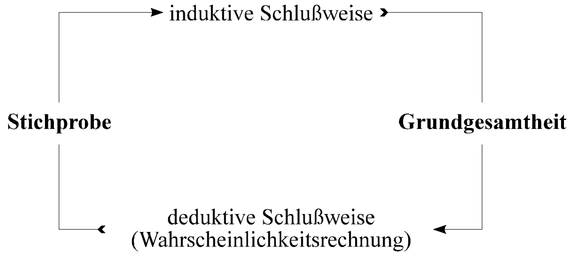
Beschreibende Statistik mit Verfahren zur Aufbereitung statistischer Daten bezogen auf die beobachteten Werte (Informationsaufbereitung)

Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik:

Zur Überprüfung allgemein gültiger Theorien; Informationsbewertung durch Inferenz- (schließende) Statistik: Wahrscheinlichkeitsaussagen über die Vereinbarkeit der in den Daten erfaßten Realität (Empirie) mit den aus einer Theorie abgeleiteten Hypothesen.

Wahrscheinlichkeitsrechnung notwendig, um von kostengünstigeren Teilerhebungen (Stichproben, 'sample') auf eine Grundgesamtheit zu schließen (induktive Statistik).

Teilerhebungen statt Vollerhebungen:
- aus Kostengründen,
- aus Zeitgründen,
- aus technischen Gründen etc.



Die Schlußweisen in der Statistik

Zum Aufbau von Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik

Wahrscheinlichkeitsrechnung und Wahrscheinlichkeitsverteilungen

- Grundzüge
- Zufallsvariablen und Wahrscheinlichkeitsverteilungen
- diskrete, stetige Verteilungen

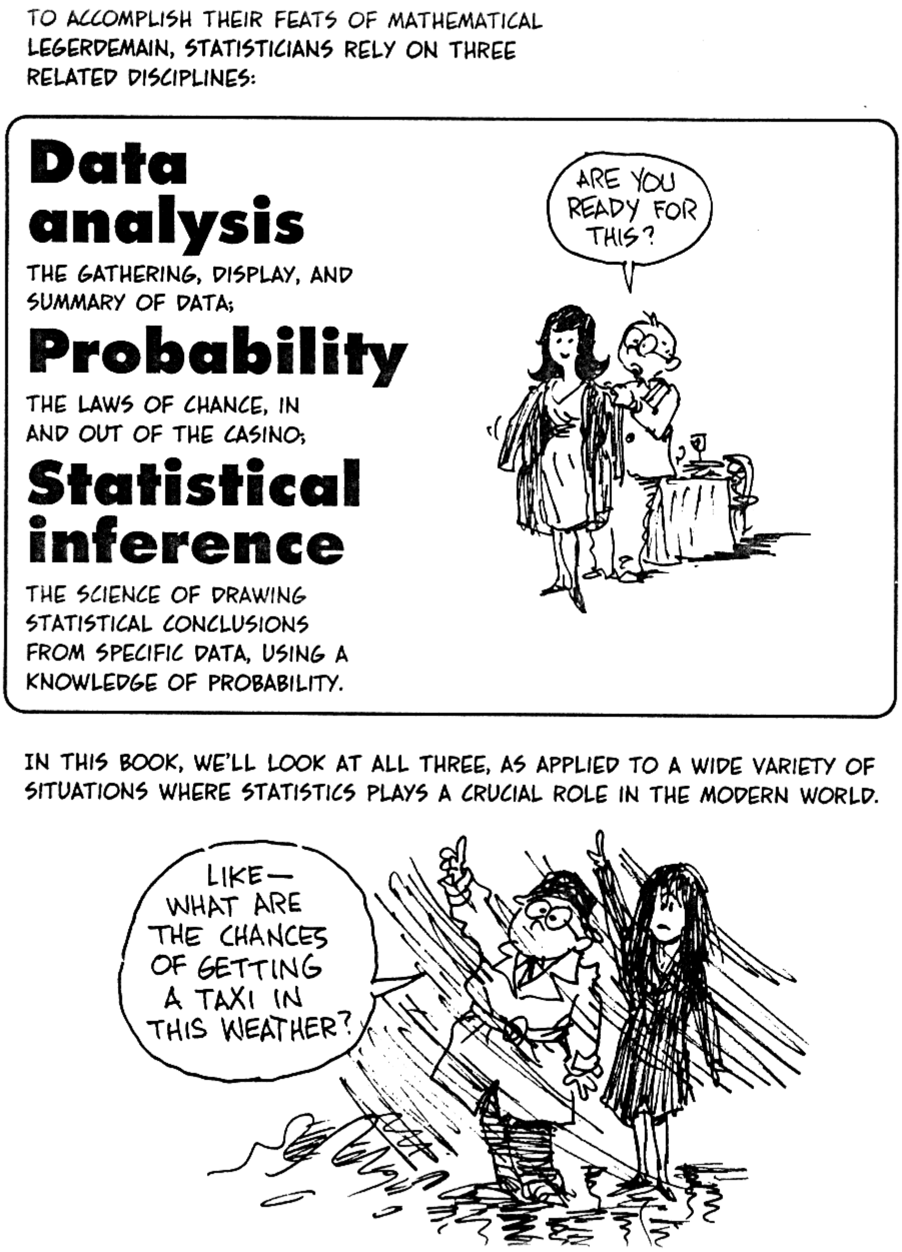
Stichprobe und Grundgesamtheit

- Induktive Statistik, Stichprobenfunktionen und Testverteilungen
- Punkt- und Intervallschätzung

Formulierung und Überprüfung von Hypothesen

- Parametertests
- Verteilungstests

Computerprogramme



Gonick, Smith 1993

I Grundzüge der Wahrscheinlichkeitsrechnung

Regeln der Wahrscheinlichkeitsrechnung als Grundlage für die induktive Statistik

Wofür wird die Wahrscheinlichkeitsrechnung benötigt?

Um die Beziehungen zwischen Stichprobe und Grundgesamtheit zu erfassen:

- Schätzung von Kenngrößen der unbekannten Grundgesamtheit, Angabe von Vertrauensbereichen (Konfidenzintervalle) aufgrund von Stichprobenergebnissen bei vorgegebener Sicherheitswahrscheinlichkeit für diese Kenngrößen (Parameter).
- Überprüfung von Hypothesen über Zusammenhänge in der Grundgesamtheit aufgrund von Stichprobenergebnissen

Da es unmöglich ist, Beziehungen zwischen sozioökonomischen Variablen in Modellen „exakt“ zu erfassen (**deterministisches Modell**), sondern jedes sozioökonomische Merkmal durch viele Faktoren mit Zufallseinflüssen beeinflußt wird, sind auch in Modellen über die Welt Zufallseinflüsse zu berücksichtigen. Man wird die Modellaussagen auf die wichtigsten Faktoren beschränken (**systematische Einflüsse**), für die übrigen im einzelnen nicht erfaßten Faktoren wird ein bestimmtes Verteilungsmodell unterstellt (**Zufall**). Damit besteht ein **stochastisches Erklärungsmodell** aus einem **systematischen und einem zufälligen Teil**. Die Wahrscheinlichkeitsrechnung ist vor allem für den Umgang mit dem zufälligen Teil und zur Fehlerabschätzung notwendig.

1 Grundbegriffe

Die Grundlagen der Wahrscheinlichkeitsrechnung wurden im 16. und 17. Jahrhundert von Blaise PASCAL (1623–1661) und Pierre FERMAT (1601–1668) gelegt; Ableitung von Wahrscheinlichkeiten für die Gewinnaussichten von Glücksspielen.

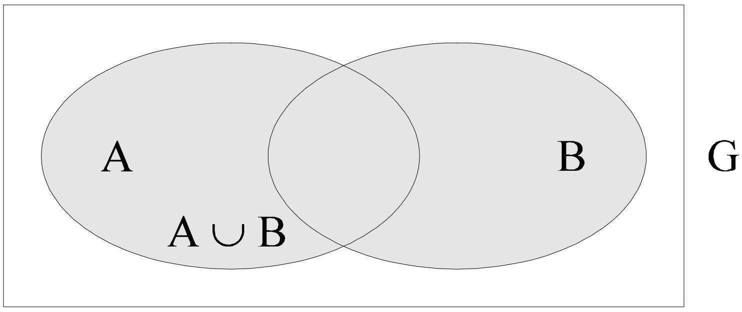
Umgangssprache: Wahrscheinlich bestehen Sie die Statistik II-Klausur.

Konkreter: Die Wahrscheinlichkeit, im Lotto mit 6 aus 49 sechs Richtige zu haben, beträgt: $\frac{1}{13\,983\,816}$.

Wichtige Begriffe:

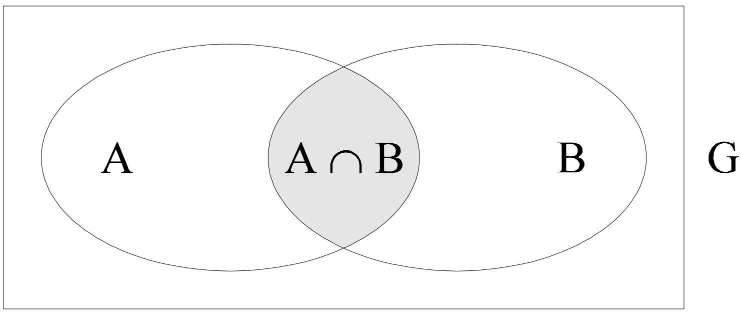
- **Zufallsexperiment:**
  Wirklich oder wenigstens gedanklich wiederholbarer Vorgang, dessen Ergebnis vom Zufall abhängt, also im voraus nicht eindeutig bestimmt werden kann.
  Beispiele: Werfen einer Münze, eines Würfels;
  Wichtig: Ergebnisse sind *unabhängig* voneinander.
- **Elementarereignisse:**
  Ergebnisse (Realisationen) des Zufallsexperiments: E_i

- **Ereignisraum G:**
  Menge der Elementarereignisse $G = \{E_1, E_2, \ldots, E_n\}$
  Beispiel: Einmaliges Werfen eines Würfels: $G = \{1, 2, 3, 4, 5, 6\}$
- **Ereignis:**
  Teilmenge von G
  Beispiel: Werfen einer ungeraden Augenzahl: $A = \{1, 3, 5\}$
- **Verknüpfung von Ereignissen:**
  Aus Ereignissen lassen sich mit bestimmten Operationen neue Ereignisse bilden:
  - **Vereinigung** zweier Ereignisse A und B: A ∪ B
    ist die Menge aller Ereignisse, *die entweder zu A, zu B oder zu A und B gemeinsam gehören*.



Venn-Diagramm

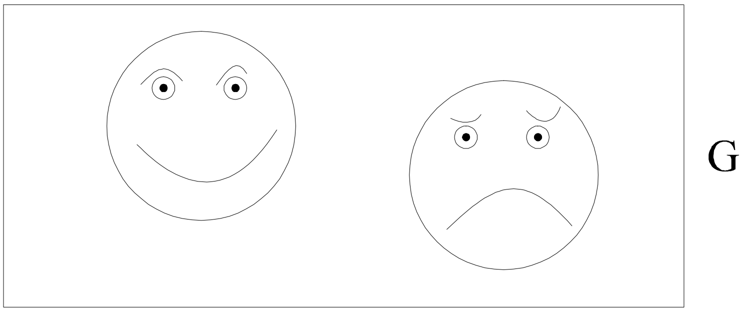
  - **Durchschnitt** zweier Ereignisse A und B: A ∩ B
    ist die Menge aller Elementarereignisse, *die sowohl zu A als auch zu B gehören*.



Der Durchschnitt zweier Ereignisse

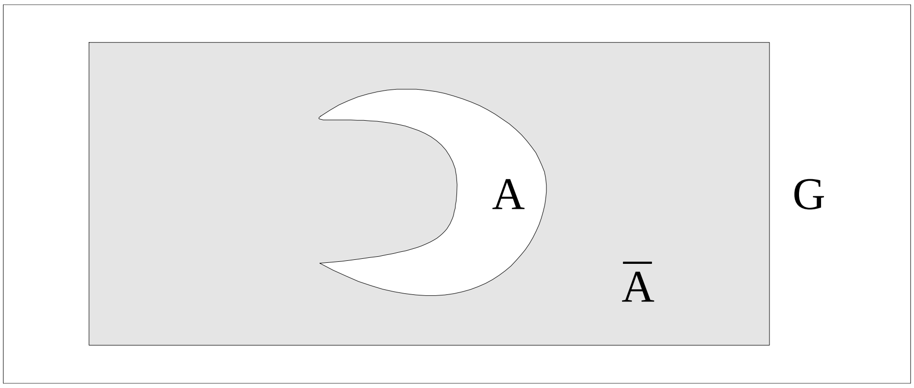
- Die Mengen A und B **schließen einander aus (A und B sind disjunkt)**, *wenn es kein Elementarereignis gibt, das zu beiden gleichzeitig gehört*:

  $A \cap B = \varnothing$ (leere Menge $\varnothing$)



Zwei disjunkte Ereignisse

- **Komplementärereignis $\overline{A}$:**
  Menge aller Elementarereignisse eines Ereignisraums G, *die nicht in A enthalten sind*.
  Es gilt: $A \cup \overline{A} = G$ und $A \cap \overline{A} = \varnothing$



Beispiele:

G = {1, 2, 3, 4, 5, 6}	Einmaliges Werfen eines Würfels
A = {1, 2, 3}	Ereignis (Augenzahl < 4)
B = {2, 4, 5, 6}	Ereignis
C = {4, 5, 6}	Ereignis (Augenzahl > 3)

Vereinigung:	$A \cup B = \{1, 2, 3, 4, 5, 6\}$
Durchschnitt:	$A \cap B = \{2\}$
Disjunkte Mengen:	A und C, da $A \cap C = \varnothing$
Komplementärereignis zu B:	$\overline{B} = \{1, 3\}$

2 Wahrscheinlichkeitsbegriffe

Die Wahrscheinlichkeit ist ein Maß zur Quantifizierung der Sicherheit bzw. Unsicherheit des Eintretens eines bestimmten Ereignisses im Rahmen eines Zufallsexperiments.

- **Klassischer Wahrscheinlichkeitsbegriff** (Pierre Simon LAPLACE (1749–1827)):
  Die Wahrscheinlichkeit für das Eintreten des Ereignisses A bei einem Zufallsexperiment ist gleich dem Verhältnis aus der Anzahl der für das Eintreten des Ereignisses günstigen Fälle und der Anzahl aller möglichen Fälle (gleichmögliche Fälle = gleichwahrscheinliche Elementarereignisse):

$$P(A)=\frac{\text{Zahl der günstigen Fälle}}{\text{Zahl der gleichmöglichen Fälle}}=\left(\frac{g}{n}\right)$$

Beispiel:

In einem Gefäß liegen sechs brauchbare und vier defekte Stücke.
Wie groß ist die Wahrscheinlichkeit, daß ein zufällig gezogenes Stück brauchbar (Ereignis A) ist?

$$P(A)=\frac{6}{10}=\frac{g}{n}=0,6$$

Voraussetzung für den klassischen Wahrscheinlichkeitsbegriff:
- endliche Anzahl der Fälle;
- nur auf Zufallsexperimente mit *gleichwahrscheinlichen Elementarereignissen* anwendbar.

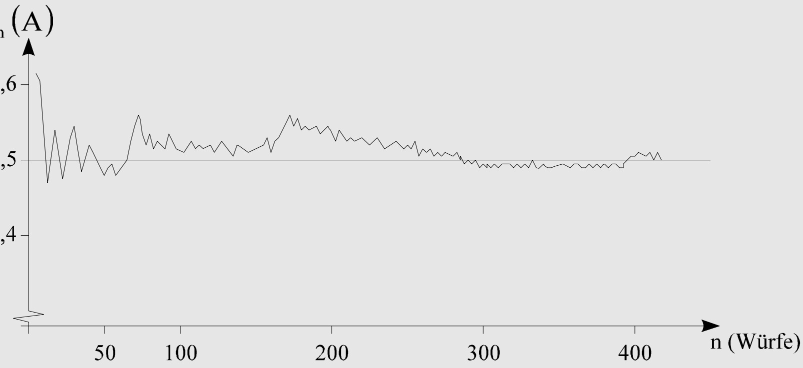
- **Statistischer Wahrscheinlichkeitsbegriff** (Richard v. MISES (1883–1953)):
  Bei einem Zufallsexperiment, das aus einer langen Folge unabhängiger Wiederholungen besteht, versteht man unter der Wahrscheinlichkeit P(A) eines Ereignisses A den **Grenzwert der relativen Häufigkeit** für das Auftreten des Ereignisses A bei unendlich häufiger Wiederholung des Zufallsexperiments:

$$P(A)=\lim_{n\to\infty}\frac{g}{n}=\lim_{n\to\infty}h_n(A)$$

$$\left(h_n(A)=\text{relative Häufigkeit von A nach n Wiederholungen}\right)$$

Beispiel:

Münzwürfe, nach jedem Wurf wird die relative Häufigkeit für „Zahl“ bis dato registriert:
A = „Zahl“



- **Subjektiver Wahrscheinlichkeitsbegriff** (Leonard J. SAVAGE):
  Wertangaben für die Wahrscheinlichkeit werden als vernünftige Glaubensaussagen interpretiert.
  $\rightarrow$ Eingang in Entscheidungsmodelle

- **Axiomatischer Wahrscheinlichkeitsbegriff** (A. N. KOLMOGOROV (geb. 1903)):
  Hier wird nicht mehr von empirischen Beobachtungen ausgegangen, sondern: Wahrscheinlichkeiten als Zuordnung von reellen Zahlen zu den Ereignissen.

  Definition der mathematischen Eigenschaften der Wahrscheinlichkeit $\rightarrow$ 3 Axiome (Axiom = mathematische Aussage, die als Grundlage einer Theorie dient und daher nicht durch diese begründbar ist):

  Gegeben sei ein System von Ereignissen R aus dem Ereignisraum G. Eine Funktion P, die jedem Ereignis A aus R eine reelle Zahl zuordnet, heißt Wahrscheinlichkeit, wenn sie folgende Eigenschaften (*Axiome*) erfüllt:

  1. P ist reell, nicht negativ: $P(A) \geq 0$, $P(A) \in \mathrm{IR}_0^+$
  2. P ist additiv: $P(A \cup B) = P(A) + P(B)$ *für* $A \cap B = \varnothing$
  3. P ist normiert: $P(G) = 1$ bzw. $0 \leq P(A) \leq 1$

  Folgerungen aus diesen Axiomen:

  - Wahrscheinlichkeit des Komplementärereignisses $\overline{A}$:
    wegen $A \cup \overline{A} = G$ und $P(A) + P(\overline{A}) = 1$
    ist $P(\overline{A}) = 1 - P(A)$
  - Wahrscheinlichkeit des unmöglichen Ereignisses:

    $P(\varnothing) = 1 - P(G) = 0$
  - Wahrscheinlichkeit für den Durchschnitt zweier sich ausschließender Ereignisse A und B:

    $P(A \cap B) = P(\varnothing) = 0$

3 Additionssatz

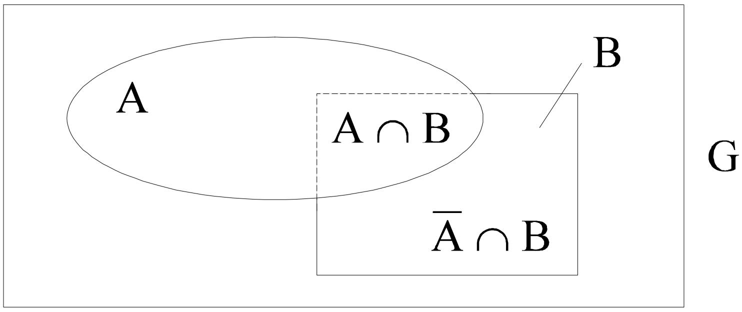
Wie hoch ist die Wahrscheinlichkeit, daß bei beliebigen, **sich nicht ausschließenden** Ereignissen eines Zufallsexperiments entweder A oder B oder A und B gemeinsam auftreten?

Also $P(A \cup B) = ?$

$A \cup B = A \cup (\overline{A} \cap B)$, wobei A und $\overline{A} \cap B$ sich ausschließen.

Aus dem 2. Axiom folgt:

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) \qquad [*]$$



Venn-Diagramm zur Ableitung des Additionssatzes

Weitere Ableitungen:

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

Daraus folgt:

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Eingesetzt in [*] ergibt sich für beliebige Ereignisse der **Additionssatz:**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (\text{A und B nicht ausschließend})$$

Schließen sich A und B gegenseitig aus, so folgt (Axiom 2):

$$P(A \cup B) = P(A) + P(B)$$

Additionssatz für 3 beliebige Ereignisse A, B, C:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

Beispiel (3 Ereignisse):

A = {1, 2, 3, 4} B = {3, 4, 5, 6} C = {2, 3, 5, 7}

$P(A) = \frac{4}{7}$ $P(B) = \frac{4}{7}$ $P(C) = \frac{4}{7}$

$A \cap B = \{3,4\}$ $A \cap C = \{2,3\}$ $B \cap C = \{3,5\}$

$P(A \cap B) = \frac{2}{7}$ $P(A \cap C) = \frac{2}{7}$ $P(B \cap C) = \frac{2}{7}$

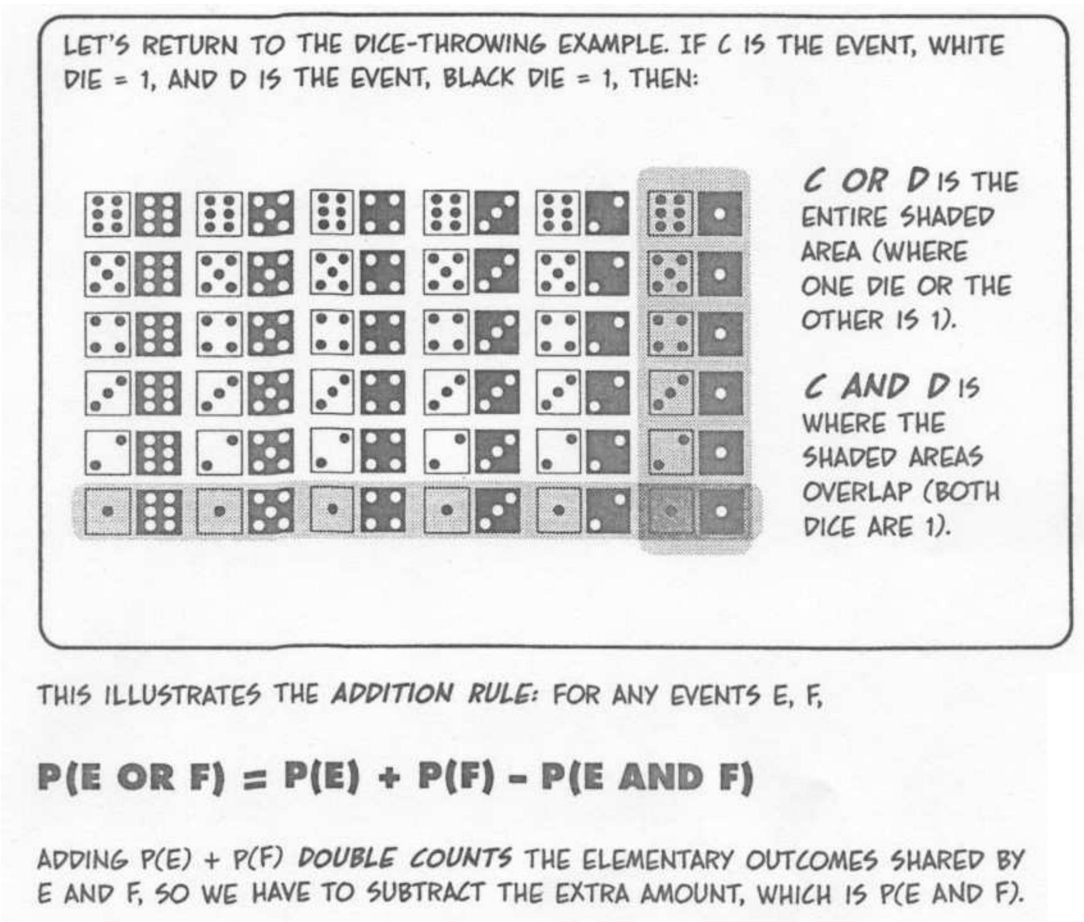
$A \cap B \cap C = \{3\}$

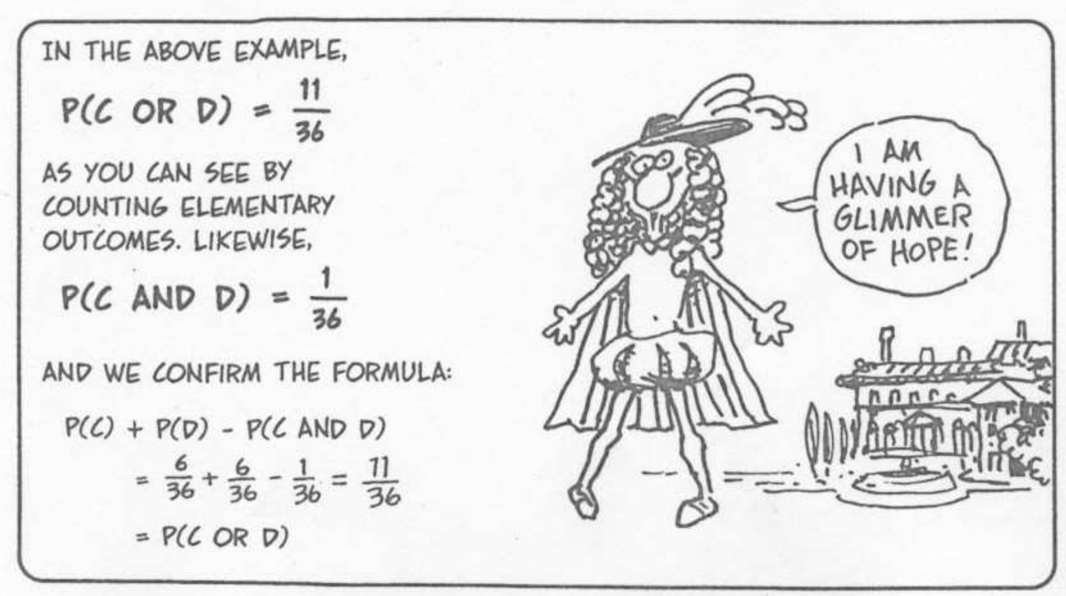
$$P(A \cap B \cap C) = \frac{1}{7}$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C) \\ &= \frac{4}{7} + \frac{4}{7} + \frac{4}{7} - \frac{2}{7} - \frac{2}{7} - \frac{2}{7} + \frac{1}{7} \\ &= 1 \end{aligned}$$

Hilfsweise:

A:	1	$\underline{2}$	~~3~~	~~4~~				$- \;\hat{=}\; A \cap B$
B:			$\cancel{3}$	4	$\cancel{5}$	6		$/ \;\hat{=}\; B \cap C$
C:		2	$\underline{3}$		5		7	$\_ \;\hat{=}\; A \cap C$
	1	2	$\underline{\underline{3}}$	4	5	6	7	$= \;\hat{=}\; A \cap B \cap C$





Gonick, Smith 1993

Beispiele (Additionssatz):

a) In einer Urne mit 200 Kugeln befinden sich 40 rote und 80 grüne Kugeln.
Wahrscheinlichkeit für das Ziehen einer roten Kugel?

$$P(\text{rot}) = \frac{40}{200} = 0{,}2$$

Wahrscheinlichkeit für das Ziehen einer grünen Kugel?

$$P(\text{grün}) = \frac{80}{200} = 0{,}4$$

Wahrscheinlichkeit für das Ziehen einer grünen oder roten Kugel?

$$P(\text{grün} \cup \text{rot}) = P(\text{grün}) + P(\text{rot}) = 0{,}4 + 0{,}2 = 0{,}6 \ ,$$

da die Ereignisse grün und rot sich ausschließen.

b) Wir betrachten nun einen Würfel:

A = {die Augenzahl < 4} = {1, 2, 3} mit P(A) = 0,5
B = {die Augenzahl ist gerade} = {2, 4, 6} mit P(B) = 0,5

Wahrscheinlichkeit für $A \cup B$?

$$A \cup B = \{\text{Augenzahl <4 oder gerade}\} = \{1, 2, 3, 4, 6\}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(2)$$
$$= \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}$$

4 Bedingte Wahrscheinlichkeit, Multiplikationssatz

Bei Eintreten von Ereignissen in Abhängigkeit von bestimmten anderen Ereignissen (sich nicht ausschließende Ereignisse):

Bedingte Wahrscheinlichkeit:

Wahrscheinlichkeit von B unter der Voraussetzung, daß Ereignis A vorher eingetreten ist, bzw. gleichzeitig mit B eintritt:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

(Also bezogen auf die Zahl der für die *Bedingung günstigen Elementarereignisse*)

Beispiel:

A = Frau B = Erwerbstätig
Wahrscheinlichkeit, aus der Menge der Frauen eine erwerbstätige Frau auszuwählen:

$$P(Erwerbstätig|Frau) = \frac{P(Frau\ und\ Erwerbstätig)}{P(Frau)}$$

Stochastische Unabhängigkeit liegt dann vor, wenn gilt

$$P(B|A) = P(B|\bar{A}) = P(B)$$

Das Eintreten des Ereignisses B hängt nicht vom Eintritt des Ereignisses A ab.

Aus der Definition der bedingten Wahrscheinlichkeit folgt der

Multiplikationssatz (Auflösen nach $P(A \cap B)$):

$$\begin{aligned} P(A \cap B) &= P(B) \cdot P(A|B) \\ &= P(A) \cdot P(B|A) \end{aligned}$$

Für **stochastisch unabhängige Ereignisse** gilt

$$P(A \cap B) = P(A) \cdot P(B)$$

Multiplikationssatz für drei Ereignisse A_1, A_2, A_3 bei **stochastischer Abhängigkeit** (allgemein):

$$\mathrm{P}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3) = \mathrm{P}(\mathrm{A}_1) \cdot \mathrm{P}(\mathrm{A}_2 | \mathrm{A}_1) \cdot \mathrm{P}(\mathrm{A}_3 | \mathrm{A}_1 \cap \mathrm{A}_2)$$

bei **stochastischer Unabhängigkeit:**

$$\mathrm{P}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3) = \mathrm{P}(\mathrm{A}_1) \cdot \mathrm{P}(\mathrm{A}_2) \cdot \mathrm{P}(\mathrm{A}_3)$$

Beispiele:

- Werfen von zwei idealen Würfeln:
  Wie hoch ist die Wahrscheinlichkeit, mit dem ersten Würfel eine Zwei und mit dem zweiten Würfel eine Sechs zu werfen?

Das Werfen beider Würfel ist stochastisch unabhängig:

$$P(\{2\}\cap\{6\})=P(2)\cdot P(6)$$

$$=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}$$

- Eine Münze wird mehrmals hintereinander geworfen:
  Wie groß ist die Wahrscheinlichkeit, dreimal Zahl hintereinander zu erhalten?

$$\mathrm{P}(\mathrm{Z}_1\cap\mathrm{Z}_2\cap\mathrm{Z}_3)=\mathrm{P}(\mathrm{Z}_1)\cdot\mathrm{P}(\mathrm{Z}_2)\cdot\mathrm{P}(\mathrm{Z}_3)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$$

- Ein Los von 50 Bauteilen hat 20 % Ausschuß. Bei der Abnahmeprüfung werden drei Bauteile nacheinander *ohne* Zurücklegen ausgewählt. Die Annahme erfolgt nur dann, wenn alle drei Bauteile einwandfrei sind.
  Wie groß ist die Annahmewahrscheinlichkeit?

  A_i ist das Ereignis "einwandfreies Bauteil im i-ten Zug".

  Es liegt stochastische Abhängigkeit vor, da die Auswahlmöglichkeiten von den Vorereignissen abhängen (ohne Zurücklegen!).

1. Zug: $\mathrm{P}(\mathrm{A}_1)=\frac{40}{50}=\frac{\text{günstige Fälle}\,(=\textit{einwandfreie Fälle})}{\text{gleichmögliche Fälle}}$

2. Zug: $P(A_2\,|\,A_1)=\frac{39}{49}$ Ein einwandfreies Bauteil ist schon gezogen worden.

3. Zug: $P(A_3\,|\,A_1\cap A_2)=\frac{38}{48}$ Zwei einwandfreie Bauteile sind schon gezogen worden.

Multiplikationssatz bei Abhängigkeit:

$$P(A_1\cap A_2\cap A_3)=P(A_1)\cdot P(A_2\,|\,A_1)\cdot P(A_3\,|\,A_1\cap A_2)$$

$$=\frac{40}{50}\cdot\frac{39}{49}\cdot\frac{38}{48}$$

$$=0{,}5041$$

Antwort: Die Wahrscheinlichkeit für drei einwandfreie Bauteile beträgt 50,41 %.

Zusammenfassung:

Additionssatz:

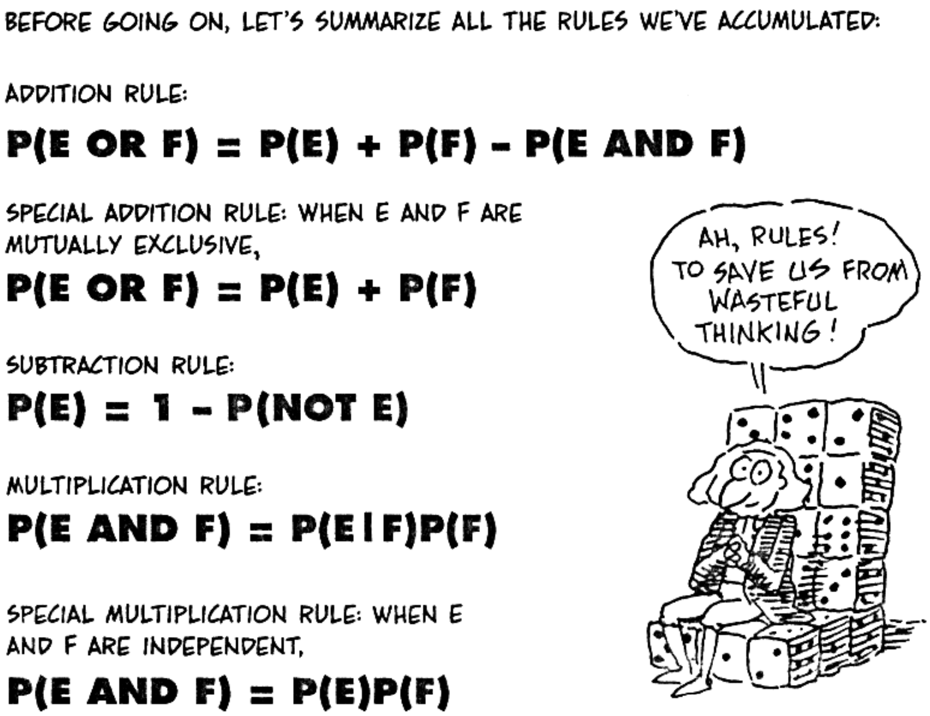
bei Verknüpfung der Ereignisse durch **„Vereinigung“, logisches „oder“ (∪)**
(„das Eine“ oder „das Andere“ oder „Sowohl als auch“)

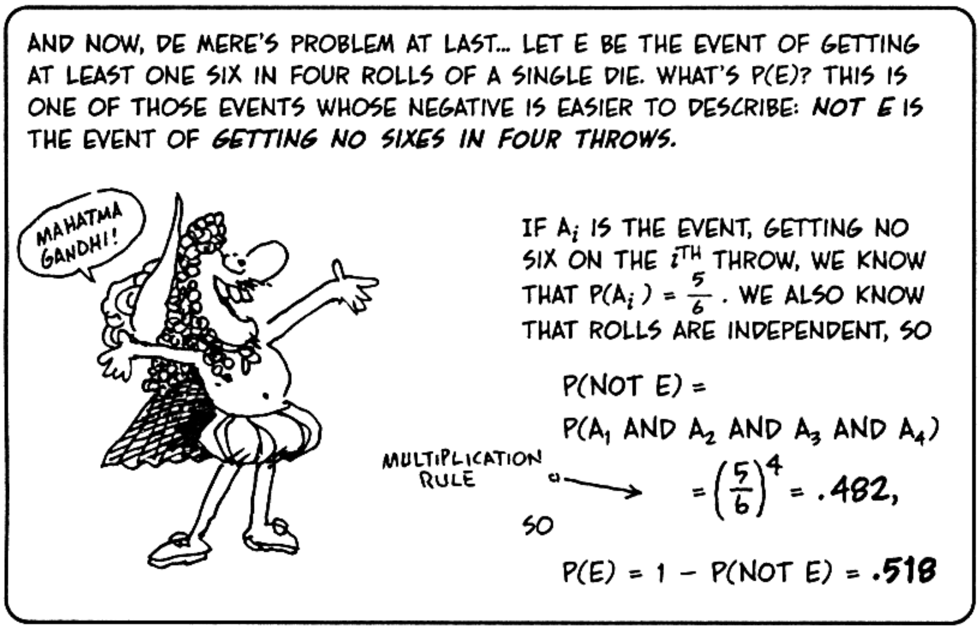
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

Multiplikationssatz:

bei Verknüpfung der Ereignisse durch „**Durchschnitt“, logisches „und“ (∩)**
(nur „Sowohl als auch“)

$P(A \cap B) = P(A) \cdot P(B | A)$





Gonick, Smith 1993

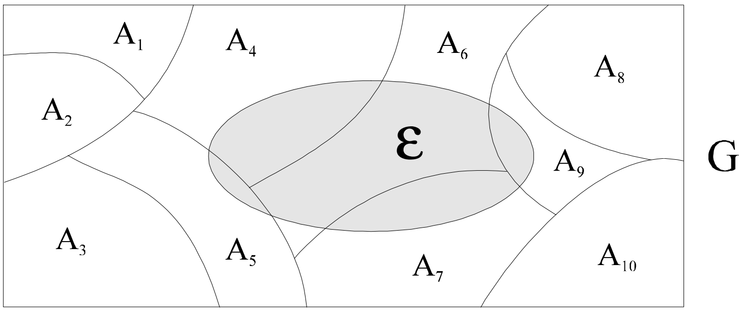
5 Theorem der totalen Wahrscheinlichkeit, Satz von Bayes

$A_1, A_2, A_3, \ldots$ seien sich gegenseitig ausschließende Ereignisse,

d.h. $\quad A_i \cap A_j = \varnothing \quad$ für $i \neq j$,

außerdem sei $A_1, A_2, A_3, \ldots$ eine Zerlegung von G,

d.h. $\quad A_1 \cup A_2 \cup A_3 \cup \ldots = G$



Einteilung des Ereignisraums

Jedes beliebige Ereignis ε läßt sich wie folgt darstellen:

$$\varepsilon = (\varepsilon \cap A_1) \cup (\varepsilon \cap A_2) \cup \ldots$$

Nach dem Additionssatz für sich gegenseitig ausschließende Ereignisse gilt:

$$P(\varepsilon) = P(\varepsilon \cap A_1) + P(\varepsilon \cap A_2) + \ldots = \sum_{i=1}^{n} P(\varepsilon \cap A_i) \qquad (*)$$

Nach dem Multiplikationssatz für sich nicht ausschließende Ereignisse gilt:

$$P(\varepsilon \cap A_i) = P(\varepsilon | A_i) \cdot P(A_i) \qquad (\text{ein Term in } (*))$$

Setzt man dies in die obige Summe (*) ein, dann ergibt sich der

Satz von der totalen Wahrscheinlichkeit:

$$P(\varepsilon) = \sum_{i=1}^{n} P(\varepsilon | A_i) \cdot P(A_i)$$

Beispiel:

In einem Betrieb werden täglich 1000 Stück eines Produktes hergestellt. Davon liefert Maschine

M_1: 100 Stück mit 5 % Ausschußanteil,
M_2: 400 Stück mit 4 % Ausschußanteil und
M_3: 500 Stück mit 2 % Ausschußanteil.

Aus einer Tagesproduktion wird ein Stück zufällig ausgewählt und überprüft.

A_i = Ereignis, daß dieses Stück auf M_i hergestellt ist;
ε = Ereignis, daß dieses Stück fehlerhaft ist.

Wie groß ist die Wahrscheinlichkeit, daß ein ausgewähltes Stück fehlerhaft ist?

$$P(\varepsilon) = P(A_1)\cdot P(\varepsilon \mid A_1) + P(A_2)\cdot P(\varepsilon \mid A_2) + P(A_3)\cdot P(\varepsilon \mid A_3)$$
$$= 0{,}1\cdot 0{,}05 + 0{,}4\cdot 0{,}04 + 0{,}5\cdot 0{,}02$$
$$= 0{,}031$$

Die Frage nun, wie groß die Wahrscheinlichkeit ist, daß ein fehlerhaftes Stück auf der Maschine M_1 gefertigt wurde, wird uns zum sogenannten Theorem von Bayes führen.

Hier ist die bedingte Wahrscheinlichkeit $P(A_j \mid \varepsilon)$ gefordert (hier: $P(A_1 \mid \varepsilon)$). Man beachte, daß dies ein Unterschied zu der bedingten Wahrscheinlichkeit $P(\varepsilon \mid A_j)$ ist!

Es gilt: $P(A_j \mid \varepsilon) = \dfrac{P(A_j \cap \varepsilon)}{P(\varepsilon)}$

Mit Hilfe des Multiplikationssatzes (wegen P (. $\cap$.)) erhält man:

$$P(A_j \mid \varepsilon) = \frac{P(A_j)\cdot P(\varepsilon \mid A_j)}{P(\varepsilon)}$$

Wird für P(ε) das Theorem der totalen Wahrscheinlichkeit eingesetzt, so ergibt sich der

Satz von Bayes:

$$P(A_j \mid \varepsilon) = \frac{P(A_j)\cdot P(\varepsilon \mid A_j)}{\sum_{i=1}^{n} P(A_i)\cdot P(\varepsilon \mid A_i)}$$

Beispiel:

Wie groß ist die Wahrscheinlichkeit, daß ein fehlerhaftes Stück auf M_1 produziert wurde?

$$P(A_1 \mid \varepsilon) = \frac{P(A_1)\cdot P(\varepsilon \mid A_1)}{\sum_{i=1}^{3} P(A_i)\cdot P(\varepsilon \mid A_i)}$$

$$= \frac{0{,}1\cdot 0{,}05}{0{,}1\cdot 0{,}05 + 0{,}4\cdot 0{,}04 + 0{,}5\cdot 0{,}02} = 0{,}1613$$

$P(A_j)$ = (unbedingte) Wahrscheinlichkeit für das Eintreten des Ereignisses A_j;

$P(A_j \mid \varepsilon)$ = bedingte Wahrscheinlichkeit für das Eintreten des Ereignisses A_j unter der Bedingung, daß das Ereignis ε eingetreten ist.

Zur Anwendung des Satzes von Bayes:

Das Theorem von Bayes wird bei Entscheidungen unter Unsicherheit bzw. in Risikosituationen angewendet. Dabei handelt es sich um Situationen der folgenden Form:

Es gibt mehrere sich gegenseitig ausschließende Zustände (Alternativen) A_1, A_2,..., A_n und Schätzungen der Wahrscheinlichkeiten $P(A_j)$ für das Eintreten der Zustände bzw. Alternativen.

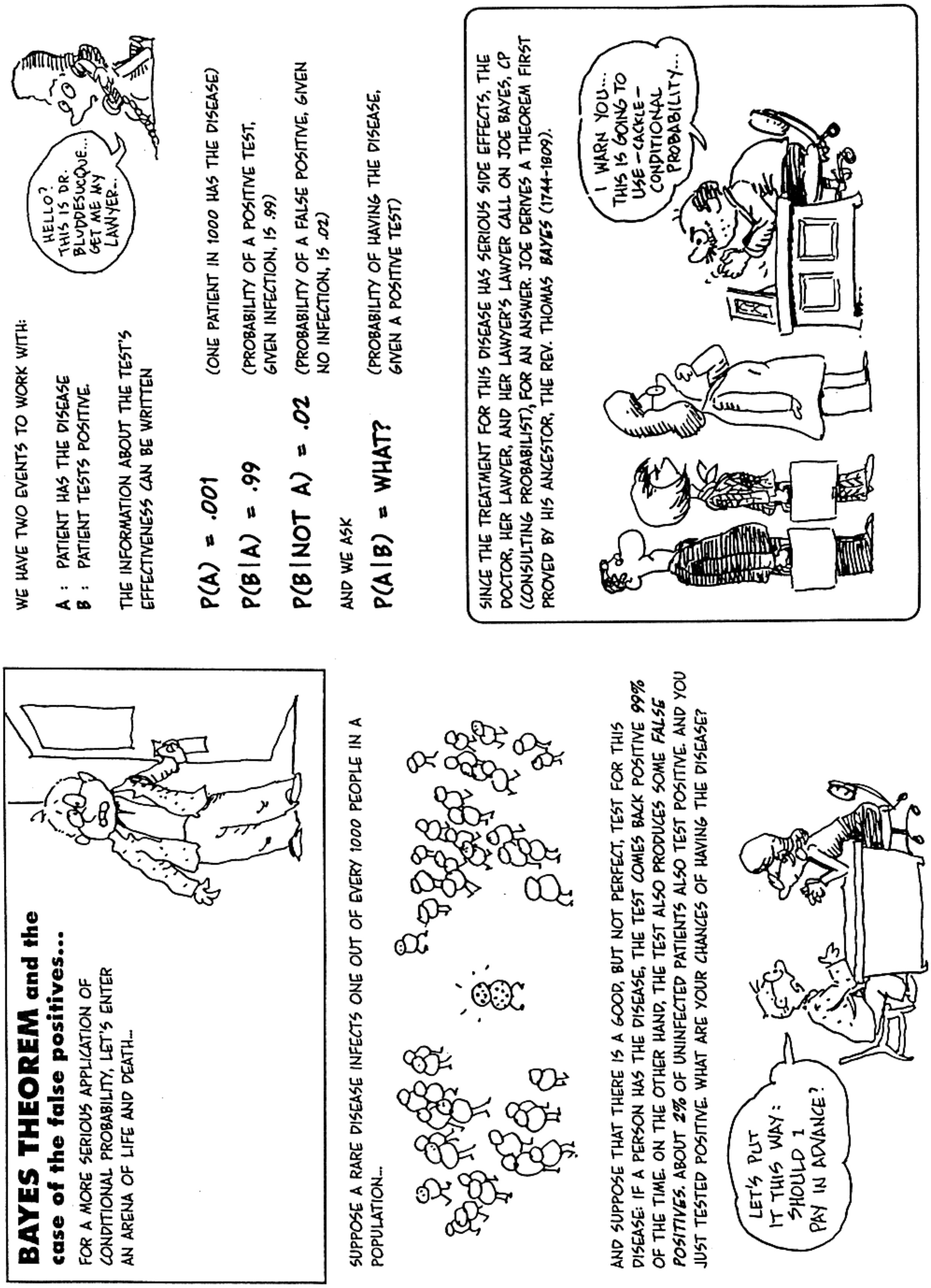
$P(A_j)$ heißen **a priori Wahrscheinlichkeiten**.

Über Realisationen von Zufallsexperimenten ermittelt man dann die Wahrscheinlichkeiten $P(\varepsilon | A_i)$.

Für das Eintreten des Ereignisses ε (oder des Komplements $\overline{\varepsilon}$) berechnet man unter Verwendung des Theorems von Bayes die bedingten Wahrscheinlichkeiten $P(A_j | \varepsilon)$.

$P(A_j | \varepsilon)$ heißen **a posteriori Wahrscheinlichkeiten**.

Die a posteriori Wahrscheinlichkeiten werden als Verbesserungen der a priori Wahrscheinlichkeiten interpretiert, weil Zusatzinformationen verwendet werden.



Gonick, Smith 1993

JOE BEGINS WITH A 2X2 TABLE, WHICH DIVIDES THE SAMPLE SPACE INTO FOUR MUTUALLY EXCLUSIVE EVENTS. IT DISPLAYS EVERY POSSIBLE COMBINATION OF DISEASE STATE AND TEST RESULT.

	A	NOT A
B	A AND B	NOT A AND B
NOT B	A AND NOT B	NOT A AND NOT B

LET'S FIND THE PROBABILITIES OF EACH EVENT IN THE TABLE:

	A	NOT A	SUM
B	P(A AND B)	P(NOT A AND B)	P(B)
NOT B	P(A AND NOT B)	P(NOT A AND NOT B)	P(NOT B)
	P(A)	P(NOT A)	1

THE PROBABILITIES IN THE MARGINS ARE FOUND BY SUMMING ACROSS ROWS AND DOWN COLUMNS.

NOW COMPUTE:

P(A AND B) = P(B|A)P(A) = (.99)(.001) = .00099

P(NOT A AND B) = P(B|NOT A)P(NOT A) = (.02)(.999) = .01998

ALLOWING US TO FILL IN SOME ENTRIES:

	A	NOT A	SUM
B	.00099	.01998	.02097
NOT B	P(A AND NOT B)	P(NOT A AND NOT B)	P(NOT B)
	.001	.999	1

WE FIND THE REMAINING PROBABILITIES BY SUBTRACTING IN THE COLUMNS, THEN ADDING ACROSS THE ROWS.

THE FINAL TABLE IS:

	A	NOT A		
B	.00099	.01998	.02097	P(B)
NOT B	.00001	.97902	.97903	P(NOT B)
	.001	.999	1	
	P(A)	P(NOT A)		

FROM WHICH WE DIRECTLY DERIVE

$$P(A|B) = \frac{P(A \text{ AND } B)}{P(B)} = \frac{.00099}{.02097} = .0472$$

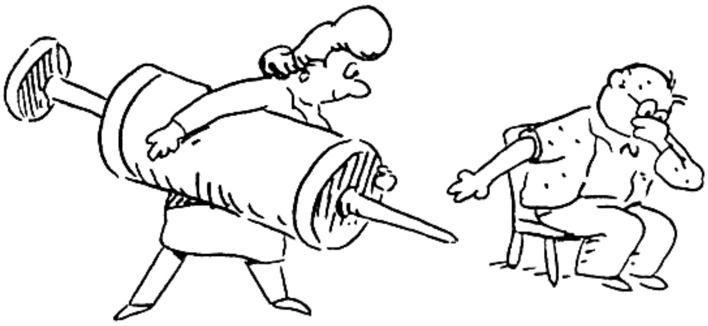
DESPITE THE HIGH ACCURACY OF THE TEST, ***LESS THAN 5%*** OF THOSE WHO TEST POSITIVE ACTUALLY HAVE THE DISEASE! THIS IS CALLED THE ***FALSE POSITIVE PARADOX.***

THIS TABLE SHOWS WHAT HAPPENS IN A GROUP OF A THOUSAND PATIENTS. ON AVERAGE, ONLY 21 PEOPLE WILL TEST POSITIVE—AND ONLY ***ONE*** OF THEM HAS THE DISEASE! 20 FALSE POSITIVES COME FROM THE ***MUCH LARGER UNINFECTED GROUP.***

	DISEASE	NO DISEASE	
TESTS POSITIVE	1	20	21
TESTS NEGATIVE	0	979	979
	1	999	1000

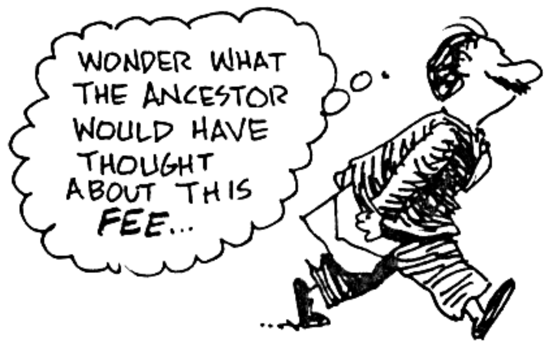
Gonick, Smith 1993

WHAT'S THE PHYSICIAN TO DO? JOE BAYES ADVISES HER NOT TO START TREATMENT ON THE BASIS OF THIS TEST ALONE. THE TEST DOES PROVIDE INFORMATION, HOWEVER: WITH A POSITIVE TEST THE PATIENT'S CHANCE OF HAVING THE DISEASE INCREASED FROM 1 IN 1000 TO 1 IN 21. THE DOCTOR FOLLOWS UP WITH MORE TESTS.



JOE BAYES COLLECTS HIS CONSULTING CHECK BEFORE ADMITTING THAT ALL THOSE STEPS HE WENT THROUGH CAN BE COMPRESSED INTO THE SINGLE FORMULA CALLED BAYES THEOREM:

$$P(A|B) = \frac{P(A)P(B)}{P(A)P(B)+P(\text{NOT } A)P(B|\text{NOT } A)}$$



IT COMPUTES P(A|B) FROM P(A) AND THE TWO CONDITIONAL PROBABILITIES P(B|A) AND P(B|NOT A). YOU CAN DERIVE IT BY NOTING THAT THE BIG FRACTION CAN BE EXPRESSED AS

$$\frac{P(A \text{ and } B)}{P(A \text{ and } B)+P(\text{NOT } A \text{ and } B)} = \frac{P(A \text{ and } B)}{P(B)} = P(A|B)$$

Gonick, Smith 1993

Vergleich zum Beispiel:

$$P\left(A_1 \mid \varepsilon\right) = P\left(\text{fehlerhaftes Stück auf } \mathrm{M}_1 \text{ produziert}\right)$$

6 Kombinatorik

Zur Berechnung aller jeweils möglichen und/oder günstigen Fälle der Wahrscheinlichkeitsrechnung werden die Ergebnisse der Kombinatorik herangezogen.[1]

KOMBINATORIK	Wiederholung (der gleichen Elemente)	
	ohne	mit
Permutation (Anordnungsmöglichkeiten)	verschiedene Elemente $P(n)=n!$	gruppenweise identische Elemente $\bar{P}(n_1,\ldots,n_K;\,n)=\dfrac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_K!}$
Variation (Auswahlmöglichkeiten, Reihenfolge wichtig)	$V(m,n)=\dfrac{n!}{(n-m)!}$	$\bar{V}(m,n)=n^m$
Kombination (Auswahlmöglichkeiten, Reihenfolge unwichtig)	$C(m,n)=\dbinom{n}{m}$	$\bar{C}(m,n)=\dbinom{n+m-1}{m}$

$\dbinom{n}{m}=\dfrac{n!}{m!(n-m)!}$; $\left(0!=1,\dbinom{n}{0}=1\right)$; n = Anzahl aller Elemente, m = Auswahlanzahl

Beispiele:

Permutation:

Alle möglichen Anordnungen, Reihenfolgen.

a) Fünf Pferde können in $P(5)=5!=1\cdot 2\cdot 3\cdot 4\cdot 5=120$ verschiedenen Anordnungen (Reihenfolgen) in das Ziel einlaufen (Permutation ohne Wiederholung).

b) Anzahl der Permutationen aus den Elementen a, b, c? (ohne Wiederholung)

abc	bac	cab
acb	bca	cba

P(3) = 3! = 1 · 2 · 3 = 6

Anzahl der Permutationen aus den Elementen a, a, c? (mit Wiederholung)

aac	~~aac~~	caa
aca	~~aca~~	~~caa~~
$n_1 = 2$	$n_2 = 1$	$n = 3$

[1] Vgl. z.B. Merz, J. 1992, Mathematik I für Wirtschaftswissenschaftler – Analysis, Skriptum zur Vorlesung, Frankfurt am Main, S. 40–45.

$$\bar{P}(2,1;3)=\frac{3!}{2!\cdot 1!}=\frac{\cancel{1}\cdot\cancel{2}\cdot 3}{\cancel{1}\cdot\cancel{2}\cdot 1}=3$$

c) Wieviele verschiedene zehnstellige Zahlen lassen sich aus den Ziffern der Zahl 7 844 673 727 bilden?

$n=10 \qquad n_1(7)=4 \qquad n_2(4)=2 \qquad n_3(8)=n_4(6)=n_5(3)=n_6(2)=1$

$$\bar{P}(4,2,1,1,1,1;10)=\frac{10!}{4!\cdot 2!}=\frac{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot 5\cdot 6\cdot 7\cdot 8\cdot 9\cdot 10}{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot\cancel{4}\cdot 1\cdot 2}=75600$$

Variation:

Die Auswahl von m aus n Elementen, wobei die Reihenfolge wichtig ist.

a) Nicole entschließt sich, vor Verlassen der UNI-DISCO noch mit drei von ihren fünf Freunden zu tanzen (immer mit einem anderen).
Wieviele Tanzmöglichkeiten hat Nicole, wenn die Reihenfolge nicht egal ist?

n = 5, m = 3

$$V(3,5)=\frac{5!}{(5-3)!}=\frac{5!}{2!}=3\cdot 4\cdot 5=60 \text{ Möglichkeiten}$$

→ Wie groß ist die Wahrscheinlichkeit für eine Auswahl (1 Tanzset), z. B., daß Nicole erst mit Freund 4, dann mit Freund 2 und schließlich mit Freund 3 tanzt?

$$P('4,2,3')=\frac{1}{60}=0,0167=1,67\,\%$$

b) Nicole ist ausnahmsweise bereit, auch mehrmals mit einem ihrer Freunde zu tanzen. Auswahl mit Wiederholung:

$$\bar{V}(3,5)=5^3=125 \textit{ Möglichkeiten}$$

c) Wieviele vierstellige Zahlen sind im Dualsystem bzw. 5er-System darstellbar?

$m=4$ bits

$n=2$ (duales System) $\qquad \bar{V}(4,2)=2^4=16 \qquad \Rightarrow 0,\ldots,15$

$n=5$ (Penta-System) $\qquad \bar{V}=(4,5)=5^4=625 \qquad \Rightarrow 0,\ldots,624$

Kombination:

Die Auswahl von m aus n Elementen, wobei die Reihenfolge uninteressant ist.

a) Ein Skatspiel hat 32 Karten. Ein Spieler erhält zehn Karten.
Wieviele voneinander verschiedene Zusammenstellungen von je zehn Karten gibt es?

Die Reihenfolge ist unwichtig, keine Wiederholung ⇒ Kombination

$$C(10,32)=\binom{32}{10}=\frac{32!}{10!\cdot(32-10)!}=64512240$$

b) Studentinnen und Studenten aus acht Fachbereichen sollen eine Delegation von fünf Studenten bilden.
Wieviele Delegationsmöglichkeiten gibt es?

Auswahl von 5 aus 8 $\Rightarrow$ n = 8 m = 5
mit Wiederholung

$$\bar{C}(5,8)=\binom{8+5-1}{5}=\binom{12}{5}=\frac{12!}{5!\cdot(12-5)!}=\frac{8\cdot 9\cdot 10\cdot 11\cdot 12}{1\cdot 2\cdot 3\cdot 4\cdot 5}=792$$

Mit der Kombinatorik (Permutation, Variation, Kombination) lassen sich also alle günstigen und gleichmöglichen Fälle berechnen $\rightarrow$ Grundlage der Wahrscheinlichkeitsrechnung mit

$$P(\ .\ )=\frac{\text{günstige Fälle}}{\text{gleichmögliche Fälle}}$$

7 Übungsaufgaben Wahrscheinlichkeitsrechnung

Für den „Prix de l´Arc de Triomphe“, dem großen Pferderennen in Longchamps bei Paris, sind die Pferde Beaujeux (B), Menes (M), Ilix (I) und Nereide (N) genannt. Es sollen Sieger und Zweitplazierte vorausgesagt werden.

a) Wie groß ist die Wahrscheinlichkeit, daß
   1. Beaujeux Sieger wird?
   2. der Einlauf (für Zweierwette) Nereide, Ilix sein wird?

1. $P(\text{'}\,Beaujeux\text{ wird Sieger'})=\text{Eins von Vieren}=\frac{1}{4}=0,25$

   bzw. Anordnungen, Reihenfolgen, Permutationen von {B, I, M, N}

$$\begin{array}{cccccccccccccccc} & \textit{Plätze} & & & & \textit{Plätze} & & & & \textit{Plätze} & & & & \textit{Plätze} & & \\ 1. & 2. & 3. & 4. & 1. & 2. & 3. & 4. & 1. & 2. & 3. & 4. & 1. & 2. & 3. & 4. \end{array}$$

$$G=\left\{\begin{array}{cccccccccccccccc} B & I & M & N & I & B & M & N & M & B & I & N & N & B & I & M \\ B & I & N & M & I & B & N & M & M & B & N & I & N & B & M & I \\ B & M & I & N & I & M & B & N & M & I & B & N & N & I & B & M \\ B & M & N & I & I & M & N & B & M & I & N & B & N & I & M & B \\ B & N & I & M & I & N & B & M & M & N & B & I & N & M & B & I \\ B & N & M & I & I & N & M & B & M & N & I & B & N & M & I & B \end{array}\right\}$$

$\rightarrow$ Permutation ohne Wiederholung

Anzahl = n! = 4! = 1 · 2 · 3 · 4 = 24

$$P(\text{'}\,B\text{ ist Erster'})=\frac{\textit{günstige}\text{ Fälle}}{\text{gleichmögliche Fälle}}=\frac{6}{24}=\frac{1}{4}=0,25$$

2. P('Reihenfolge N I'), 'sowohl N als auch I'
   $\rightarrow$ Multiplikationssatz, stochastische Abhängigkeit ('ohne' Zurücklegen)

$$P(N \cap I) = P(I) \cdot P(N \mid I)$$

$$\left.\begin{aligned} P(I) &= \frac{1}{4} \\ P(N \mid I) &= \frac{1}{3} \end{aligned}\right\} \Rightarrow P(N \cap I) = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$$

Alternative Berechnungsweise:

{B, I, M, N}

$$\textit{Gleichmögliche}\text{ Fälle: } \mathrm{G} = \begin{Bmatrix} BI & BM & BN \\ IB & IM & \mathbf{IN} \\ MB & MI & MN \\ NB & NI & NM \end{Bmatrix}$$

Anzahl der Fälle = 12

Auswahl m = 2 aus n = 4 ohne Wiederholung, Reihenfolge wichtig

Variation:

$$V(m,n) = \frac{n!}{(n-m)!} = \frac{4!}{(4-2)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2} = 12$$

$$P('NI') = \frac{\textit{günstige}\text{ Fälle}}{\text{gleichmögliche Fälle}} = \frac{\mathbf{1}}{12} = 8{,}3\%$$

Die Wahrscheinlichkeit, daß sich ein Einlauf Nereide vor Ilix ergibt, beträgt 8,3 %.

b) Die vier Pferde treffen am nächsten Tag erneut aufeinander.
Wie groß ist die Wahrscheinlichkeit, daß an beiden Tagen

1. B der Sieger ist?
2. die Reihenfolge B, N gilt?
3. die Reihenfolge B, M, N, I gilt?

1. P('Beaujeux ist der Sieger an beiden Tagen') = ?

   Logisches UND: $\cap \Rightarrow$ Multiplikationssatz, stochastisch unabhängig

   $$P(\text{'}\textit{Beaujeux}\text{ ist der Sieger an beiden Tagen'})$$
   $$= P(\text{'B ist der Sieger am Tag 1'}) \cdot P(\text{'B ist der Sieger am Tag 2'})$$
   $$= \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$

Alternative Berechnungsweise:

Alle möglichen Fälle: Auswahl m = 2 aus n = 4 mit Wiederholung
Reihenfolge wichtig $\Rightarrow \overline{V}(2,4)$

	Tage							
	1.	2.	1.	2.	1.	2.	1.	2.
	B	B	I	B	M	B	N	B
Elementarereignisse	B	I	I	I	M	I	N	I
	B	M	I	M	M	M	N	M
	B	N	I	N	M	N	N	N

$$\overline{V}(2,4)=4^2=16 \quad \Rightarrow P(...)=\frac{1}{\overline{V}(2,4)}=\frac{1}{16}$$

2. P('B, N an beiden Tagen') = ? (⇒ Multiplikationssatz)

$$P('B,\text{ N an beiden Tagen'})$$
$$=P(\text{'B, N am Tag 1'})\cdot P(\text{'B, N am Tag 2'})$$
$$=\frac{1}{12}\cdot\frac{1}{12}=\frac{1}{144}=0{,}0069$$

Alternative Berechnungsweise:

Alle möglichen Fälle: Auswahl ... (komplex) zusammengesetzt

Elementarereignisse: Tag 1: V(2 , 4) = 12 (siehe vorne); jedes Ereignis, das am ersten Tag möglich war, kann auch am zweiten Tag eintreten:

$$\Rightarrow 12\cdot 12=144 \quad \Rightarrow P(...)=\frac{1}{144}$$

3. P('B, M, N, I an beiden Tagen') = ?

'Durchschnitt', logisches UND: ∩ ⇒ Multiplikationssatz, stochastisch unabhängig

$$P(...)=P(\text{'B, M, N, I erster Tag'})\cdot P(\text{'B, M, N, I zweiter Tag'})$$

P('B, M, N, I'): Permutation ohne Wiederholung ⇒ n!
n! = 4! = 1·2·3·4 = 24

$$P('B,\text{ M, N, I'})=\frac{1}{24} \qquad (\textit{siehe}\text{ vorne})$$

$$\Rightarrow P(...)=\frac{1}{24}\cdot\frac{1}{24}=\frac{1}{576}=0{,}0017$$

Alternative Berechnungsweisen:

komplex zusammengesetztes Ereignis, es ist relativ schwierig (da umfangreich), alle Möglichkeiten darzustellen.

Schlußfolgerungen:

- Manchmal ist es sehr komplex, alle(!) Möglichkeiten zusammenzustellen und dann die für die LAPLACE-Wahrscheinlichkeit günstigen Möglichkeiten auszuwählen.
- Deshalb: Additionssatz, Multiplikationssatz anwenden; Die dafür notwendigen Teilwahrscheinlichkeiten können vereinfacht über die Kombinatorik berechnet werden.

Keyconcepts

Zufallsexperiment
Ereignis
Wahrscheinlichkeitsbegriff
Kolmogorov'sche Axiome
Bedingte/unbedingte Wahrscheinlichkeit
Additionssatz
Multiplikationssatz
Satz von Bayes
A priori Wahrscheinlichkeit
A posteriori Wahrscheinlichkeit
Satz der totalen Wahrscheinlichkeit

II Zufallsvariablen und Wahrscheinlichkeitsverteilungen

Beschreibung der möglichen Ergebnisse eines zufälligen Prozesses, Berechnung der möglichen Ausgänge und der dazugehörigen Wahrscheinlichkeiten

1 Zufallsvariablen

Untersucht man bei einem Zufallsexperiment *'zweimaliges Werfen einer Münze'* die Frage, wie oft *'Wappen'* erscheint, so sind hier die möglichen Werte 0, 1 oder 2.

Variablen, wie die 'Anzahl Wappen' bei einem mehrmaligen Münzwurf oder das Ziehen von roten und schwarzen Kugeln bei einem Urnenexperiment (mit/ohne Zurücklegen), **deren Werte vom Zufall abhängen**, nennt man **Zufallsvariablen**.

Die beobachteten Ausprägungen bestimmter Merkmale der Wirtschaft und Gesellschaft sind, wenn nur eine zufällige Auswahl von Merkmalsträgern befragt werden, das Ergebnis eines zufälligen Prozesses und damit Zufallsvariablen.

Zufallsvariable:	X, Y, Z,... (große Buchstaben) 'random variables'
Realisation, Ausprägung davon:	x, y, z,... (kleine Buchstaben)
Prob(X = x) = P(X = x):	Wahrscheinlichkeit, daß X die Ausprägung x annimmt.

Beispiel:

Zweimaliges Werfen einer Münze:
Zufallsvariable: X ('Anzahl Zahl')
Ausprägungen: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$.

Definitionsbereich einer Zufallsvariablen = Ereignisraum G;
Wertebereich = Menge der reellen Zahlen (im allgemeinen);

Diskrete Zufallsvariable:

im Wertebereich liegen nur endlich oder abzählbar unendlich viele Werte x_1, x_2, ... (zählbar) (z. B. Zahl der 'faulen' Kunden, Anzahl der Lottogewinner);

Stetige Zufallsvariable:

im Wertebereich liegt jeder beliebige Zahlenwert eines Intervalls (kontinuierliche Zufallsvariable, nicht zählbar)
(z. B. Temperatur, Bearbeitungszeit)

2 Wahrscheinlichkeitsfunktion und Verteilungsfunktion diskreter Zufallsvariablen

Der Prozeß, der die Ausprägungen der Zufallsvariablen generiert (der sogenannte „datengenerierende Prozeß), kann durch seine Wahrscheinlichkeitsfunktion beschrieben werden:

Die **Wahrscheinlichkeitsfunktion ('probability function') f(X)** listet alle Ergebnisse x_i (i = 1,...), die die Zufallsvariable X annimmt, mit ihren Eintrittswahrscheinlichkeiten auf.

$f(x_i) = Prob(X = x_i) = P(X = x_i)$ (Wahrscheinlichkeit, mit der X die Ausprägung x_i annimmt)

Eigenschaften: 1. $0 \le f(x_i) \le 1;$

2. $\sum_i f(x_i) = 1;$

Beispiel:

Dreimaliges Werfen einer idealen Münze.
Der Ereignisraum G dieses Zufallsexperiments besteht aus acht Ereignissen:
$\left(\bar{V}(m,n) = n^m \Rightarrow \bar{V}(3,2) = 2^3 = 8\right)$ mit

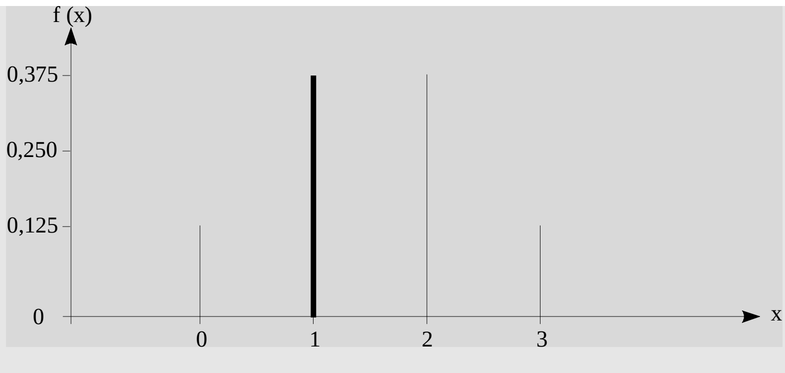
$G = \{(Z,Z,Z);(Z,Z,W);(Z,W,Z);(W,Z,Z);(Z,W,W);(W,Z,W);(W,W,Z);(W,W,W)\}$.

Zufallsvariable X sei 'Anzahl der Wappen' (diskret):
Auszählen ⇒ Zuordnung x_i aus dem Wertebereich $R_{x_i} = \{0,1,2,3\}$.

G		$x_i \in R_{x_i}$		$f(x_i)$
(Z, Z, Z)	→	$x_1 = 0$	→	$f(x_1 = 0) = 0{,}125$ (1/8)
$\left.\begin{matrix} Z, & Z, & W \\ Z, & W, & Z \\ W, & Z, & Z \end{matrix}\right\}$	→	$x_2 = 1$	→	$f(x_2 = 1) = 0{,}375$ (3/8)
$\left.\begin{matrix} Z, & W, & W \\ W, & Z, & W \\ W, & W, & Z \end{matrix}\right\}$	→	$x_3 = 2$	→	$f(x_3 = 2) = 0{,}375$ (3/8)
(W, W, W)	→	$x_4 = 3$	→	$f(x_4 = 3) = 0{,}125$ (1/8)

Also:

$X = x_i$	0	1	2	3
$P(X = x_i) = f(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$



Wahrscheinlichkeitsfunktion der diskreten Zufallsvariablen 'Anzahl der Wappen'

Die **Verteilungsfunktion ('cumulative distribution function (cdf)') F(X)** gibt die Wahrscheinlichkeit an, daß die Zufallsvariable **X höchstens den Wert x** annimmt:

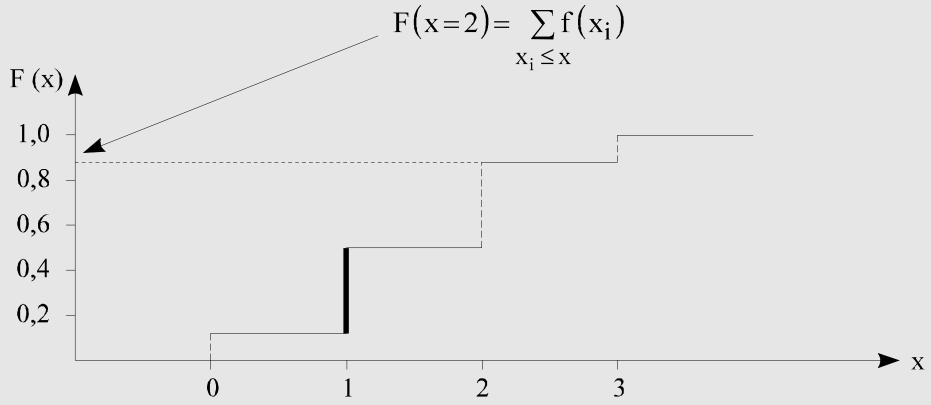
$$F(X) = P(X \le x) = \sum_{x_i \le x} f(x_i); \;\; mit \;\; f(x_i) = F(x_i) - F(x_{i-1});$$

Beispiel:

X = 'Anzahl Wappen' bei dreimaligem Münzwurf:

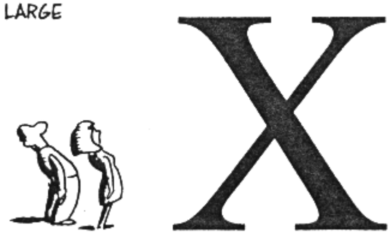
x	0	1	2	3
F (X)	0,125	0,500	0,875	1,000

diskret → **Treppenfunktion**



Verteilungsfunktion der diskreten Zufallsvariablen 'Anzahl der Wappen'

THE KEY IDEA IS THE ***RANDOM VARIABLE***, WHICH WE WRITE AS A LARGE



A RANDOM VARIABLE IS DEFINED AS THE ***NUMERICAL OUTCOME OF A RANDOM EXPERIMENT.***

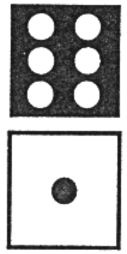
FOR EXAMPLE, IMAGINE DRAWING ONE STUDENT AT RANDOM FROM THE STUDENT BODY. THAT'S THE RANDOM EXPERIMENT. THE STUDENT'S ***HEIGHT, WEIGHT, FAMILY INCOME, S.A.T. SCORE, AND GRADE POINT AVERAGE*** ARE ALL NUMERICAL ***VARIABLES*** DESCRIBING PROPERTIES OF THE RANDOMLY SELECTED STUDENT. THEY'RE ALL ***RANDOM VARIABLES.***

ANOTHER EXAMPLE: TOSS TWO COINS (THE RANDOM EXPERIMENT) AND RECORD THE ***NUMBER*** OF HEADS: 0, 1, OR 2.

OUTCOME	TT	HT TH	HH
x	0	1	2

NOTE THE NOTATION! THE VARIABLE IS WRITTEN WITH A CAPITAL X. THE LOWERCASE x REPRESENTS A SINGLE VALUE OF X, FOR EXAMPLE $x=2$, IF HEADS COMES UP TWICE.

ANOTHER EXAMPLE IS BASED ON THE FAMILIAR TOSS OF TWO DICE. LET Y REPRESENT THE SUM OF THE DOTS ON THE TWO DICE. FOR THIS RANDOM VARIABLE, Y CAN BE ANY NUMBER BETWEEN 2 AND 12.



$Y=7$

NOW WE WANT TO LOOK AT THE ***PROBABILITIES*** OF THE OUTCOMES. FOR THE PROBABILITY THAT THE RANDOM VARIABLE X HAS THE VALUE x, WE WRITE $Pr(X=x)$, OR JUST $p(x)$. FOR THE COIN-FLIPPING RANDOM VARIABLE X, WE CAN MAKE THE TABLE:

x	0	1	2
$Pr(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

THIS TABLE IS CALLED THE ***PROBABILITY DISTRIBUTION*** OF THE RANDOM VARIABLE X.

FOR THE RANDOM VARIABLE Y (THE SUM OF TWO DICE), THE PROBABILITY DISTRIBUTION LOOKS LIKE THIS:

y	2	3	4	5	6	7	8	9	10	11	12
$Pr(Y=y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Gonick, Smith 1993

3 Dichte- und Verteilungsfunktion stetiger Zufallsvariablen

Bei einer stetigen (kontinuierlichen) Zufallsvariablen X kann man die Wahrscheinlichkeit **nur für Intervalle** angeben.

Wahrscheinlichkeitsdichte (Dichtefunktion)

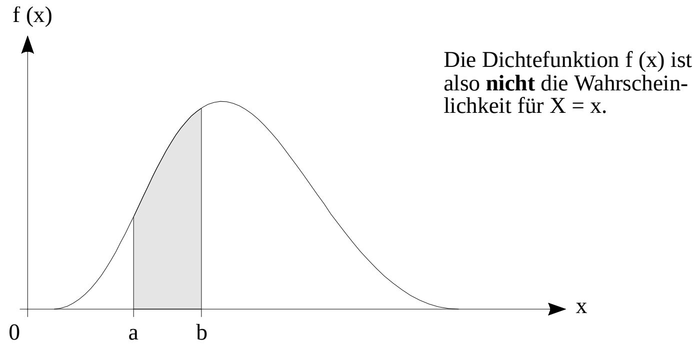
('probability distribution function' (pdf)) (density function)

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx \geq 0$$

Eigenschaften:

1. $P(a \leq x \leq b) \geq 0;\ f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1.$

Die **Fläche** unter der Dichtefunktion f(x) in den Grenzen a und b gibt die Wahrscheinlichkeit P(a ≤ x ≤ b) an:



Wahrscheinlichkeitsdichte (Dichtefunktion, pdf)

Die **Verteilungsfunktion F(x)** gibt bei einer stetigen Zufallsvariablen wieder die Wahrscheinlichkeit an, daß **X höchstens den Wert x** annimmt. Die Verteilungsfunkton ist hier dann keine Treppenfunktion, sondern auch stetig.

Verteilungsfunktion:

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

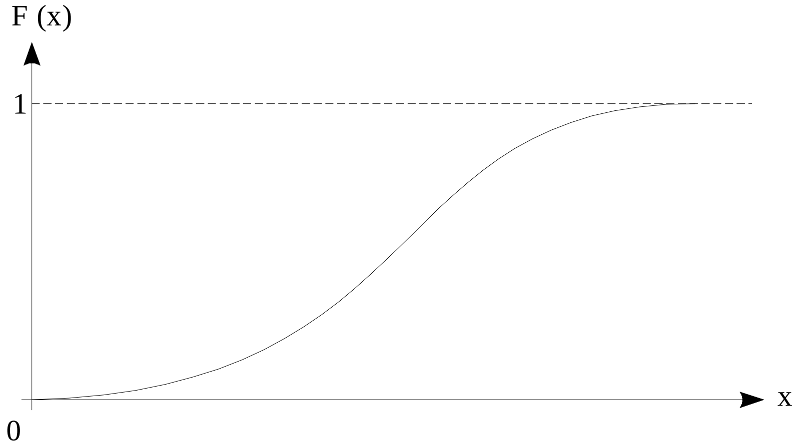
Eigenschaften von F(x):

1. $0 \leq F(x) \leq 1;$
2. $F(x)$ *monoton wachsend* $\Rightarrow$ *für* $x_1 < x_2$ *gilt* $F(x_1) \leq F(x_2);$
3. $\lim_{x \to -\infty} F(x) = 0;$
4. $\lim_{x \to +\infty} F(x) = 1.$

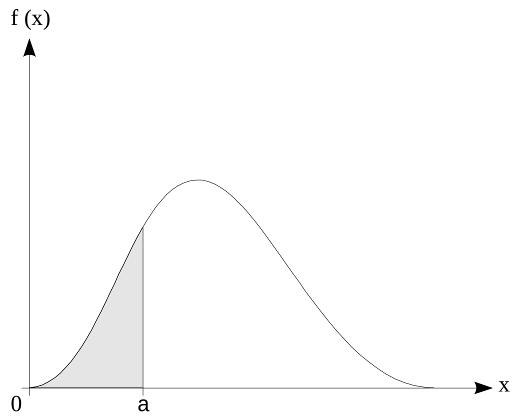
Es gilt:

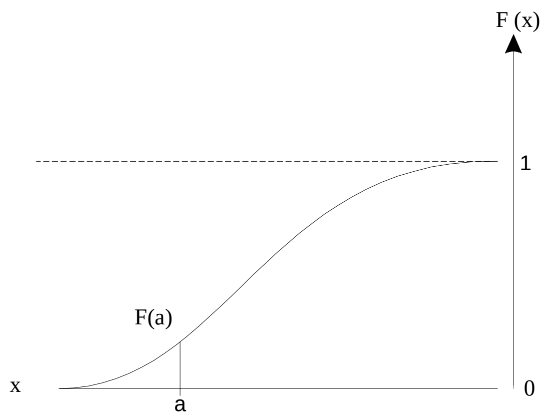
$$f(x) = \frac{dF(x)}{dx} = F'(x) \qquad 1.\ \textit{Ableitung der 'Stammfunktion' } F(x)$$

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$
$$= F(b) - F(a)$$
$$P(X = a) = \int_a^a f(x)\,dx$$
$$= F(a) - F(a) = 0$$



Verteilungsfunktion F(x) (cdf) einer stetigen Zufallsvariablen





$$P(X \le a) = F(a) = \int_{-\infty}^{a} f(t)\,dt$$

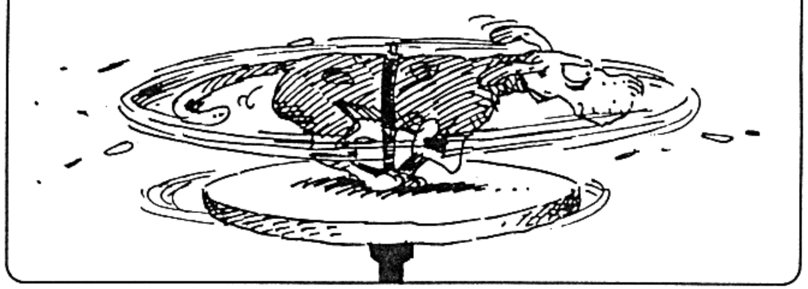
Die Fläche unter f(x) bis a Höhe von F(x) bei $X = a$.

OUR EXAMPLES SO FAR HAVE BEEN *DISCRETE* RANDOM VARIABLES. THEIR OUTCOMES ARE A SET OF ISOLATED ("DISCRETE") VALUES, LIKE THOSE WE SAW IN CHAPTER 3, BUT THERE ARE ALSO

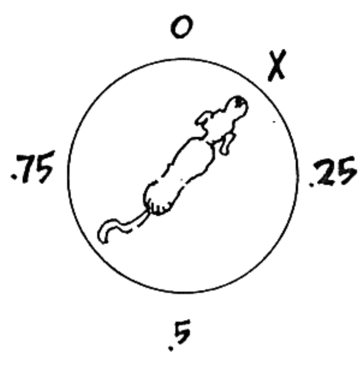
Continuous Random Variables

LET'S IMAGINE A RANDOM EXPERIMENT IN WHICH ***ALL OUTCOMES HAVE PROBABILITY ZERO.*** THAT'S RIGHT, $p(x) = 0$ FOR EVERY x.

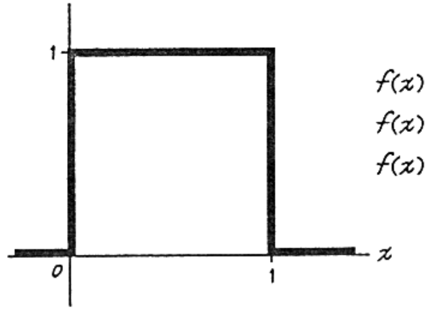
A SIMPLE EXAMPLE IS A BALANCED, SPINNING POINTER. IT CAN STOP ANYWHERE IN THE CIRCLE. IF X REPRESENTS THE PROPORTION OF THE TOTAL CIRCUMFERENCE IT LANDS ON, THE RANDOM VARIABLE X CAN TAKE ON ANY VALUE BETWEEN 0 AND 1—AN *INFINITE* RANGE OF VALUES.



SOME PROBABILITIES ARE EASY TO FIND, LIKE THE PROBABILITY THAT X FALLS WITHIN A RANGE: FOR EXAMPLE, $Pr(.25 \leq X \leq .75) = .5$, BECAUSE IT'S HALF THE CIRCLE. BUT WHAT ABOUT $Pr(X = .5)$? SINCE X CAN TAKE ON AN INFINITE NUMBER OF VALUES, AND ALL OF THESE VALUES ARE EQUALLY LIKELY, THE PROBABILITY THAT X IS EXACTLY .5 (OR EXACTLY ANYTHING) IS PRECISELY 0.



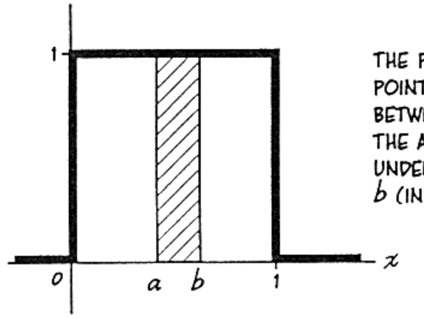
HOW CAN WE DRAW A PICTURE OF THIS? BY ANALOGY WITH THE CASE OF DISCRETE PROBABILITIES, WE TRY TO SEE CONTINUOUS PROBABILITIES AS ***AREAS UNDER SOMETHING.*** FOR THE SPINNING POINTER, THE "SOMETHING" LOOKS LIKE THIS:



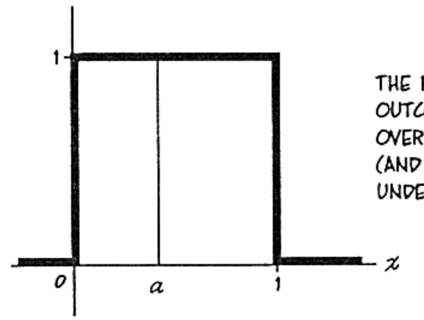
$f(x) = 0$ WHEN $x < 0$

$f(x) = 1$ WHEN $0 \leq x \leq 1$

$f(x) = 0$ WHEN $x > 1$



THE PROBABILITY THAT THE POINTER POINTS ANYWHERE BETWEEN a AND b IS PRECISELY THE AREA OF THE SHADED REGION UNDER THE CURVE BETWEEN a AND b (IN THIS CASE, $b-a$).



THE PROBABILITY OF AN EXACT OUTCOME, HOWEVER, IS THE "AREA" OVER A POINT, WHICH IS ***ZERO.*** (AND NOTE THAT THE TOTAL AREA UNDER THE CURVE IS EXACTLY 1.)

Gonick, Smith 1993

4 Parameter eindimensionaler Wahrscheinlichkeitsverteilungen

Wie die Häufigkeitsverteilungen in der deskriptiven Statistik, so lassen sich auch die Wahrscheinlichkeitsverteilungen von Zufallsvariablen durch entsprechende Maßzahlen charakterisieren.

4.1 Erwartungswert und Varianz

Erwartungswert (≙ arithmetisches Mittel)

Aus

$$\bar{x} = \frac{1}{n}\left(x_1 n_1 + x_2 n_2 + \ldots\right) = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \ldots = \sum_i x_i \frac{n_i}{n}$$

mit

$$\lim_{n\to\infty} \frac{n_i}{n} = P_i = f(x_i)$$

$$\lim_{n\to\infty} \bar{x} = \lim_{n\to\infty} \sum_i x_i \frac{n_i}{n} = \sum_i x_i\, f(x_i)$$

$$E(X) = \mu = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n) = \sum_i x_i\, f(x_i)$$

$$E(X) = \begin{cases} \sum_i x_i\, f(x_i) & x_i \text{ diskret} \\ \int_{-\infty}^{+\infty} x\, f(x)\, dx = \mu & x \text{ stetig } (\text{kontinuierlich}) \end{cases}$$

Beispiel:

Die stetige Zufallsvariable sei die in Minuten gemessene Verspätung eines Busses an einer bestimmten Haltestelle und habe die folgende Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{2} - \frac{1}{8}x & \textit{für } 0 \le \mathrm{x} \le 4 \\ 0 & \textit{sonst} \end{cases}$$

1. Wie groß ist die Wahrscheinlichkeit, daß X im Intervall [1, 2] liegt?

$$P(1 \le x \le 2) = \int_1^2 f(x)\, dx = \int_1^2 \left(\frac{1}{2} - \frac{1}{8}x\right) dx$$

$$= \left[\frac{1}{2}x - \frac{1}{16}x^2\right]_1^2 = 1 - \frac{4}{16} - \left(\frac{1}{2} - \frac{1}{16}\right) = \frac{5}{16} = 0{,}3125$$

2. Wie lautet die Verteilungsfunktion F(x)?

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x}\left(\frac{1}{2} - \frac{1}{8}t\right)dt = \left[\frac{1}{2}t - \frac{1}{16}t^2\right]_0^x$$

$$= \frac{1}{2}x - \frac{1}{16}x^2$$

also

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{2}x - \frac{1}{16}x^2 & \textit{für } 0 \le x \le 4 \\ 1 & \text{für } x > 4 \end{cases}$$

Die gesuchte Wahrscheinlichkeit (aus 1) von ein bis zwei Minuten Verspätung findet man auch über die Verteilungsfunktion:

$$P(1 \le x \le 2) = F(2) - F(1) = 0{,}75 - 0{,}4375 = 0{,}3125 \qquad (\text{wie oben!})$$

3. Wie groß ist im Mittel die Bus-Verspätung?

$$E(X) = \mu = \int_{-\infty}^{+\infty} x\, f(x)\,dx = \int_{a}^{b} x\, f(x)\,dx = \int_{0}^{4} x\left[\frac{1}{2} - \frac{1}{8}x\right]dx$$

$$= \int_{0}^{4}\left(\frac{1}{2}x - \frac{1}{8}x^2\right)dx = \left[\frac{1}{4}x^2 - \frac{1}{24}x^3\right]_0^4$$

$$= \left(4 - 2\frac{2}{3}\right) - 0 = 1\frac{1}{3}\ \textit{Minuten}$$

Im Mittel beträgt die Bus-Verspätung $1,\overline{33}$ Minuten.

Varianz

$$Var(X) = \sigma_x^2 = E\left\{\left[X - E(X)\right]^2\right\} = E\left(X^2\right) - \mu^2$$

$$Var(X) = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i) = \sum_i x_i^2 f(x_i) - \mu^2 & \text{X diskret} \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2 & X \text{ stetig} \end{cases}$$

Standardabweichung

$$\sigma_X = +\sqrt{\sigma_X^2}$$

Beispiele:

a) **diskrete Zufallsvariable**:

Bei einem Würfelspiel wird die gewürfelte Augenzahl als Gewinn ausgezahlt, höchstens jedoch 4 Gold-DOLLARS.

Wie groß ist der Erwartungswert der ausgezahlten Gewinnsumme?

X = 'ausgezahlte Gewinnsumme'

G	$x_i \in R_{x_i}$	$f(x_i)$	$x_i\, f(x_i)$
(1)	$x_1 = 1$	$\frac{1}{6}$	$\frac{1}{6}$
(2)	$x_2 = 2$	$\frac{1}{6}$	$\frac{2}{6}$
(3)	$x_3 = 3$	$\frac{1}{6}$	$\frac{3}{6}$
$(4),(5),(6)$	$x_4 = 4$	$\frac{3}{6}$	$\frac{12}{6}$
$\sum_i$		1	$\frac{18}{6} = 3$

$$\Rightarrow E(X) = \sum_i x_i\, f(x_i) = \mu = \frac{18}{6} = 3$$

Auf lange Sicht werden bei diesem Würfelspiel im Durchschnitt 3 Gold-DOLLARS pro Spiel als Gewinnsumme ausgezahlt.

Wie groß ist die Varianz der ausgezahlten Gewinnsumme?

1. Möglichkeit:

$$\sum_{i=1}^{4} (x_i - \mu)^2 \cdot f(x_i) = \sum_i (x_i - 3)^2 \cdot f(x_i)$$

$$= (1-3)^2 \cdot \frac{1}{6} + (2-3)^2 \cdot \frac{1}{6} + (3-3)^2 \cdot \frac{1}{6} + (4-3)^2 \cdot \frac{3}{6}$$

$$= \frac{4}{6} + \frac{1}{6} + 0 + \frac{3}{6} = 1\frac{1}{3}$$

2. Möglichkeit:

$$\sum_{i=1}^{4} x_i^2 \cdot f(x_i) - \mu^2 = \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{48}{6} - 9$$

$$= 10\frac{1}{3} - 9 = 1\frac{1}{3}$$

b) **stetige Zufallsvariable:**

Wie groß ist die Varianz im Falle des Bus-Verspätungsbeispiels?

1. Möglichkeit:

$$Var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\,dx = \int_0^4 \left(x - \frac{4}{3}\right)^2 \cdot \left(\frac{1}{2} - \frac{1}{8}x\right) dx$$

$$= \int_0^4 \left(x^2 - \frac{8}{3}x + \frac{16}{9}\right) \cdot \left(\frac{1}{2} - \frac{1}{8}x\right) dx$$

$$= \int_0^4 \left(-\frac{1}{8}x^3 + \frac{5}{6}x^2 - \frac{14}{9}x + \frac{8}{9}\right) dx$$

$$= \left[-\frac{1}{32}x^4 + \frac{5}{18}x^3 - \frac{7}{9}x^2 + \frac{8}{9}x\right]_0^4$$

$$= -\frac{256}{32} + \frac{320}{18} - \frac{112}{9} + \frac{32}{9} = -\frac{256}{32} + \frac{80}{9} = \frac{256}{288} = \frac{8}{9}$$

2. Möglichkeit:

$$Var(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2$$

$$= \int_0^4 x^2 \left(\frac{1}{2} - \frac{1}{8}x\right) dx - \left(\frac{4}{3}\right)^2$$

$$= \int_0^4 \left(\frac{1}{2}x^2 - \frac{1}{8}x^3\right) dx - \left(\frac{4}{3}\right)^2$$

$$= \left[\frac{1}{6}x^3 - \frac{1}{32}x^4\right]_0^4 - \left(\frac{4}{3}\right)^2$$

$$= 10\frac{2}{3} - 8 - 1\frac{7}{9} = \frac{8}{9} \quad \left[\text{Minuten}^2\right]$$

Standardabweichung

$$\sigma_x = \sqrt{\frac{8}{9}} = 0{,}9428 \text{ Minuten Streuung}$$

Tschebyscheffsche Ungleichung zur Abschätzung von P(X) in einem Intervall[2]

$$P(\mu - c\cdot\sigma \le x \le \mu + c\cdot\sigma) \ge 1 - \frac{1}{c^2}$$

2 Vgl. Schwarze 1991, S. 67 ff.

Eigenschaften des Erwartungswertoperators E bzw. des Varianzoperators

(Expected value, expectation, expectation operator E)

1. $E(a \cdot X + b) = a \cdot E(X) + b$ $\quad a, b$ Konstanten
2. $E\left((a \cdot X)^2\right) = a^2 \cdot E\left(X^2\right)$ $\quad X, Y$ Zufallsvariablen
3. $Var(a \cdot X + b) = a^2 \cdot Var(X)$

4. $E(X + Y) = E(X) + E(Y)$
5. $Var(X + Y) = Var(X) + Var(Y) + 2 \cdot Cov(X, Y)$

Unabhängigkeit zweier Zufallsvariablen X und Y:

6. $E(X \cdot Y) = E(X) \cdot E(Y)$
7. $Cov(X, Y) = 0$

Wenn zwei Zufallsvariablen unabhängig sind, dann ist ihre Kovarianz gleich Null. Es kann aber **nicht** der Umkehrschluß 'Kovarianz = 0 $\Rightarrow$ Unabhängigkeit' gezogen werden.

4.2 Konzept der Momente: Schiefe und Exzeß

Neben Erwartungswert und Varianz existieren analog zur deskriptiven Statistik

Median = Wert der Zufallsvariablen mit P(X) = 0,50 = $\frac{1}{2}$

Modus = Realisation x mit f(x) = max

Erwartungswert und Varianz sind Spezialfälle einer allgemeinen Klasse von Parametern zur Charakterisierung von Wahrscheinlichkeitsverteilungen:

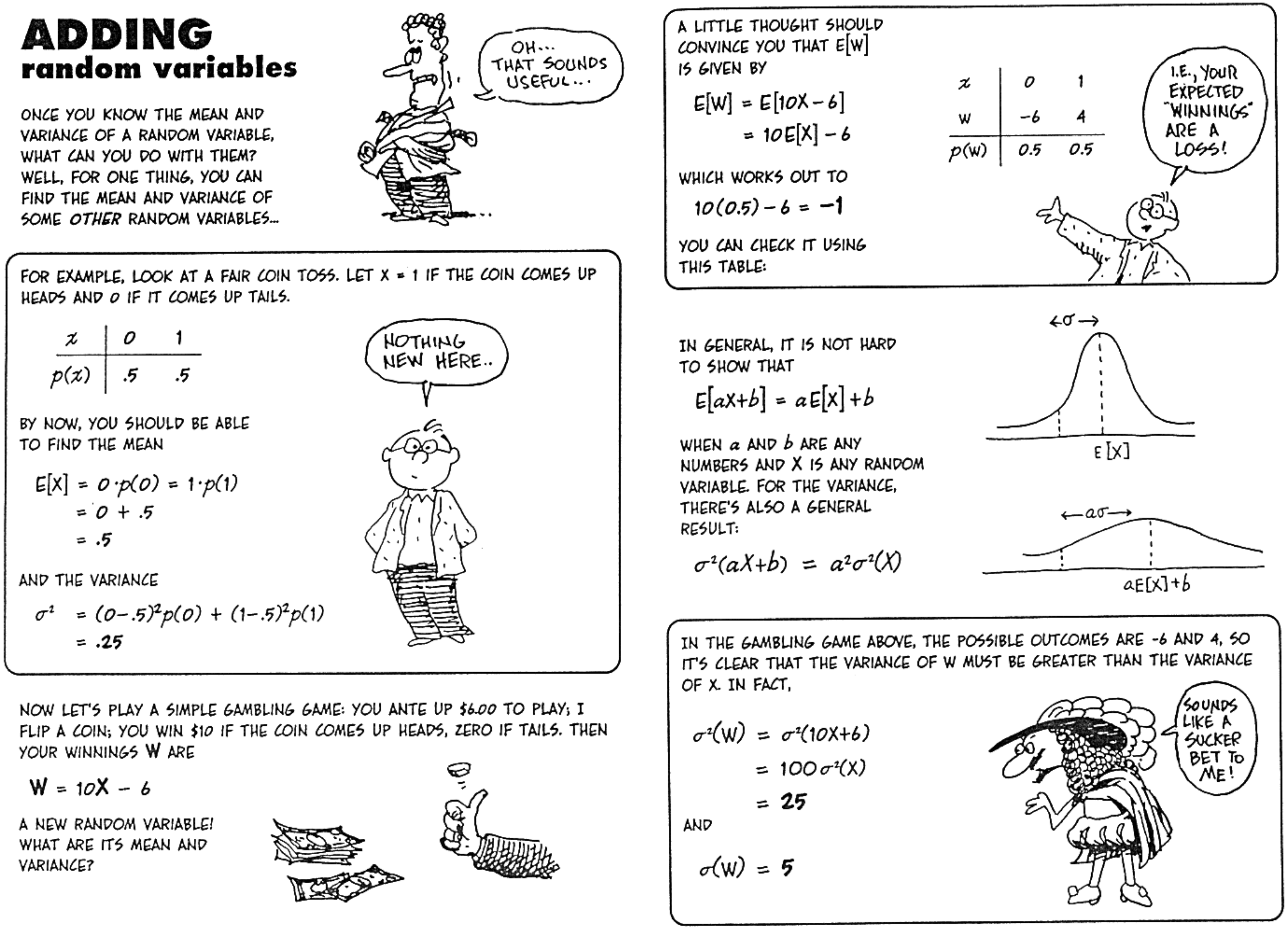
- Momente um Null
- zentrale Momente n-ter Ordnung

Zentrale Momente n-ter Ordnung:

$$E\left((X - E(X))^n\right) = \begin{cases} \sum_i \left(x_i - E(X)\right)^n \cdot f(x_i) & \textit{diskreter}\text{ Fall} \\ \int_{-\infty}^{+\infty} \left(x - E(X)\right)^n \cdot f(x)\, dx & \textit{stetiger}\text{ Fall} \end{cases}$$

Momente um Null: lediglich E(X) = 0;

Zwei besondere (und bekannte) Fälle: Schiefe ('skewness')
Exzeß, Wölbung ('kurtosis')



ADDING random variables

ONCE YOU KNOW THE MEAN AND VARIANCE OF A RANDOM VARIABLE, WHAT CAN YOU DO WITH THEM? WELL, FOR ONE THING, YOU CAN FIND THE MEAN AND VARIANCE OF SOME *OTHER* RANDOM VARIABLES...

OH... THAT SOUNDS USEFUL...

FOR EXAMPLE, LOOK AT A FAIR COIN TOSS. LET X = 1 IF THE COIN COMES UP HEADS AND 0 IF IT COMES UP TAILS.

x	0	1
$p(x)$	.5	.5

BY NOW, YOU SHOULD BE ABLE TO FIND THE MEAN

$$E[X] = 0\cdot p(0) = 1\cdot p(1)$$
$$= 0 + .5$$
$$= .5$$

AND THE VARIANCE

$$\sigma^2 = (0-.5)^2 p(0) + (1-.5)^2 p(1)$$
$$= \mathbf{.25}$$

NOTHING NEW HERE..

NOW LET'S PLAY A SIMPLE GAMBLING GAME: YOU ANTE UP $6.00 TO PLAY; I FLIP A COIN; YOU WIN $10 IF THE COIN COMES UP HEADS, ZERO IF TAILS. THEN YOUR WINNINGS **W** ARE

$$W = 10X - 6$$

A NEW RANDOM VARIABLE! WHAT ARE ITS MEAN AND VARIANCE?

A LITTLE THOUGHT SHOULD CONVINCE YOU THAT E[W] IS GIVEN BY

$$E[W] = E[10X-6]$$
$$= 10E[X] - 6$$

WHICH WORKS OUT TO

$$10(0.5) - 6 = \mathbf{-1}$$

YOU CAN CHECK IT USING THIS TABLE:

x	0	1
W	−6	4
$p(W)$	0.5	0.5

I.E., YOUR EXPECTED "WINNINGS" ARE A LOSS!

IN GENERAL, IT IS NOT HARD TO SHOW THAT

$$E[aX+b] = aE[X]+b$$

WHEN a AND b ARE ANY NUMBERS AND X IS ANY RANDOM VARIABLE. FOR THE VARIANCE, THERE'S ALSO A GENERAL RESULT:

$$\sigma^2(aX+b) = a^2\sigma^2(X)$$

$\leftarrow\sigma\rightarrow$

E[X]

$\leftarrow a\sigma\rightarrow$

aE[X]+b

IN THE GAMBLING GAME ABOVE, THE POSSIBLE OUTCOMES ARE -6 AND 4, SO IT'S CLEAR THAT THE VARIANCE OF W MUST BE GREATER THAN THE VARIANCE OF X. IN FACT,

$$\sigma^2(W) = \sigma^2(10X+6)$$
$$= 100\sigma^2(X)$$
$$= \mathbf{25}$$

AND

$$\sigma(W) = \mathbf{5}$$

SOUNDS LIKE A SUCKER BET TO ME!

Gonick, Smith 1993

Schiefe

$$E\left[(X-\mu)^3\right] \qquad E(X)=\mu$$

Maß der Asymmetrie: Schiefe negativ → linksschiefe/rechtssteile Verteilung,'long tail' in der negativen Richtung;

Bei Symmetrie

$$f(x-\mu)=f(x+\mu) \Rightarrow Schiefe=0$$

Exzeß (Wölbung)

$$E\left((X-\mu)^4\right)$$

Maß für die Wölbung ('thickness of tail'), je größer die Wölbung, desto flacher ist die Verteilung

Generell also

$$\mu_r=E\left((X-\mu)^r\right)$$

Da μ_r „explodiert", wenn r groß wird: ⇒ Normierung $\frac{\mu_r}{\sigma^r}$

standardisierte Schiefe = $\frac{\mu_3}{\sigma^3}$

standardisierte Wölbung = $\frac{\mu_4}{\sigma^4}-3$ $\left(\begin{array}{l}\text{bei Basierung auf 'Normalverteilung',}\\ \text{NV: Schiefe = 0 und Wölbung = 3}\end{array}\right)$

Zentrierte Zufallsvariable

$Y = X - \mu$ mit $E(Y) = 0$

Standardisierte Zufallsvariable

$$Z=\frac{X-\mu}{\sigma}$$

Es gilt $E(Z) = 0$ und $Var(Z) = 1$.

Beispiele:

Wie ist die Schiefe und die Wölbung für das Busverspätungsbeispiel zu charakterisieren?

Bisher: Wahrscheinlichkeitsdichte:

$$f(x)=\begin{cases}\frac{1}{2}-\frac{1}{8}x & \text{für } 0\le x\le 4\\ 0 & \text{sonst}\end{cases}$$

Verteilungsfunktion:

$$F(x)=\begin{cases} 0 & \textit{für } \mathrm{x}<0 \\ \frac{1}{2}x-\frac{1}{16}x^2 & \textit{für } 0\le x\le 4 \\ 1 & \textit{für } \mathrm{x}>4 \end{cases}$$

$$E(X)=1\frac{1}{3}$$

$$Var(X)=\frac{8}{9}$$

Welchen Wert hat die Schiefe $E\left((X-\mu)^3\right)$?

$$E\left((X-\mu)^3\right)=\int_{-\infty}^{+\infty}(x-\mu)^3\cdot f(x)\,dx$$

$$=\int_0^4\left(x-\frac{4}{3}\right)^3\cdot f(x)\,dx$$

$$=\int_0^4\left(x-\frac{4}{3}\right)^2\cdot\left(x-\frac{4}{3}\right)\cdot\left(\frac{1}{2}-\frac{1}{8}x\right)dx$$

$$=\int_0^4\left(x^2-\frac{8}{3}x+\frac{16}{9}\right)\cdot\left(\frac{1}{2}x-\frac{2}{3}-\frac{1}{8}x^2+\frac{1}{6}x\right)dx$$

$$=\int_0^4\left(\frac{1}{2}x^3-\frac{4}{3}x^2+\frac{8}{9}x-\frac{2}{3}x^2+\frac{16}{9}x-\frac{32}{27}-\frac{1}{8}x^4+\frac{1}{3}x^3-\frac{2}{9}x^2+\frac{1}{6}x^3-\frac{4}{9}x^2+\frac{8}{27}x\right)dx$$

$$=\int_0^4\left(-\frac{1}{8}x^4+x^3-\frac{24}{9}x^2+\frac{80}{27}x-\frac{32}{27}\right)dx$$

$$=\left[-\frac{1}{40}x^5+\frac{1}{4}x^4-\frac{8}{9}x^3+\frac{40}{27}x^2-\frac{32}{27}x\right]_0^4$$

$$=0,47$$

Schiefe = 0,47 > 0 $\Rightarrow$ linkssteil

normiert:

$$\frac{\mu_3}{\sigma^3}=\frac{0,47}{0,9428^3}=0,56$$

Welchen Wert hat die Wölbung (Exzeß) $E\left((X-\mu)^4\right)$?

$$E\left((X-\mu)^4\right)=\int_{-\infty}^{+\infty}(x-\mu)^4\cdot f(x)\,dx$$

$$=\int_0^4\left(x-\frac{4}{3}\right)^4\cdot\left(\frac{1}{2}-\frac{1}{8}x\right)dx$$

$$=\ldots$$

$$=\int_0^4\left(-\frac{1}{8}x^5+\frac{7}{6}x^4-4x^3+\frac{176}{27}x^2-\frac{416}{81}x+\frac{128}{81}\right)dx$$

$$=\left[-\frac{1}{48}x^6+\frac{7}{30}x^5-x^4+\frac{176}{81}x^3-\frac{208}{81}x^2+\frac{128}{81}x\right]_0^4$$

$$=1{,}8963$$

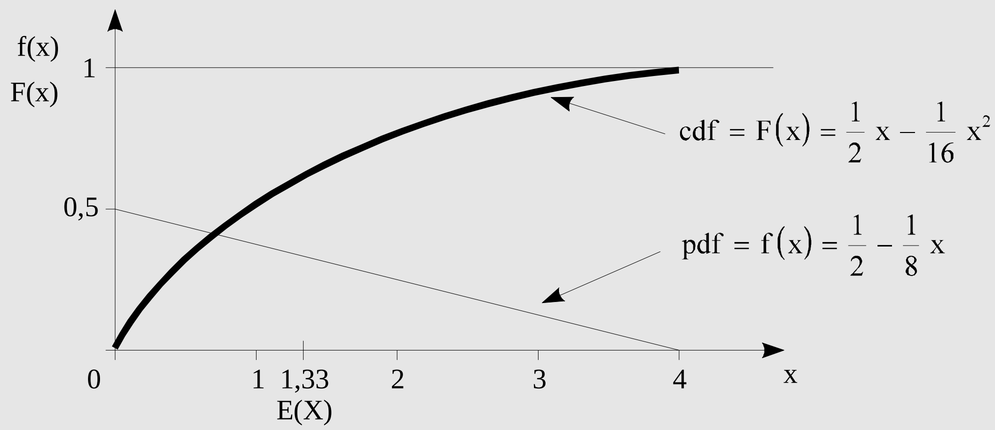
Wölbung = 1,8963

normiert:

$$\frac{\mu_4}{\sigma^4}-3=\frac{1{,}8963}{0{,}9428^4}-3=2{,}40-3=-0{,}60$$

Wölbung ist weniger flach, also stärker gewölbt, als die Normalverteilung (Normalverteilung ist flacher).

Graphische Zusammenfassung des Bus-Verspätungsbeispiels:



---

5 Mehrdimensionale Wahrscheinlichkeitsverteilungen

Zwei und mehr Zufallsvariablen werden gleichzeitig betrachtet (analog mehrdimensionaler Häufigkeitsverteilungen in der Deskription):

z.B. - gleichzeitiges Würfeln mit zwei Würfeln;
- Körpergröße und -umfang $\rightarrow$ Konfektionsgröße;

$\rightarrow$ zweidimensionale Zufallsvariable (X, Y)
X, Y können jeweils diskret oder stetig sein.

5.1 Wahrscheinlichkeitsverteilungen zweidimensionaler Zufallsvariablen

Die **Wahrscheinlichkeitsfunktion der diskreten Zufallsvariablen** (X, Y) gibt die Wahrscheinlichkeit an, daß die Zufallsvariable einen bestimmten Wert (x_i, y_j) mit der Wahrscheinlichkeit P_{ij} annimmt:

$$P\left(X=x_i \textbf{ und } \mathrm{Y}=\mathrm{y}_{\mathrm{j}}\right)=P\left(x_i,y_j\right)=P_{ij}=f\left(x_i,y_j\right) \qquad \Big| \sum_i \sum_j P_{ij}=1,$$

analog für stetige Zufallsvariablen (X, Y), so daß für die **gemeinsame Wahrscheinlichkeits- bzw. Dichtefunktion** ('joint density function') gilt:

$$\mathrm{Prob}\left(a\le x\le b \textbf{ und } \mathrm{c}\le y\le d\right)=\begin{cases}\displaystyle\sum_{a\le x\le b}\ \sum_{c\le y\le d} f\left(x,y\right) & \textit{diskret}\\ \displaystyle\int_a^b\int_c^d f\left(x,y\right)dy\,dx & \textit{stetig}\end{cases}$$

Für die **gemeinsame Verteilungsfunktion F(x, y) (cdf)** gilt:

$$F\left(x,y\right)=\mathrm{Prob}\left(X\le x,Y\le y\right)=\begin{cases}\displaystyle\sum_{X\le x}\ \sum_{Y\le y} f\left(x,y\right) & \text{diskret}\\ \displaystyle\int_{-\infty}^{x}\int_{-\infty}^{y} f\left(t,s\right)ds\,dt & \text{stetig}\end{cases}$$

Randverteilungen (marginale Verteilungen)

Unter der Randverteilung versteht man die Verteilung von (X, Y), die unabhängig davon ist, ob X (oder Y) als Bezugsgröße angenommen wird.

Randwahrscheinlichkeits- bzw. Randdichtefunktion

$$f_y\left(x\right)=\begin{cases}\displaystyle\sum_{j=-\infty}^{+\infty} f\left(x,y_j\right) & \text{diskret}\\ \displaystyle\int_{-\infty}^{+\infty} f\left(x,y\right)dy & \text{stetig}\end{cases}$$

Zwei Zufallsvariablen sind **statistisch (stochastisch) unabhängig**, wenn gilt:

$$f\left(x,y\right)=f_x\left(y\right)\cdot f_y\left(x\right) \qquad f\left(x_1,\ldots,x_n\right)=f_1\left(x_1\right)\cdot\ldots\cdot f_n\left(x_n\right)$$

bzw. *bzw.*

$$F\left(x,y\right)=F_x\left(y\right)\cdot F_y\left(x\right) \qquad F\left(x_1,\ldots,x_n\right)=F_1\left(x_1\right)\cdot\ldots\cdot F_n\left(x_n\right)$$

Bedingte Verteilung

Verteilung einer Zufallsvariablen X unter der Bedingung, daß Y einen bestimmten Wert Y = y annimmt.

Bedingte Wahrscheinlichkeits- bzw. Dichtefunktion

$$f(x|y) = \frac{f(x,y)}{f_x(y)}$$

Bedingte Verteilungsfunktion

$$F(x|y) = \frac{F(x,y)}{F_x(y)}$$

Beispiel:

Das PORTFOLIO einer Investmentgruppe bestehe aus Derivaten und Aktien. Eine Einschätzung der Gewinnentwicklung hat die verbundenen Wahrscheinlichkeiten von Derivaten und Aktien zu berücksichtigen.

5.2 Parameter zweidimensionaler Verteilungen: Erwartungswert, Kovarianz und Konzept der Momente

Definition in Bezug auf die Randverteilungen:

Erwartungswert

$$E(X) = \begin{cases} \sum_i x_i\, f_y(x_i) = \sum_i x_i \cdot \left[\sum_j f(x, y_j)\right] = \sum_i \sum_j x_i\, f(x_i, y_j) \\ \int_{-\infty}^{+\infty} x\, f_y(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f(x,y)\, dy\, dx \end{cases}$$

Varianz

$$\mathrm{Var}(X) = \begin{cases} \sum_i (x_i - E(X))^2\, f_y(x) = \sum_i \sum_j (x_i - E(X))^2\, f(x_i, y_j) \\ \int_{-\infty}^{\infty} (x-\mu)^2\, f_y(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu)^2\, f(x,y)\, dy\, dx \end{cases}$$

Für die Streuung zwischen zwei Zufallsvariablen X, Y gilt

Kovarianz

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E\left[(X-\mu_x)\cdot(Y-\mu_y)\right] = \sum_i \sum_j (x_i - \mu_x)(y_j - \mu_y)\, f(x_i, y_j) \\ &= E(X\,Y) - \mu_x\, \mu_y \\ &= \sigma_{xy} \end{aligned}$$

Wenn **X und Y unabhängig** sind, dann ist

$$f(x,y) = f_y(x) \cdot f_x(y)$$

und

$$\begin{aligned}\sigma_{xy} &= \sum_x \sum_y f_y(x) \cdot f_x(y) \cdot (x-\mu_x) \cdot (y-\mu_y) \\ &= \sum_x [x-\mu_x] \cdot f_y(x) \cdot \sum_y [y-\mu_y] \cdot f_x(y) \\ &= E(X-\mu_x) \cdot E(Y-\mu_y) \\ &= 0\end{aligned}$$

Das Vorzeichen der Kovarianz gibt die Richtung des Zusammenhangs zwischen X und Y an. Durch eine Normierung kann die Unabhängigkeit von der Skalierung erreicht werden.

Korrelationskoeffizient

$$r(X,Y) = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$\left(-1 \leq \rho_{xy} \leq 1\right)$ Interpretation analog zum nichtstochastischen Fall (Deskription)

Konzept der Momente

$$\mu_{rs} = E\left[(X-\mu_x)^r \cdot \left(Y-\mu_y\right)^s\right]$$

z. B. **zentrales Moment der Ordnung** (r, s) = (1 ; 1):

$$\mathrm{Cov}(X,Y) = E\left[\left(X-E(X)\right) \cdot \left(Y-E(Y)\right)\right] = E(XY) - E(X) \cdot E(Y)$$

6 Stochastische Simulation und Pseudo-Zufallszahlen

Simulation: Zielgerichtetes Experimentieren an Modellen;

Stochastisches Modell: Überlagerung eines deterministischen Teils mit einer **Zufallsschwankung ε**:

$$\mathrm{Y} = \sum_k \beta_k \cdot x_k + \varepsilon$$

→ Stochastische Simulation: aus vielen Computerläufen können dann der Erwartungswert E(Y) und andere Größen wie Var (Y) etc. gebildet werden.[3]

Pseudo-Zufallszahlen sind generierte Zufallszahlen.

Ein besonderes Problem von Pseudo-Zufallszahlen (-generatoren) ist die Vermeidung von wiederkehrenden Zyklen!

Es gibt verschiedene Methoden, um eine 'zuverlässige' Zufallszahl zu generieren.

[3] Vgl. Yang, Robinson 1986, S. 5 f.

Beispiele:

- **Gleichverteilte Zufallsvariable z.B. mit der linearen Kongruentenmethode**

Startwert I_0 (ungerade Integer-Zahl)

$$I_n = a \cdot I_{n-1} + c \; (\mathrm{mod}\; m)$$

$$U_n = \frac{I_n}{m}$$

mod (m) bezeichnet hier den modulo-Operator, der als Ergebnis den ganzzahligen Rest einer Division durch m liefert (Beispiel: 125 (mod 100) = 25).

Die lineare Kongruentenmethode liefert eine Sequenz von gleichverteilten Pseudozufallszahlen ('uniform pseudo random numbers') $U_1, U_2, U_3, \ldots$

Beispielsweise verwendet das IBM/370-System folgende Werte für a, c und m:

$a = 7^5 = 16807;$

$c = 0;$

$m = 2^{31} - 1.$

Diese Werte erfüllen die Bedingungen einer 'zuverlässigen' Zufallszahl.

- **Normalverteilte Zufallsvariable**

Dies stellt ein besonderes Problem dar, da keine geschlossene Form der Wahrscheinlichkeitsfunktion (Integral!) existiert. Dieses Problem wird im Abschnitt "Verteilungen" noch einmal aufgegriffen.

$w = \sqrt{-2 \ln U}$ U = Zufallszahl, **gleichverteilt** (0,1)

Für $0 < U \leq 0{,}5$

$$X = \Phi^{-1}(U) = \frac{a_0 + a_1 \cdot w + a_2 \cdot w^2}{1 + b_1 \cdot w + b_2 \cdot w^2 + b_3 \cdot w^3} - w$$

mit $a_0 = 2{,}515517$ $a_1 = 0{,}802853$ $a_2 = 0{,}0103028$

$b_1 = 1{,}432788$ $b_2 = 0{,}189269$ $b_3 = 0{,}001308$

für $0{,}5 < U < 1$:[4]

$$X = -\Phi^{-1}(1 - U)$$

wobei Φ^{-1} die Inverse der Standard-Normalverteilungsfunktion ist (vgl. auch IV.4)

Computerprogramme und Pseudo-Zufallszahlen

In manchen Computerprogrammen gibt es vorbereitete Prozeduren, die nach Aufruf eine Zufallszahl produzieren, die einer bestimmten Verteilung entspricht:

4 Vgl. Yang, Robinson 1986, S. 53.

Beispiel: Econometrics Toolkit ET

RNU (0 , 1)	gleichverteilt
RNG (μ , σ)	lognormalverteilt
RNP (μ)	Poisson-verteilt
etc.	

ET - Aufruf z.B.: CALC; Z = RNU (0 , 1) $

Zur Generierung von Pseudo-Zufallszahlen in Computerprogrammen vgl. auch den Abschnitt X.

Keyconcepts

Zufallsvariable
Ausprägungen einer Zufallsvariablen
Dichte-/Wahrscheinlichkeitsfunktion
Verteilungsfunktion
Erwartungswert
Varianz
Kovarianz
Schiefe
Wölbung
Mehrdimensionale Wahrscheinlichkeitsverteilung

III Diskrete Verteilungen

Darstellung der Wahrscheinlichkeiten des Auftretens bestimmter Ausprägungen einer diskreten Zufallsvariablen

Sowohl in der deskriptiven, wie auch in der induktiven Statistik lassen sich viele Berechnungen durch die Verwendung von Wahrscheinlichkeitsverteilungen vereinfachen.

In der deskriptiven Statistik dienen die theoretischen Verteilungen vor allem zur näherungsweisen Beschreibung empirisch beobachteter Verteilungen. Die Vereinfachung wird am besten durch ein Beispiel deutlich: Würde man beispielsweise die Einkommensverteilung in einer Gesellschaft beschreiben wollen, ohne auf eine theoretische Verteilung zurückzugreifen, so wäre es erforderlich jedes einzelne individuelle Einkommen zu erheben, eventuell in Guppen zu aggregieren und auszuwerten. Findet man hingegen eine theoretische Verteilung, mit der sich die Einkommensverteilung in der Gesellschaft „gut“ beschreiben lässt (im Fall von Einkommensverteilungen z.B. die Lognormalverteilung), ist es häufig nur noch erforderlich einige wenige Parameter zu berechnen (bei der Lognormalverteilung den Erwartungswert und die Varianz), um Aussagen über einzelne Fragen (Wieviel Prozent der Gesellschaft verdienen weniger als … €?) treffen zu können.

In der schließenden (induktiven) Statistik liegt der Vorteil der Verwendung von Wahrscheinlichkeitsverteilungen vor allem darin, dass sie geschlossene, analytische Lösungen für Ergebnisse von Zufallsexperimenten bieten, diese also nicht näherungsweise (iterativ) bestimmt werden müssen.

Diskrete Verteilungen: Zufallsvariable X ist diskret
Stetige Verteilungen: Zufallsvariable X ist stetig

Vor wichtigen Verteilungen für stetige Zufallsvariablen werden zunächst wichtige Verteilungen (in geschlossener Form) für diskrete Zufallsvariablen behandelt.

1 Gleichverteilung

Ist X eine diskrete Zufallsvariable mit x_i (i=1,...,n) als Realisationen, die mit den Wahrscheinlichkeiten $f(x_i)$ auftreten, dann heißt

X gleichverteilt, wenn jeder Wert x_i gleich wahrscheinlich ist.

Wahrscheinlichkeitsfunktion

$$f(x_i)=\frac{1}{n} \qquad (i=1,\ldots,n)$$

Verteilungsfunktion

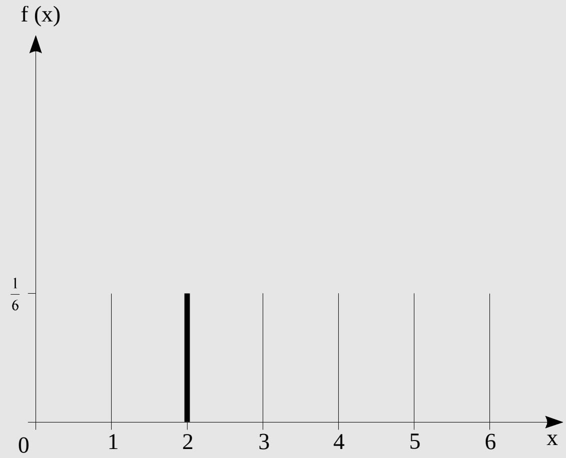
$$F(x_i) = \begin{cases} 0 & \textit{für } \mathrm{x} < \mathrm{x}_1 \\ \dfrac{i}{n} & \textit{für } \mathrm{x}_\mathrm{i} \leq x < x_{i+1} \quad (i = 1, \ldots, n-1) \\ 1 & \textit{für } \mathrm{x}_\mathrm{n} \leq x \end{cases}$$

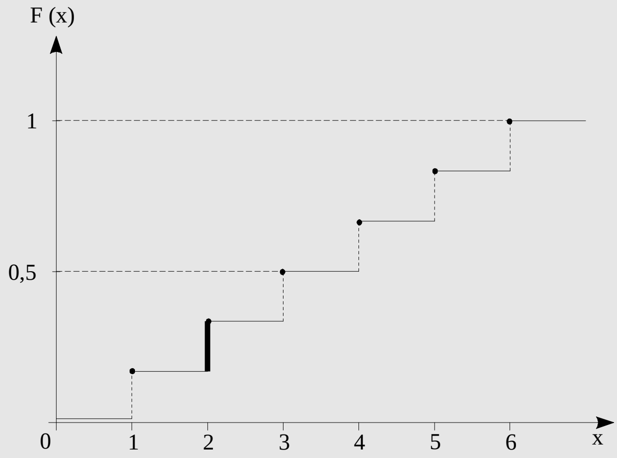
Beispiel:

Würfeln mit einem fairen Würfel:

$$f(x_i) = \frac{1}{6} \qquad x_i = i \qquad (i = 1, \ldots, 6)$$

$$F(x_i) = \begin{cases} 0 & \textit{für } x < 1 \\ \dfrac{i}{6} & \textit{für } 1 \leq x < 6 \\ 1 & \textit{für } 6 \leq x \end{cases}$$





2 Das Urnenmodell und das Bernoulli-Experiment

Zur gedanklichen Vereinfachung werden viele Zufallsexperimente in das sogenannte **Urnenmodell** überführt:
Elementarereignisse: (gleich große) grüne, rote, schwarze etc. Kugeln;

Ziehen mit Zurücklegen: unabhängige Ereignisse
Ziehen ohne Zurücklegen: abhängige Ereignisse

Für den Ziehenden: zufälliges Ereignis aus einer bestimmten Verteilung;

Bernoulli-Experiment (J. Bernoulli (1654–1705))

- Für jeden Versuch gibt es nur zwei mögliche Ergebnisse: A und das Komplement $\overline{\mathrm{A}}$.
- Die Erfolgswahrscheinlichkeiten p bzw. $1-\mathrm{p}$ der Ereignisse A bzw. $\overline{\mathrm{A}}$ sind konstant, ändern sich also von Versuch zu Versuch nicht.
- Die einzelnen Versuche sind voneinander unabhängig.

Dies **entspricht einem Urnenmodell** mit z.B. nur schwarzen und weißen Kugeln, wobei p = f(x) = Anteil schwarzer und 1 - p = Anteil weißer Kugeln ist.
Man zieht also mit Zurücklegen, damit p bzw. 1 - p unverändert bleiben.

Beispiel:

Die Ereignisse „Wappen“ oder „Zahl“ beim Werfen einer Münze;

Eine Folge von Bernoulli-Versuchen führt zur Binomialverteilung.

3 Binomialverteilung

Wie groß ist die Wahrscheinlichkeit, daß bei n unabhängigen Wiederholungen des Zufallsexperiments x-mal das Ereignis A und damit $(n-x)$-mal das Ereignis $\overline{\mathrm{A}}$ eintritt?

Also: n Versuche, wobei A mit der Wahrscheinlichkeit p und $\overline{\mathrm{A}}$ mit der Wahrscheinlichkeit $1-\mathrm{p}$ eintritt.

X = Anzahl der Ereignisse A bei n Versuchen

Gesucht: P(X = x) = ?

Eine Realisation:

$$\underbrace{A, A, A, \ldots}_{\text{x-mal}} \qquad \underbrace{\bar{A}, \bar{A}, \bar{A}, \ldots}_{(n-x)\text{-mal}}$$

Da die einzelnen Versuche unabhängig voneinander sind, gilt nach dem Multiplikationssatz (logisches UND):

$$\underbrace{\mathrm{p} \cdot \mathrm{p} \cdot \mathrm{p} \cdot \ldots}_{\text{x-mal}} \qquad \underbrace{(1-p) \cdot (1-p) \cdot (1-p) \cdot \ldots}_{(n-x)\text{-mal}}$$

$$P(\text{x-mal A}) = p^x \qquad P\left((n-x)\text{-mal } \overline{\mathrm{A}}\right) = (1-p)^{n-x}$$

P(für *eine bestimmte* Realisation) = $p^x \cdot (1-p)^{n-x}$

WE BEGIN WITH THE DISCRETE ONE, CALLED THE ***BINOMIAL*** RANDOM VARIABLE. SUPPOSE WE HAVE A RANDOM PROCESS WITH JUST ***TWO POSSIBLE OUTCOMES:*** A HEADS-OR-TAILS COIN TOSS, A WIN-OR-LOSE FOOTBALL GAME, A PASS-OR-FAIL AUTOMOTIVE SMOG INSPECTION. WE ARBITRARILY CALL ONE OF THESE OUTCOMES A *SUCCESS* AND THE OTHER A *FAILURE.*

WHAT WE DO IS TO REPEAT THIS EXPERIMENT... WELL, REPEATEDLY. SUCH A REPEATABLE EXPERIMENT IS CALLED A

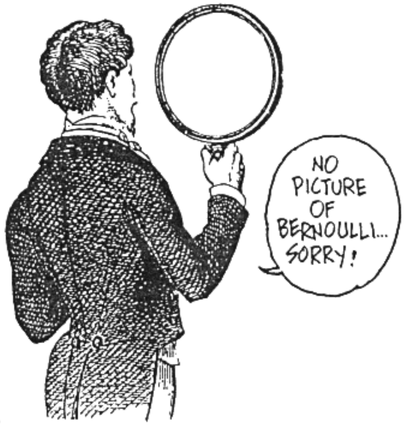
Bernoulli trial,

PROVIDED IT HAS THESE CRITICAL PROPERTIES:

1) THE RESULT OF EACH TRIAL MAY BE EITHER A SUCCESS OR A FAILURE

2) THE PROBABILITY p OF SUCCESS IS THE SAME IN EVERY TRIAL.

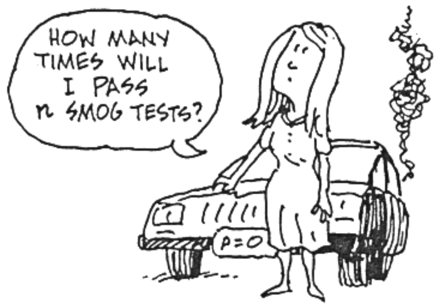
3) THE TRIALS ARE *INDEPENDENT:* THE OUTCOME OF ONE TRIAL HAS NO INFLUENCE ON LATER OUTCOMES.



STARTING WITH A BERNOULLI TRIAL, WITH PROBABILITY OF SUCCESS p, LET'S BUILD A NEW RANDOM VARIABLE BY REPEATING THE BERNOULLI TRIAL.

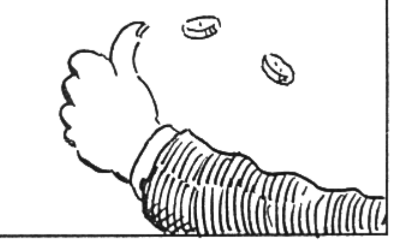
The binomial random variable

X IS THE ***NUMBER OF SUCCESSES*** IN n REPEATED BERNOULLI TRIALS WITH PROBABILITY p OF SUCCESS.



AN EXAMPLE OF A BINOMIAL RANDOM VARIABLE IS THE NUMBER OF HEADS (SUCCESSES) IN TWO FLIPS OF A COIN. HERE $n = 2$ AND $p = .5$

k = NUMBER OF SUCCESSES	0	1	2
$Pr(X=k)$	.25	.5	.25



ANOTHER EXAMPLE IS DE MERE'S FIRST GAMBLE: TOSSING A SINGLE DIE FOUR TIMES IN A ROW. SUCCESS MEANS ROLLING A 6. THE DISTRIBUTION IS:

Gonick, Smith 1993

Alle Realisationen:

Wieviele verschiedene Anordnungen (Reihenfolgen, Möglichkeiten) gibt es, bei n Versuchen x-mal das Ereignis A zu erhalten?

Kombinatorik:
Auswahl x aus n, Reihenfolge **un**wichtig, **ohne** Wiederholung $(AA \hat{=} AA)$.

$$Kombination:\ C(x,n)=\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

Es gibt damit $\binom{\text{n}}{\text{x}}$ verschiedene Möglichkeiten, um bei n Versuchen x-mal das Ereignis A zu erhalten. Jede Anordnung, Folge, Möglichkeit besitzt die gleiche Wahrscheinlichkeit $p^x \cdot (1-p)^{n-x}$.

Da sich die einzelnen Folgen gegenseitig ausschließen, addieren sich die Wahrscheinlichkeiten für die einzelnen Reihenfolgen (Additionssatz für sich ausschließende Ereignisse (ODER)).

Wahrscheinlichkeitsfunktion der Binomialverteilung

$$P(X=x)=f_B(x|\ n,p)=\binom{n}{x}p^x(1-p)^{n-x}$$

Verteilungsfunktion

$$P(X\leq x)=F_B(x|\ n,p)=\sum_{k=0}^{X}\binom{n}{k}p^k(1-p)^{n-k}$$

Beispiel:

a) Wie groß ist die Wahrscheinlichkeit bei n = 5 Würfen einer Münze **genau x = 2** mal Ziffer zu werfen?

$$P(X=2)=f_B(2|5;0,5)=\binom{5}{2}\cdot 0,5^2\cdot(1-0,5)^{5-2}=0,3125$$

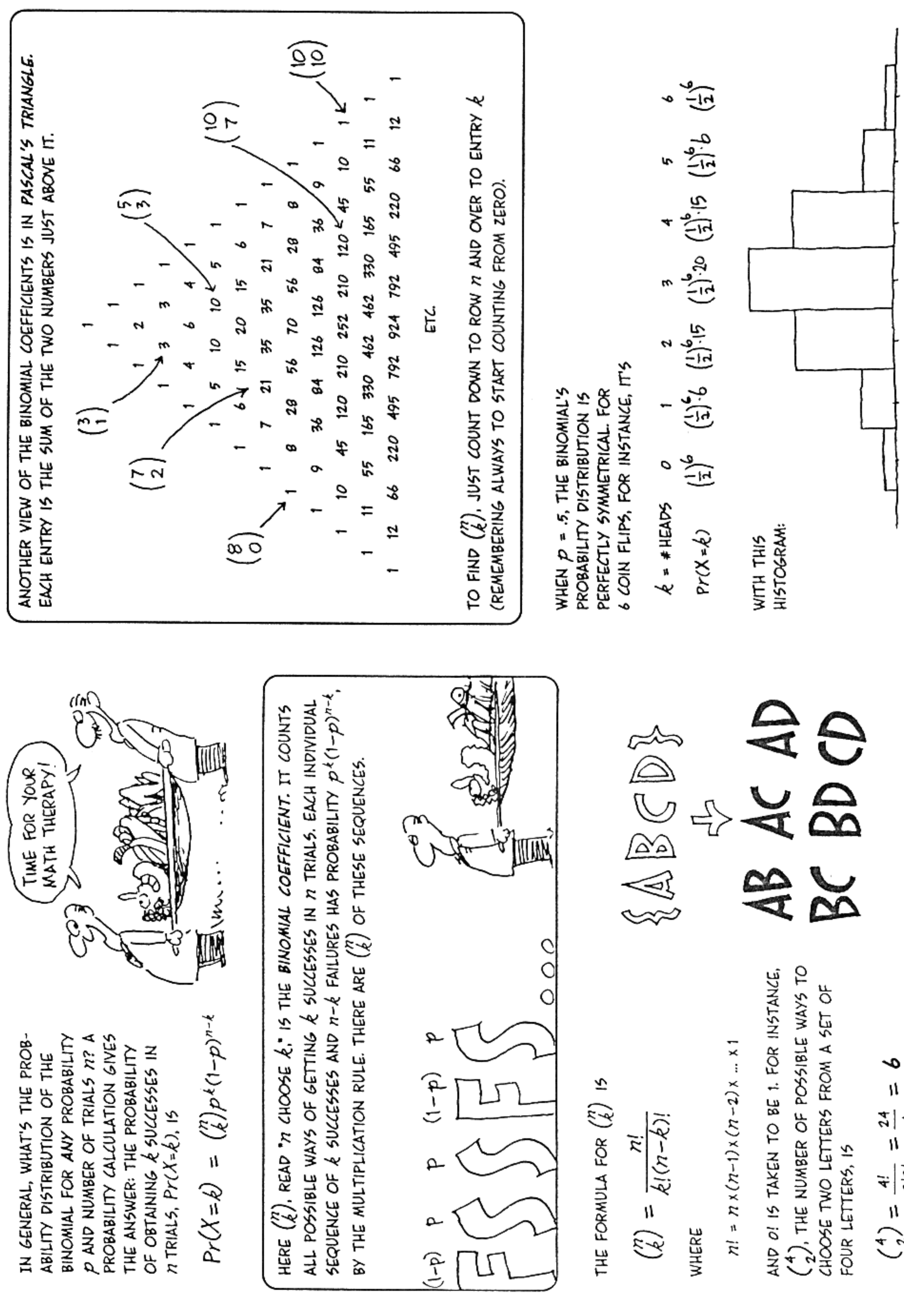
b) Wie groß ist die Wahrscheinlichkeit bei n = 5 Würfen **höchstens zweimal** Ziffer zu werfen (X ≤ x = 2)?

$$P(X\leq x=2)=F_B(2|5;0,5)=\sum_{k=0}^{2}\binom{n}{k}p^k(1-p)^{n-k}$$

$$F_B(2|5;0,5)=F(x\leq 2)=\binom{5}{0}\cdot 0,5^0\cdot(1-0,5)^{5-0}+\binom{5}{1}\cdot 0,5^1\cdot(1-0,5)^{5-1}+\binom{5}{2}\cdot 0,5^2\cdot(1-0,5)^{5-2}$$

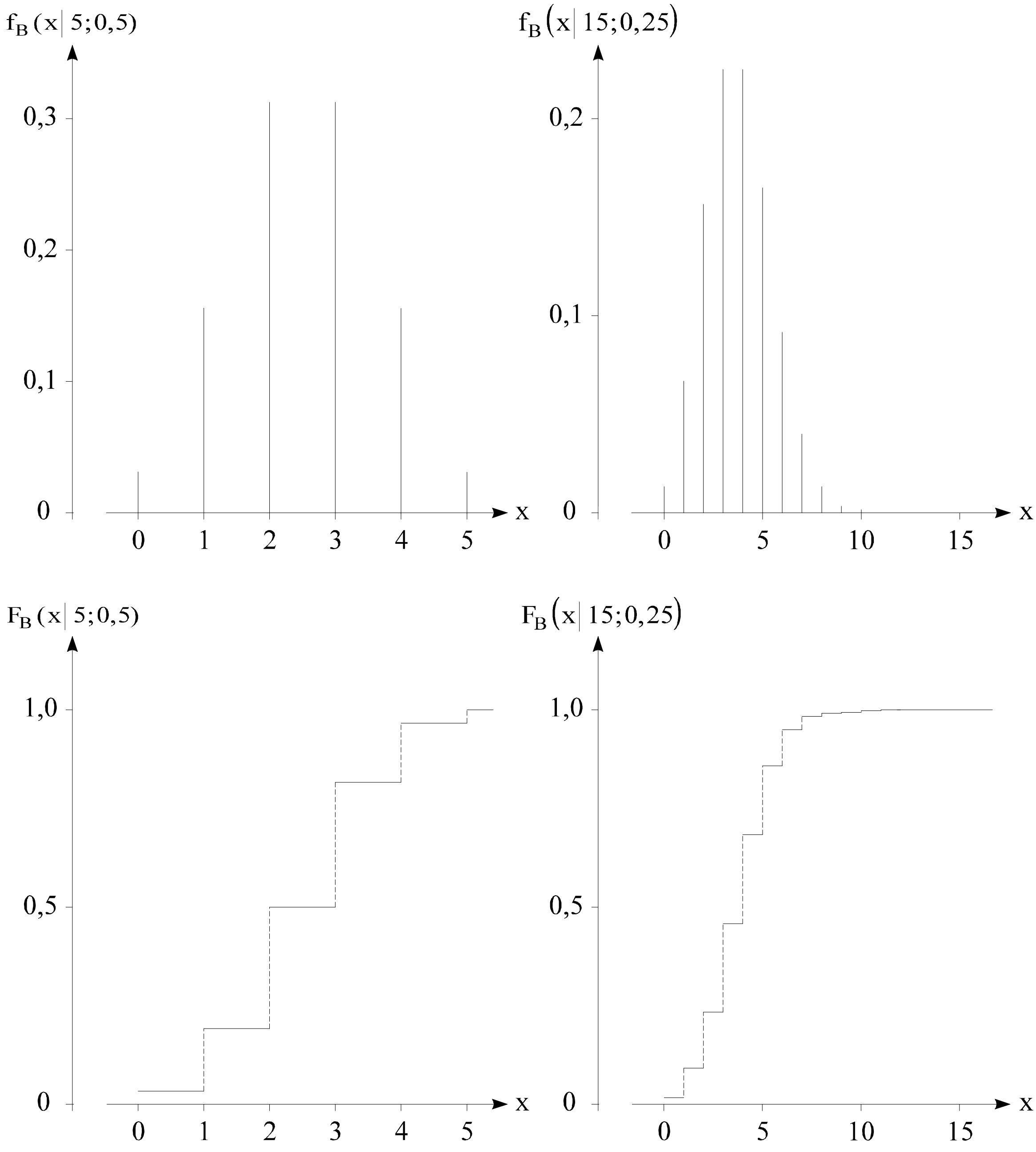
$$=1\cdot 1\cdot 0,03125+\frac{5!}{1!\cdot(5-1)!}\cdot 0,5\cdot 0,0625+0,3125$$

$$=0,5$$



Gonick, Smith 1993

Graphische Beispiele (Binomialverteilung):



Die Gestalt der Binomialverteilung hängt also von den Parametern n und p ab!

Erwartungswert

$$E(X)=n\cdot p$$

Varianz

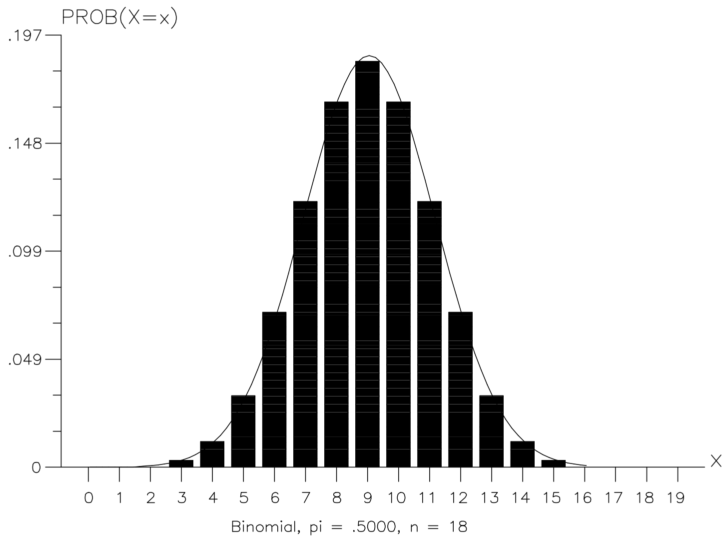
$$Var(X)=n\cdot p\cdot(1-p) \qquad (1-p)=q$$

Binomialverteilungen sind unimodale Verteilungen, die
für $p = q$ symmetrisch
$p < 0{,}5$ linkssteil sind.

ET: Probability, Discrete Distributions, Binomial Distribution

Discrete
1 Binomial
2 Poisson
3 Geometric
4 Hypergeom.
5 N**E**g. Binom

Probabilities of x successes in n independent trials from a population in which the prob. of a success on any draw is π. x = 0,1,..,n=the number of successes
p[x]=nCx π^x(1-π)^(n-x), Mean=nπ,Variance =nπ(1π)
(Notation: nCx = n!/[x!(n-x)!] · x^y = x to y power.)



x	Prob[X=x]	x	Prob[X=x]	x	Prob[X=x]	x	Prob[X=x]
0	.000000	5	.032684	10	.166924	15	.003113
1	.000000	6	.070816	11	.121399	16	.000584
2	.000584	7	.121399	12	.070816	17	.000000
3	.003113	8	.166924	13	.032684	18	.000000
4	.011673	9	.185471	14	.011673		

Success probability π	.500	0 < π < 1
Sample size, n	18	0 < n ≤ 99

Mean= 9.0000, Std. Deviation= 2.121

Move the cursor to the computation you want with ↑ or ↓, enter the value, then press Enter. Esc=Exit

Prob	P[X=x]	what is x? **6**	P= **.071**	1-P=**.929**
Cumul	P[X≤x]	what is x?	P=	1-P=
Range	P[a≤X≤b]	what is a?	b?	P=
Invrs	x for P[X≤x]=P:	what is P?		X=

Source: W. Greene, Econometrics Toolkit, Version 3.0, Econometric Software, Inc. New York 1992, S. 57 ff.

Beispiel:

Zur Fertigung eines Werkstückes stehen acht Maschinen zur Verfügung. Man hat festgestellt, daß die einzelnen Maschinen mit einer Wahrscheinlichkeit von p = 0,8 (unabhängig voneinander) arbeiten.

a) Wieviele Maschinen arbeiten langfristig im Durchschnitt?

b) Wie groß ist die Streuung?

a) Binomialverteilung: $n = 8$ $p = 0{,}8$

$$E(X) = \mu = n \cdot p = 8 \cdot 0{,}8 = 6{,}4$$

6,4 Maschinen arbeiten im Durchschnitt.

b) $Var(X) = \sigma^2 = n \cdot p \cdot (1-p) = n \cdot p \cdot q = 8 \cdot 0{,}8 \cdot 0{,}2 = 1{,}28$

$$\sigma = \sqrt{1{,}28} = 1{,}13$$

Die Streuung beträgt 1,13 Maschinen.

Satz:

Sind X und Y unabhängige Zufallsvariablen und binomialverteilt mit B(n , p) bzw. B(m, p), so ist die Zufallsvariable X + Y ebenfalls binomialverteilt mit B(m + n, p).

Binomialverteilung und Tabellenwerte

Formelsammlung (Anhang):

f_B, F_B für n = 1,..., 20 mit jeweils
x = 0,..., n
und p = 0,05 ; 0,10 ;...; 0,50

Für Werte p > 0,5: $x \rightarrow x^* = n - x$
$p \rightarrow p^* = 1 - p$

Beispiel:

Eine Urne ist zu 70 % mit roten Kugeln gefüllt. Zwölf Kugeln werden nacheinander mit Zurücklegen gezogen.
Gesucht: P(X = x = 8 rote Kugeln)

n = 12 x = 8 p = 0,7
da p > 0,5

Wahrscheinlichkeit P (X = 8) = f_B:

$$f_B(x \mid n, p) = f_B(n - x \mid n, 1 - p)$$

$$f_B(4 \mid 12; 0{,}3) = 0{,}2311$$

Wahrscheinlichkeit P (X ≤ 8) = F_B:

$$\begin{aligned} F_B(x|\,n,p) &= 1 - F_B(n-1-x|\,n, 1-p) \\ &= 1 - F_B(12-1-8|\,12;0,3) \\ &= 1 - F_B(3|\,12;0,3) \\ &= 1 - 0{,}4925 \\ &= 0{,}5075 \end{aligned}$$

4 Hypergeometrische Verteilung

Urnenmodell **ohne** Zurücklegen:

Aus einer Urne mit M schwarzen und N – M weißen Kugeln (also insgesamt N Kugeln) wird die Zufallsstichprobe n entnommen, ohne daß die einzelne Kugel zurückgelegt wird.

Wie groß ist die Wahrscheinlichkeit, daß unter den n gezogenen Kugeln x schwarze zu finden sind?

1. Die Anzahl der Möglichkeiten einer Auswahl von x schwarzen Kugeln aus M schwarzen Kugeln berechnet man durch (Kombinatorik):

   $$\binom{M}{x}$$

   Analog: Anzahl der Möglichkeiten (Reihenfolgen), aus N – M weißen Kugeln (n – x) weiße Kugeln zu ziehen:

   $$\binom{N-M}{n-x}$$

2. Da jede einzelne Möglichkeit 'x aus M' mit jeder einzelnen Möglichkeit 'n – x aus N – M' kombiniert werden kann, errechnet sich die Gesamtanzahl aller Möglichkeiten, daß genau x der n Kugeln schwarz sind durch:

   $$\binom{M}{x} \cdot \binom{N-M}{n-x}$$

3. Als Gesamtzahl der Möglichkeiten, aus einer Urne mit N Kugeln eine Auswahl (Stichprobe) mit n zu ziehen, ergibt sich:

   $$\binom{N}{n}$$

4. Mit Hilfe des Laplaceschen Wahrscheinlichkeitsbegriffs ergibt sich dann die

Hypergeometrische Verteilung (Wahrscheinlichkeitsfunktion, pdf)

$$f_H\left(x\middle|\, N, n, M\right) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$$

Verteilungsfunktion (Summation der Einzelwahrscheinlichkeiten, P(X ≤ x), cdf)

$$F_H\left(x\middle|\, N, n, M\right) = \sum_{k=0}^{x}\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$$

X ist verteilt H(N, n, M).

Erwartungswert

$$E(X) = n\cdot\frac{M}{N}$$

Varianz

$$Var(X) = n\cdot\frac{M}{N}\cdot\frac{N-M}{N}\cdot\frac{N-n}{N-1}$$

Approximation:

Für große N, M, N – M und dazu relativ kleinen Wert n kann die hypergeometrische Verteilung gut durch die Binomialverteilung angenähert (approximiert) werden.

Faustregel:

$$\frac{n}{N}\leq 0{,}05 \qquad \left(\frac{M}{N}\mathrel{\widehat{=}} p \text{ der Binomialverteilung}\right)$$

Graphisches Beispiel:

Computerprogramm Econometrics Toolkit ET:

Befehlsequenz : PROB → Hypergeom. → Parameter:

hier	ET	
N	P	('*population*')
M	m	('*successes*')
n	n	*Stichprobe*

Beispiele:

1. N = P = 15
   M = m = 8
   n = 5

Mit welcher Wahrscheinlichkeit zieht man genau drei schwarze Kugeln aus einer entsprechenden Urne mit $\frac{8}{15}$ = 53,33 % schwarzen Kugeln?

$$pdf:\quad P\left(x=3\,\middle|\,N=15,n=5,M=8\right)=\frac{\binom{8}{3}\cdot\binom{7}{2}}{\binom{15}{5}}=39,2\,\%$$

cdf: maximal drei schwarze Kugeln

$$F_H\left(x\le 3\,\middle|\,N=15,n=5,M=8\right)=\sum_{k=0}^{3}\frac{\binom{8}{k}\cdot\binom{7}{5-k}}{\binom{15}{5}}=81,8\,\%$$

(ET: Set Parameters → Cumulative → p=15, n=5, m=8 → Probabilities → x=3 ⇒ F_H)

2. P(6 Richtige im Lotto '6 aus 49') = ?

   N = 49 M = 6 N – M = 43 n = x = 6

$$f_H\left(x=6\,\middle|\,N=49,n=6,M=6\right)=\frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}=\frac{\binom{6}{6}\cdot\binom{43}{0}}{\binom{49}{6}}$$

$$=\frac{1}{\binom{N}{n}}=\frac{1}{\binom{49}{6}}=\frac{1}{\frac{49!}{6!(49-6)!}}=7,15\cdot 10^{-8}$$

   P(5 Richtige im Lotto '6 aus 49') = ?

   x = 5 , sonst wie oben:

$$f_H\left(x=5\,\middle|\,N,n,M\right)=0,0000184=1,845\cdot 10^{-5}$$

3. P(mindestens 5 Richtige) = ?

$$=7,15\cdot 10^{-8}+1,845\cdot 10^{-5}=1,85215\cdot 10^{-5}$$

5 Poissonverteilung

Es handelt sich hier wiederum um ein Bernoulli-Experiment, wobei aber die Erfolgswahrscheinlichkeit sehr klein und die Anzahl der Experimente sehr groß ist. Dann ist die Poissonverteilung (S. O. POISSON (1837)) Grenzfall der Binomialverteilung. In diesem Fall wäre die Binomialverteilung nur sehr aufwendig zu berechnen:

$p \to 0$
$n \to \infty$
$n \cdot p = \mu = E(X) = \text{konstant}$

$$\Rightarrow \lim_{\substack{p\to 0 \\ n\to\infty \\ n\cdot p = const.}} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = \frac{\mu^x \cdot e^{-\mu}}{x!} \qquad (\textit{ohne}\text{ Beweis!})$$

Wahrscheinlichkeitsfunktion der Poissonverteilung

$$f_p(x|\mu) = \frac{\mu^x \cdot e^{-\mu}}{x!} \qquad e = 2,7182...$$

Verteilungsfunktion der Poissonverteilung

$$F_p(x|\mu) = \sum_{k=0}^{x} \frac{\mu^k \cdot e^{-\mu}}{k!}$$

Erwartungswert

$$E(X) = \mu$$

Varianz

$$Var(X) = \mu$$

Aufeinanderfolgende Glieder der Wahrscheinlichkeitsfunktion lassen sich mit der folgenden **Rekursionsformel** berechnen:

$$f(x+1|\mu) = \frac{\mu}{x+1} \cdot f(x|\mu)$$

Poissonverteilung: für Modelle mit Zähldaten ('discrete choice' - Modelle)

Approximation der Binomialverteilung durch Poissonverteilung:

Faustregel: $n > 100$ und $p < 0,05$

Beispiel:

In einer Untersuchung zur Umweltbelastung der bundesdeutschen Bevölkerung wurde in 100 Arztpraxen die Zahl der Patienten gezählt, die in einem Jahr nachweislich durch erhebliche Umweltbelastungen tödlich erkrankt sind. Um das Datenmaterial in weitere Überlegungen

einzubringen, sollte versucht werden, die empirischen Häufigkeiten durch eine theoretische Verteilung zu approximieren.
Als Kandidat wurde die Poisson-Verteilung vorgeschlagen. Die Poisson-Verteilung habe den Parameter μ = 0,63.

Zahl der tödlich erkrankten Patienten	N_i	$\frac{n_i}{n}$	p_{x_i}	n_i (Poisson)
0	54	0,54	0,533	53
1	32	0,32	0,335	34
2	11	0,11	0,100	10
3	2	0,02	0,022	2
4	1	0,01	0,004	0
5	0	0,00	0,000	0

Relativ gute Approximation des Datenmaterials durch eine Poisson-Verteilung mit μ = 0,63.

6 Geometrische Verteilung

Fragt man nun, wie oft man den Bernoulli-Prozeß durchführen muß, bis man zum ersten Mal das Ereignis A beobachtet, so gelangt man zur geometrischen Verteilung. Die Zufallsvariable X bezeichne die Anzahl der Versuche bis einschließlich zum ersten Auftreten von A. Dann gilt:

Wahrscheinlichkeitsfunktion der geometrischen Verteilung

$$f_g(x|\,p) = p \cdot (1-p)^{x-1} \qquad x = 1, 2, \ldots$$

Verteilungsfunktion der geometrischen Verteilung

$$F_g(x|\,p) = \begin{cases} 0 & \textit{für } \mathrm{x} < 1 \\ 1-(1-p)^m & \textit{für } \mathrm{m} \le \mathrm{x} < \mathrm{m}+1; \quad \mathrm{m} = 1, 2, \ldots \end{cases}$$

Erwartungswert

$$E(X) = \frac{1}{p}$$

Varianz

$$Var(X) = \frac{1-p}{p^2}$$

Beispiel:

Beim Roulette beträgt die Wahrscheinlichkeit, daß eine rote Zahl fällt

$$P(rot) = p = \frac{18}{37}.$$

Wie groß ist die Wahrscheinlichkeit, daß beim 10. Spiel zum ersten Mal Rot eintritt?

$$f_g(10|\, p) = \frac{18}{37} \cdot \left(\frac{19}{37}\right)^9 = 0,001208$$

Der Erwartungswert beträgt

$$E(X) = \frac{1}{p} = \frac{37}{18} = 2,05556.$$

Die Varianz beträgt

$$Var(X) = \frac{\frac{19}{37}}{\left(\frac{18}{37}\right)^2} = 2{,}16975.$$

Im Mittel ist etwas seltener als jedes zweite Spiel rot.

---

7 Multinomialverteilung und allgemeine Hypergeometrische Verteilung

Anwendung bei mehrdimensionalen Ereignissen A_1, A_2, ..., A_k

Modelle mit Zurücklegen: Multinomialverteilung

Wenn k mögliche Ereignisse A_1, A_2, ..., A_k des einzelnen Versuchs existieren – Voraussetzung: Unabhängigkeit der Versuche und konstante Eintrittswahrscheinlichkeiten p_1, p_2, ..., p_k der Ereignisse - ist die Wahrscheinlichkeit, daß bei n Versuchen genau x_1 mal A_1, x_2-mal A_2, ..., x_k mal A_k eintritt, gegeben durch die **Multinomialverteilung**

$$f_M(x_1, ..., x_k \,|\, n, p_1, ..., p_k) = \frac{n!}{x_1! \cdot x_2! ... x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot ... \, p_k^{x_k}$$

$$mit \sum_{i=1}^{k} x_i = n \quad und \sum_{i=1}^{k} p_i = 1$$

Voraussetzung: Ziehen mit Zurücklegen bzw. unendlich große Grundgesamtheit

$X = (X_1, ..., X_k)$ *ist verteilt* $M(n, p_1, ..., p_k)$

Spezialfall: Binomialverteilung mit k = 2, $x_2 = n - x_1$, $p_2 = 1 - p_1$

Erwartungswert

$$E(X_i) = n \cdot p_i$$

Varianz

$$Var(X_i) = n \cdot p_i \cdot (1 - p_i)$$

Beispiel:

Bei einem Fragebogen wird nach dem Alter der befragten Person in Klassen (0-20, 20-40, 40-60 Jahre, älter als 60 Jahre) gefragt. Der Anteil der Bevölkerung in der i-ten Klasse beträgt p_i (i = 1,...,4) mit $\sum_i p_i = 1$.

Es werden 1000 Personen befragt. Wie groß ist die Wahrscheinlichkeit, daß höchstens 10% der Befragten bis zu 20 Jahre und höchstens 10% älter als 60 Jahre sind? (Hartung 1982, S. 210)

Ist $x_i^l = 1$ oder 0, je nachdem ob die l-te Person in der i-ten Altersklasse liegt oder nicht, so ist $x^l = \left(x_1^l, ..., x_4^l\right) \sim M\left(1, p_1, \ldots, p_4\right)$.

Die Zufallsvariable $X = \sum_{l=1}^{1000} X^l$ *ist dann* $M\left(1000, p_1, ..., p_4\right)$ *verteilt*.

Gesucht:

$$\begin{aligned} P\left(X_1 \le 100, X_4 \le 100\right) &= P\left(X_1 \le 100, X_2 + X_3 = 1000 - X_1 - X_4, X_4 \le 100\right) \\ &= \sum_{i=0}^{100} \sum_{j=0}^{100} P\left(X_1 = i, X_2 + X_3 = 1000 - i - j, X_4 = j\right) \\ &= \sum_{i=0}^{100} \sum_{j=0}^{100} \frac{1000!}{i! \, j!} \cdot p_1^i \cdot p_4^i \left(p_2 + p_3\right)^{1000-i-j} \end{aligned}$$

Modelle ohne Zurücklegen: Allgemeine Hypergeometrische Verteilung

Werden Stichproben ohne Zurücklegen gezogen, dann ist die Binomialverteilung durch die Allgemeine Hypergeometrische Verteilung zu ersetzen. Werden z.B. fünf Kugeln aus einer Urne mit $N_1 = 5$ weißen und $N_2 = 10$ schwarzen Kugeln gezogen, dann beträgt die Wahrscheinlichkeit für ($x_1 = 2$ aus fünf weißen und $x_2 = 3$ aus zehn schwarzen Kugeln) $(n = x_1 + x_2, N_1 + N_2 = N)$

$$f_{AH}\left(x_1, x_2 \mid n, N_1, N_2\right) = \frac{\binom{N_1}{x_1}\cdot\binom{N_2}{x_2}}{\binom{N}{n}} = \frac{\binom{N_1}{x_1}\cdot\binom{N-N_1}{n-x_1}}{\binom{N}{n}}$$

$$= \frac{\binom{5}{2}\cdot\binom{10}{3}}{\binom{15}{5}} = \frac{\left(\frac{5!}{2!\cdot 3!}\right)\cdot\left(\frac{10!}{3!\cdot 7!}\right)}{\left(\frac{15!}{5!\cdot 10!}\right)}$$

$$= 0{,}3996$$

Bei kleineren $\frac{n}{N}$ entspricht diese Verteilung der Binomialverteilung.

Allgemein gilt für $x_1, \ldots, x_k$ Beobachtungen der Ereignisse $A_1, \ldots, A_k$:

$$f_{AH}\left(x_1, ..., x_k \mid n, N_1, N_2, ..., N_k\right) = \frac{\binom{N_1}{x_1}\cdot\binom{N_2}{x_2}\cdots\binom{N_k}{x_k}}{\binom{N}{n}}$$

$$mit \sum_{i=1}^{k} x_i = n \quad und \sum_{i=1}^{k} N_i = N$$

$X = (X_1,...X_k)$ ist verteilt AH (n, $N_1,...,N_k$)

Erwartungswert

$$E\left(X_i\right) = n\cdot\frac{N_i}{N}$$

Varianz

$$Var\left(X_i\right) = n\cdot\frac{N_i}{N}\cdot\left(1-\frac{N_i}{N}\right)\cdot\frac{N-n}{N-1}$$

Beispiele:

a) ET: für $k = 2$ $\quad \left(N = P\left(ET\right); N_1 = m\left(ET\right), \frac{N_1}{N} = \pi\left(ET\right)\right)$

b) Zahlenlotto (Urnenmodell ohne Zurücklegen):

z. B. '6 aus 49', Wahrscheinlichkeit für fünf Richtige:

$x_1 = 5, N_1 = 6; x_2 = 1, N_2 = N - N_1 = 49 - 6 = 43, n = x_1 + x_2 = 6$

$$f_{AH}\left(5 \text{ von } 6, 1 \text{ von } 43\right) = f_{AH}\left(5,1 \mid 6,6,43\right)$$

$$= \frac{\binom{6}{5}\cdot\binom{43}{1}}{\binom{49}{6}} = \frac{\binom{6}{5}\cdot\binom{49-6}{6-5}}{\binom{49}{6}}$$

$$= \frac{6\cdot 43}{13983816} = 0,0000184499 = 0,18\cdot 10^{-4}$$

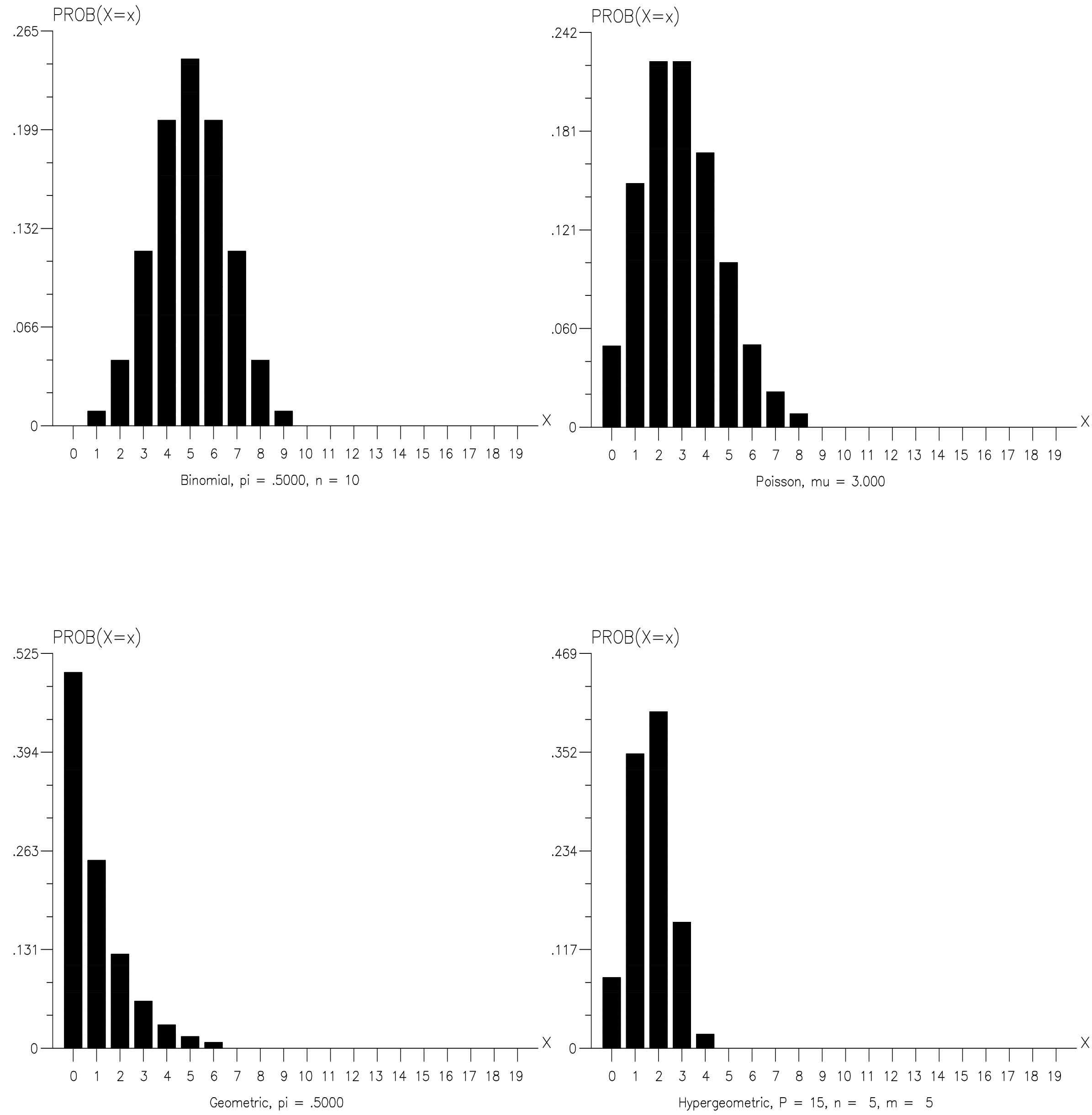
Wahrscheinlichkeit für sechs Richtige:

$$f_{AH}\left(6,0 \mid 6,6,43\right) = \frac{\binom{6}{6}\cdot\binom{43}{0}}{\binom{49}{6}} = \frac{1\cdot 1}{13983816}$$

$$= 0,0000000715 = 0,72\cdot 10^{-7}$$

c) $k > 2$: rote, grüne, schwarze Kugeln $\left(A_1, A_2, A_3\right)$, unterschiedliche Parteien etc.

ET: Discrete Distributions



Source: W. Greene, Econometrics Toolkit, Version 3.0, Econometric Software, Inc. New York 1992, S. 56 ff.

Keyconcepts

Diskrete Merkmale/Verteilungen
Stetige Merkmale/Verteilungen
Parameter von Wahrscheinlichkeitsverteilungen
Bernoulli Experiment
Binomial-, Hypergeometrische-, Poisson-, Geometrische-Verteilung
Multinomialverteilung, Allgemeine Hypergeometrische Verteilung

IV Stetige Verteilungen

Darstellung der Wahrscheinlichkeiten des Auftretens bestimmter Ausprägungen einer stetigen Zufallsvariablen

Stetige Verteilung auf der Basis **stetiger Zufallsvariablen**.

Von besonderem Interesse:

- Dichtefunktion (Wahrscheinlichkeitsdichte)
- Verteilungsfunktion ('cumulative distribution function', cdf) und zentrale Parameter wie Erwartungswert und Varianz

1 Gleichverteilung

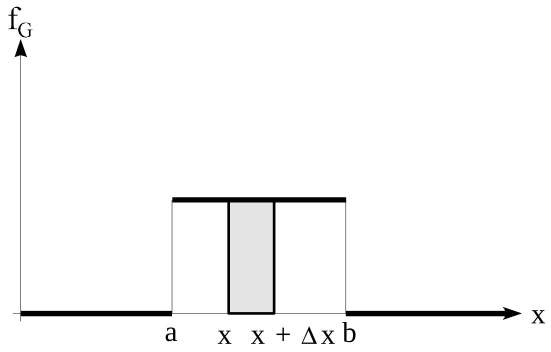
Gleichverteilung: Jeder Wert der stetigen Zufallsvariablen X hat die gleiche Eintrittswahrscheinlichkeit.

Dichtefunktion der Gleichverteilung

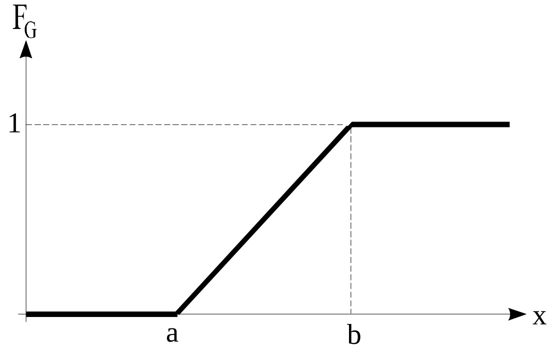
$$f_g\left(x \middle| a,b\right)=\begin{cases}\dfrac{1}{b-a} & \textit{für } a\leq x\leq b\\ 0 & \textit{sonst}\end{cases}$$

Verteilungsfunktion

$$F_g\left(x \middle| a,b\right)=\begin{cases}0 & \textit{für } x< a\\ \dfrac{x-a}{b-a} & \textit{für } a\leq x\leq b\\ 1 & \textit{für } x>b\end{cases}$$



Dichtefunktion



Verteilungsfunktion

Erwartungswert

$$E(X)=\frac{a+b}{2}$$

Varianz

$$Var(X)=\frac{(b-a)^2}{12}$$

Beispiele:

a) Zeigen Sie, daß sich die Verteilungsfunktion einer gleichverteilten Zufallsvariablen mit F_g ergibt?

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_{a}^{x}\frac{1}{b-a}\,dt=\left[\frac{t}{b-a}\right]_a^x=\frac{x}{b-a}-\frac{a}{b-a}=\frac{x-a}{b-a}$$

b) Warum haben E(X) und Var(X) einer gleichverteilten Zufallsvariaben X obiges Ergebnis?

$$E(X)=\int_{-\infty}^{+\infty} f(x)\,x\,dx=\int_a^b f(x)\,x\,dx=\int_a^b\frac{1}{b-a}\,x\,dx=\left[\frac{1}{b-a}\cdot\frac{x^2}{2}\right]_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{(b+a)(b-a)}{2(b-a)}=\frac{b+a}{2}=\mu$$

$$Var(X)=\int_{-\infty}^{+\infty}(x-E(X))^2\,f(x)\,dx=\int_{-\infty}^{+\infty}x^2\,f(x)\,dx-\mu^2$$

$$=\int_a^b\frac{1}{b-a}x^2\,dx-\left(\frac{b+a}{2}\right)^2=\frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b-\left(\frac{b+a}{2}\right)^2$$

$$=\frac{b^3-a^3}{3(b-a)}-\frac{(b+a)^2}{4}=\frac{4(b^3-a^3)-3(b-a)(b+a)^2}{12(b-a)}$$

$$=\frac{(4b^3-4a^3-3ba^2-6ab^2-3b^3+3a^3+6a^2b+3ab^2)}{12(b-a)}$$

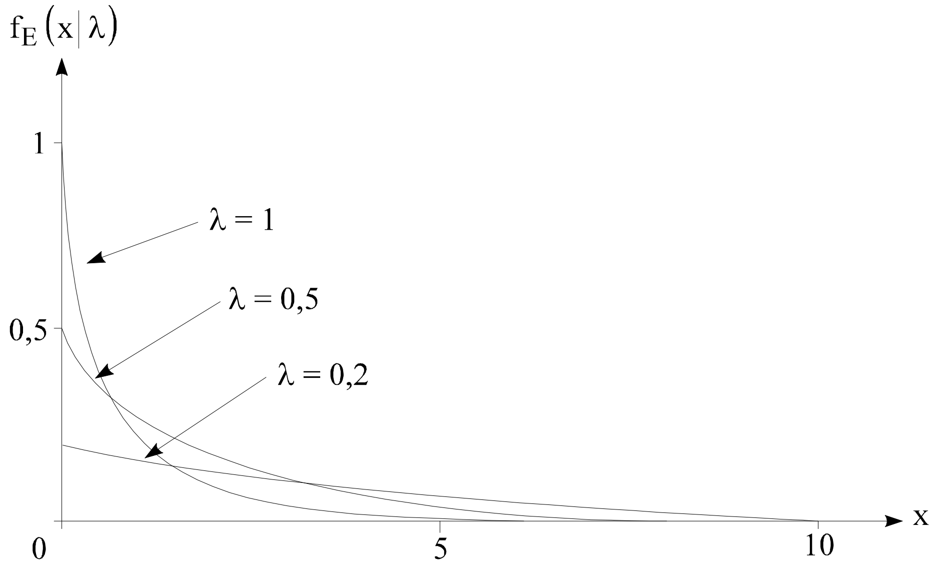
$$=\frac{(b^3-a^3+3a^2b-3ab^2)}{12(b-a)}=\frac{(b-a)(b-a)^2}{12(b-a)}=\frac{(b-a)^2}{12}\qquad q.e.d.$$

2 Exponentialverteilung

Dichtefunktion einer exponentialverteilten Zufallsvariable

$$f_E(x|\lambda)=\begin{cases}\lambda e^{-\lambda x} & \textit{für } x\geq 0 \;\; \textit{mit } \lambda>0\\ 0 & \textit{sonst}\end{cases}$$

Die Exponentialverteilung ist nur von λ abhängig!



Dichtefunktion der Exponentialverteilung für λ = 0.2; 0,5; 1

Die Verteilungsfunktion (cdf) P (X ≤ x) erhält man durch Integrieren:

$$F_E(x|\lambda)=\begin{cases}0 & \textit{für } x<0\\ 1-e^{-\lambda x} & \textit{für } x\geq 0\end{cases}$$

Erwartungswert

$$E(X)=\frac{1}{\lambda}$$

Varianz

$$Var(X)=\frac{1}{\lambda^2}$$

3 Gammaverteilung

Die Gamma (Γ)-Verteilung wird bspw. verwendet zur Analyse von Einkommensverteilungen und Produktionsfunktionen.

Dichtefunktion der Gammaverteilung

$$f_G = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1}$$

wobei $\Gamma(r) = \int_0^\infty t^r e^{-t} dt, \quad \Gamma(r) = (r-1)!$ (r ganzzahlig) Gammafunktion heißt.

Viele bekannte Verteilungen sind Spezialfälle wie z. B.:

- Exponentialverteilung (r = 1)
- Chi-Quadratverteilung ($\lambda = \frac{1}{2}$; $r = \frac{n}{2}$)

Erwartungswert

$$E(X) = \frac{r}{\lambda}$$

Varianz

$$Var(X) = \frac{r}{\lambda^2}$$

4 Normalverteilung

Bedeutung der Normalverteilung:

- als **empirische Verteilung**:
  Sozioökonomische Faktoren verhalten sich annähernd normalverteilt, (QUETELET, 19. Jhdt);
- als **Verteilungsmodell im Zusammenhang mit dem zentralen Grenzwertsatz**:
  In einer Urne befinden sich viele Kugeln mit unterschiedlichem Gewicht. Man zieht nun Stichproben mit Zurücklegen. Bestimmt man dann das durchschnittliche Gewicht der Kugeln in den Stichproben, also das arithmetische Mittel, so werden diese Mittelwerte für jede Stichprobe verschieden sein. In diesem Sinne stellen die Mittelwerte also eine Zufallsvariable dar, die bei genügend großem Stichprobenumfang normalverteilt ist. Dies ist davon unabhängig, wie die Gewichte aller Kugeln in der Urne verteilt sind.
- als **mathematische Basisverteilung**:
  Die Normalverteilung ist die Basisverteilung der χ^2-, der t- und der F-Verteilung (Vgl. auch Abschnitt V.3). Die Normalverteilung liefert häufig auch gute Approximationen für andere Verteilungen (Bsp. Binomialverteilung).
- als **statistische Fehlertheorie**:
  Messungen von ökonomischen Variablen werden häufig von Zufallsfehlern überlagert, wobei oft viele Fehlerfaktoren existieren. Man will zum Beispiel x_i messen, kann aber nur y_i beobachten, wobei der folgende Zusammenhang gilt: $y_i = x_i + \varepsilon_i$. Wenn viele Wiederholungsmessungen durchgeführt werden, kann man für ε_i eine Standardnormalverteilung an-

nehmen ($\varepsilon_i \sim N(0,\ 1)$). Gemäß der Beziehung $y_i = x_i + \varepsilon_i$ läßt sich ε_i bestimmen: $\varepsilon_i = y_i - x_i$.

Die Normalverteilung ist daher die wichtigste statistische Verteilung.

Grundlegende Arbeiten stammen von GAUß aus den Jahren 1809 und 1816 (Gaußsche Glokken- oder Fehlerkurve).

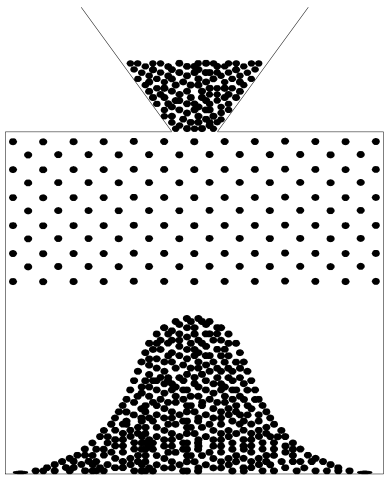
Dichtefunktion (symmetrisch) der Normalverteilung

$$f_N(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \text{für} -\infty < x < +\infty;$$

$$\sigma > 0;$$
$$e = 2,7182....;$$
$$\pi = 3,1415\ldots;$$

X heißt normalverteilt mit den Parametern μ und σ^2 : $X \sim N(\mu, \sigma^2)$

Eigenschaften: symmetrisch, glockenförmig,
Modus = Median = Erwartungswert
Maximum bei: $x = \mu$
Wendepunkte: $x = \mu \pm \sigma$
zwischen den Wendepunkten: ca. 2/3 der Fläche



Nagelbrett und Normalverteilung

Nagelbrett und Normalverteilung

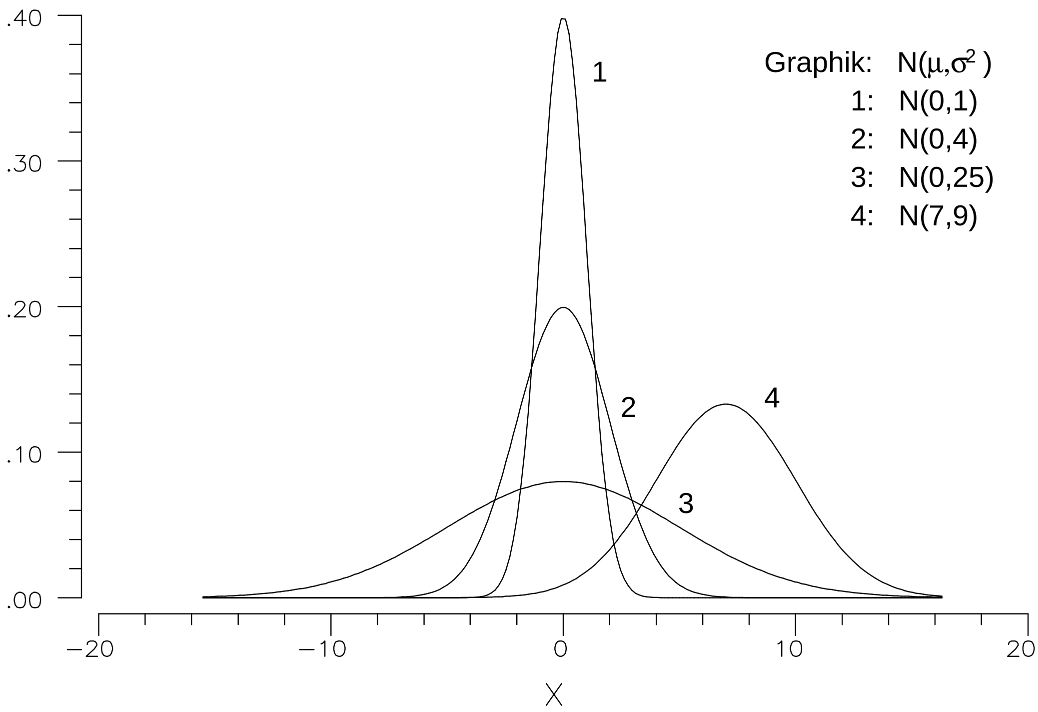
Nägel bewirken, daß die Kugeln mit p = 0,5 rechts oder links fallen. Für großes n nähert sich also die Binominalverteilung an die Normalverteilung (Glockenkurve).

Graphisches Beispiel:

Computerprogramm Econometrics Toolkit ET:

Befehlssequenz: PROB → Normal → Graphic

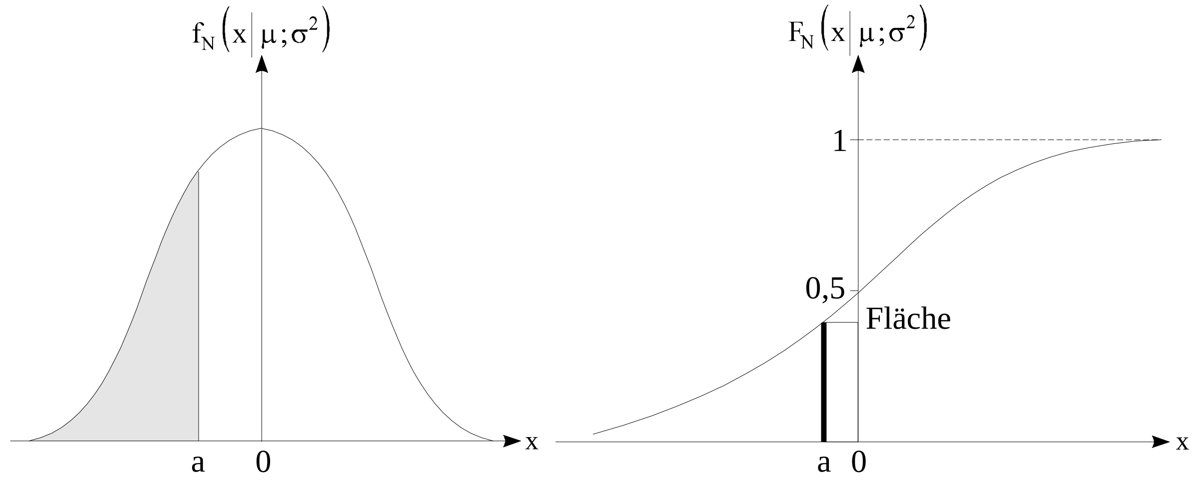
Mehrere Grafiken: Multiple plots



Dichtefunktionen von Normalverteilungen mit verschiedenen Parametern μ und σ^2

Verteilungsfunktion (nicht mehr elementar)

$$F_N(x \mid \mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$



Dichtefunktion Verteilungsfunktion

Wichtige Transformationseigenschaft

Ist X normalverteilt, dann ist auch jede lineare Transformation von X normalverteilt:

$$\mathrm{X} \sim N\left(\mu,\sigma^2\right) \quad \rightarrow \quad a+b\ \mathrm{X} \sim N\left(a+b\,\mu, b^2\,\sigma^2\right)$$

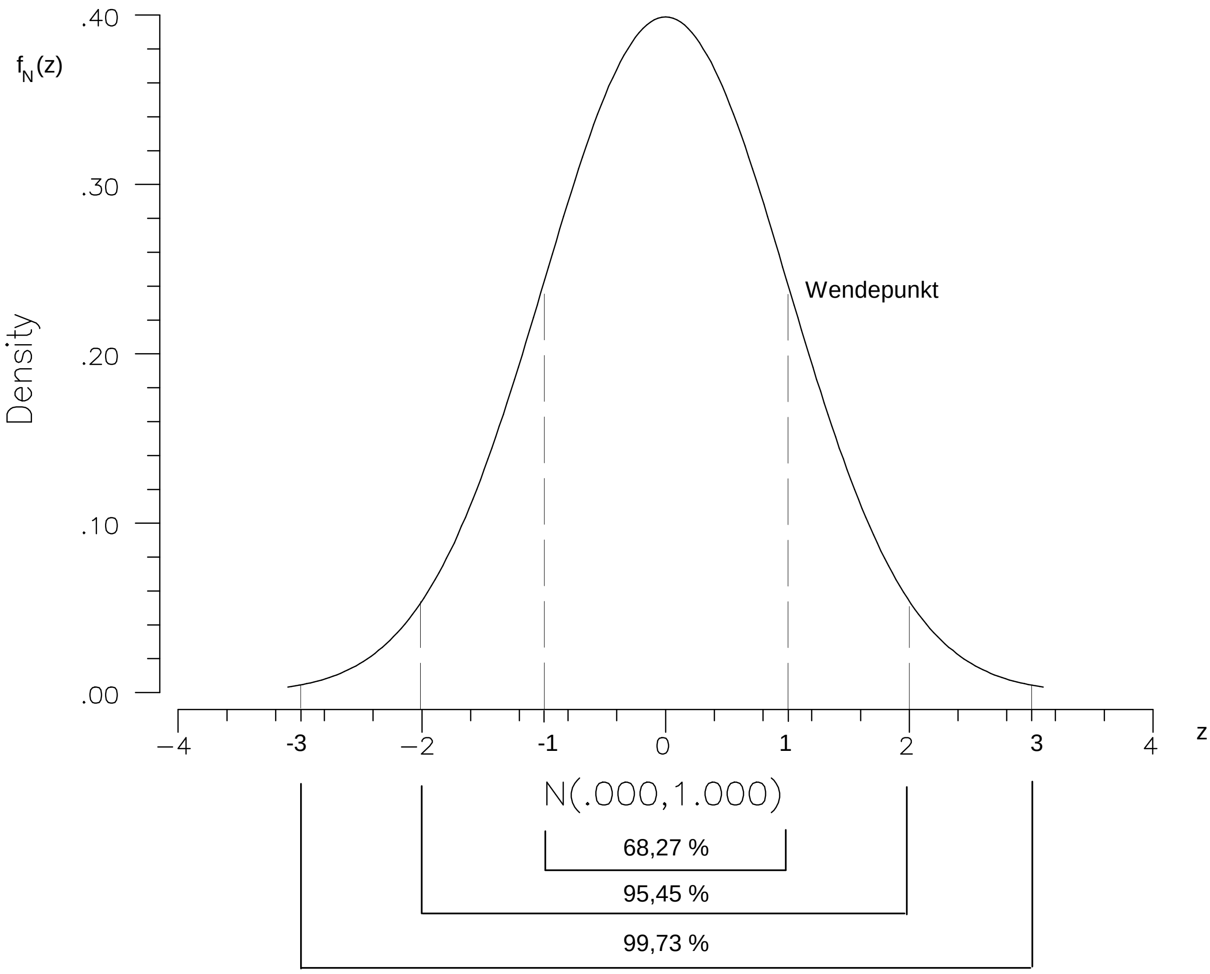
Eine besondere Transformation:

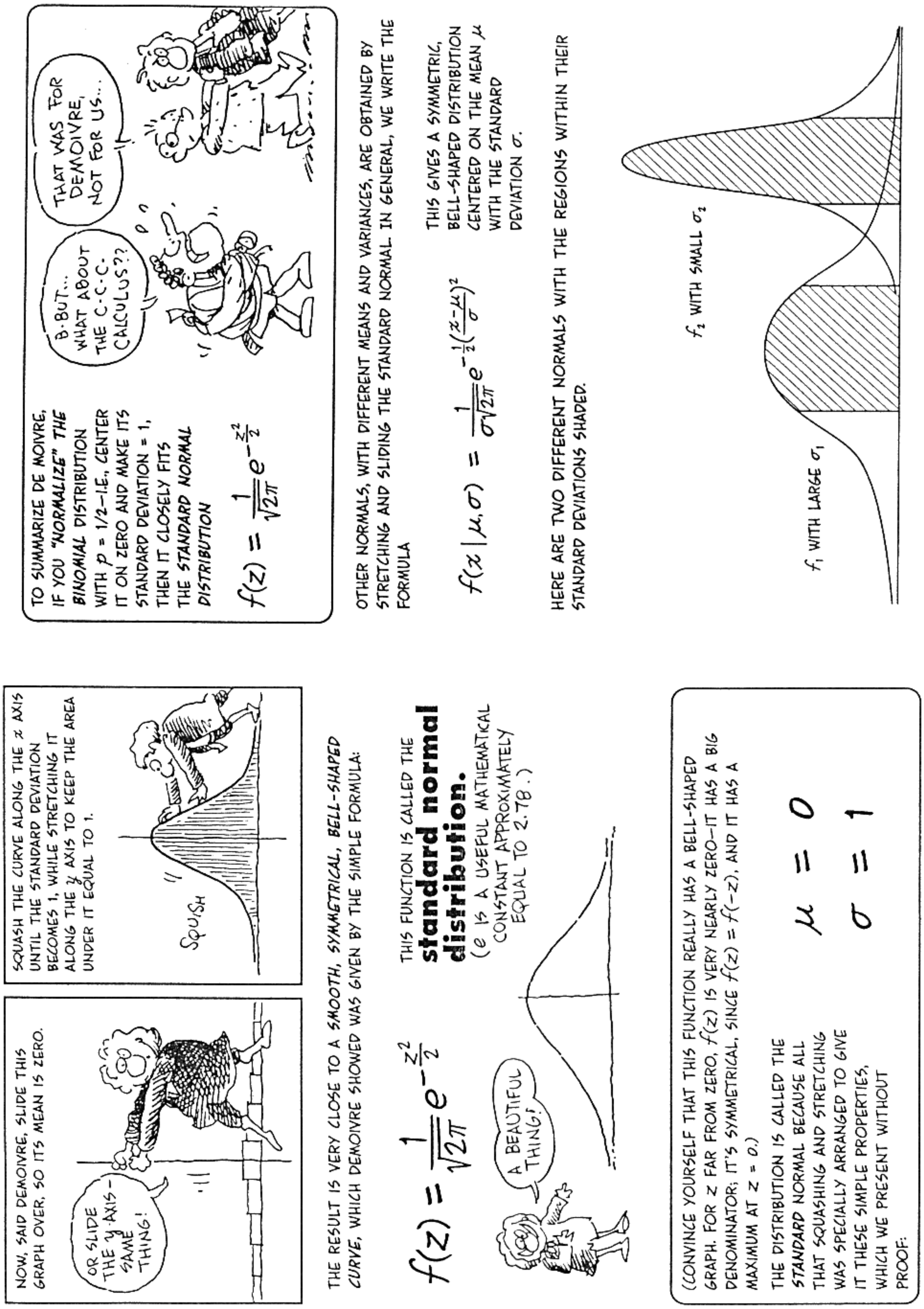
$$a=\frac{-\mu}{\sigma} \quad ; \quad b=\frac{1}{\sigma} \quad \Rightarrow \quad Z=\frac{X-\mu}{\sigma}$$

Z ist standardnormalverteilt mit $\mu = 0$, $\sigma^2 = 1$; N(0, 1)

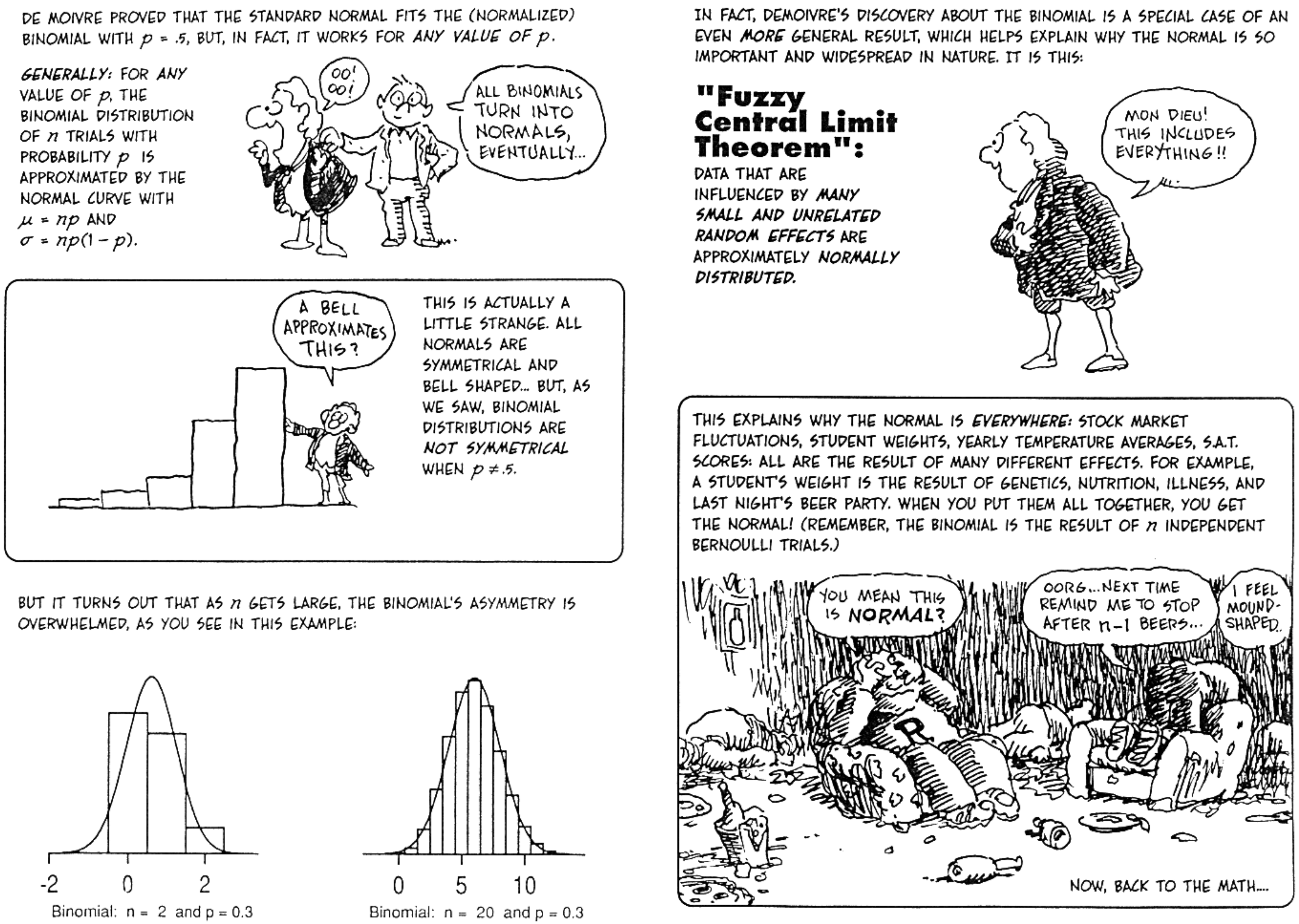
Dichtefunktion Standardnormalverteilung Z ~ N(0,1)

$$\phi(z)=f_N(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$





Gonick, Smith 1993



Gonick, Smith 1993

Verteilungsfunktion Standardnormalverteilung

$$\Phi(Z) = F_N(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

Tabellenwerte:

Es gilt: $\text{Prob}(a < X < b) = \text{Prob}\left(\frac{\text{a - }\mu}{\sigma} < \frac{\text{X - }\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$

$-$ *also* $F_N(x) = F_N(z) \qquad z = \dfrac{x-\mu}{\sigma}$

$-\ \Phi(-z) = 1 - \Phi(z)$ *Symmetrie*

$-\ f(-z) = f(z)$

$-\ z_p = -z_{1-p}$

Beispiele:

Der Durchmesser von Brettern sei normalverteilt mit μ = 1,2 cm und σ^2 = 0,0016 cm².

a) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser eines zufällig ausgewählten Brettes größer als 1,28 cm ist?

$$P(X > 1{,}28) = 1 - P(X \le 1{,}28) = 1 - P(Z \le z),$$

$$\textit{wobei } z = \frac{x-\mu}{\sigma} = \frac{1{,}28 - 1{,}2}{0{,}04} = 2$$

$$\begin{aligned} P(X > 1{,}28) &= 1 - P(Z \le 2) = 1 - F_N(2) \\ &= 1 - 0{,}9773 \\ &= 0{,}0227 \end{aligned}$$

über ET oder Tabelle

$$\textit{Tabelle}: \ F_Z, F_1, F_2 \Rightarrow F_Z = F(z=2) = 0{,}9773$$

b) Wie groß ist die Wahrscheinlichkeit, daß der Durchmesser zwischen 1,18 und 1,26 cm liegt?

$$P(1{,}18 \le X \le 1{,}26) = P(z_u \le Z \le z_o);$$

$$z_u = \frac{1{,}18 - 1{,}2}{0{,}04} = -0{,}5 \qquad z_o = \frac{1{,}26 - 1{,}2}{0{,}04} = 1{,}5$$

$$\begin{aligned} P(1{,}18 \le X \le 1{,}26) &= P(-0{,}5 \le Z \le 1{,}5) \\ &= F_N(1{,}5) - F_N(-0{,}5) \\ &= F_N(1{,}5) - (1 - F_N(0{,}5)) \\ &= 0{,}9332 - 0{,}3085 \\ &= 0{,}6247 \end{aligned}$$

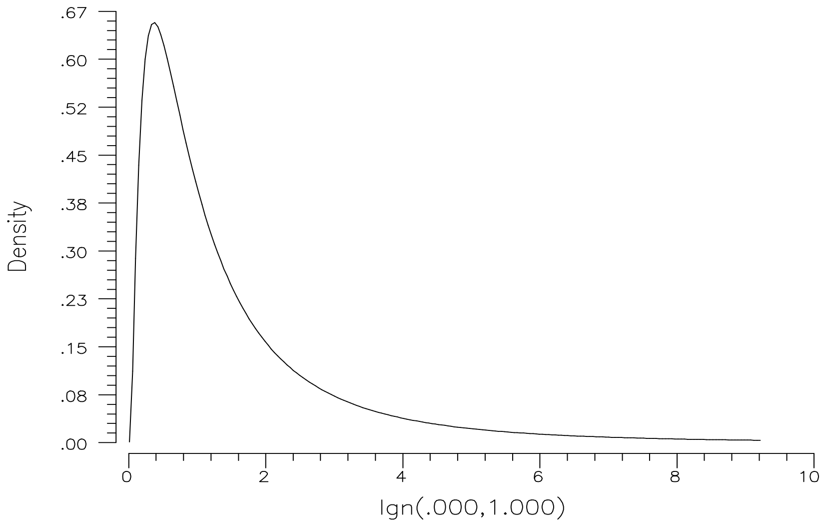
Eine weitere geläufige Transformation der Normalverteilung führt zur Lognormalverteilung. Wenn der Logarithmus von X normalverteilt ist, ist X selbst lognormalverteilt:

$$\ln X \sim N(\mu,\sigma) \to X \sim LN(\mu,\sigma)$$

Die entstehende linkssteile Verteilung kann recht gut zur Beschreibung einer Vielzahl empirischer Verteilungen, wie bspw. Vermögens-, Einkommens-, aber auch Firmengrößen verteilungen, verwendet werden**.**

Dichtefunktion Lognormalverteilung X~LN

$$f_{LN}(x|\ \mu,\sigma^2)=\frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}} \qquad \textit{für } x>0;\sigma>0;e=2,7182....;\ \pi=3{,}1415\ldots;$$



Dichtefunktion der Exponentialverteilung für μ = 0, σ = 1

Verteilungsfunktion (cdf) P (X ≤ x)

$$F_{LN}(X)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^x \frac{e^{\frac{-(\ln t-\mu)^2}{2\sigma^2}}}{t}dt$$

Erwartungswert

$$E(\ln X)=\mu \to E(X)=e^{(\mu+\frac{\sigma^2}{2})}$$

Varianz

$$Var(\ln X)=\sigma^2 \to Var(X)=e^{(2\mu+\sigma^2)}*(e^{\sigma^2}-1)$$

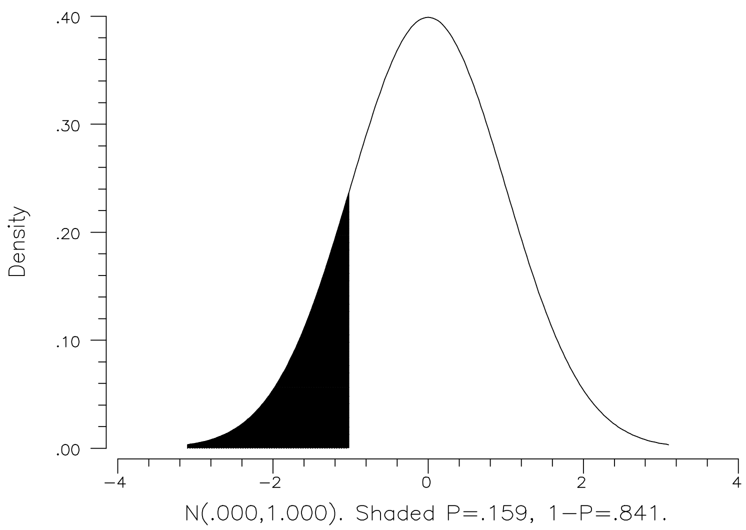
ET: Probability, Continuous Distributions, Normal Distribution

```
Continuous
6 Normal
7 t
8 F
9 Chi squared
0 Lognormal
A Beta
```

```
Mean, μ                    .000
Standard deviation, σ  1.000    σ < 0

   Mean=  0.000, Std. Deviation= 1.000
```

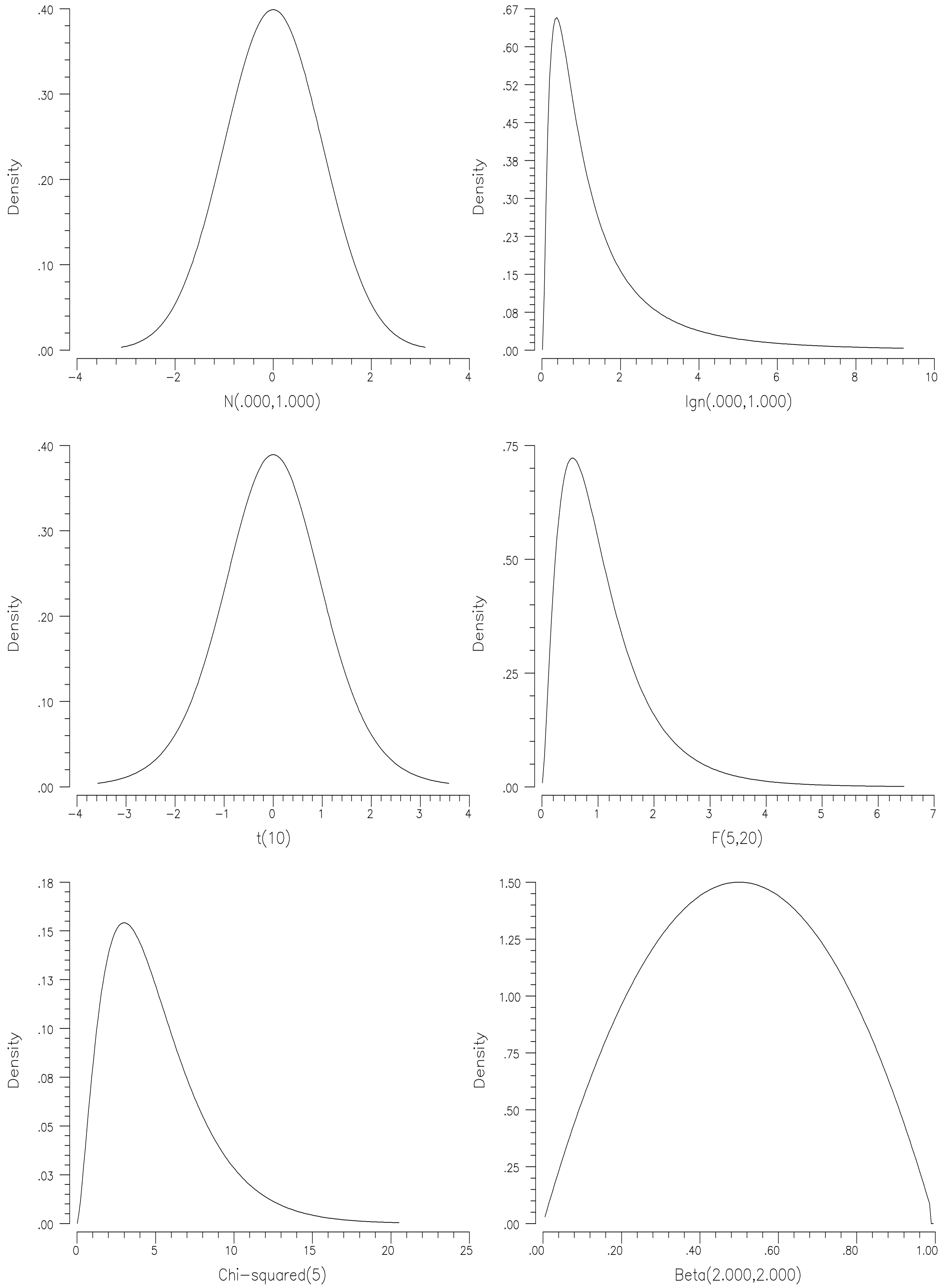
```
Set parameters
Probabilities
Graph distn.
Multiple plots

Use ↑↓↵ or ltr.
ESC=exit to menu
```

```
Move the cursor to the computation you want with ↑
or ↓, enter the value, then press Enter.  Esc=Exit
Enter P,C,R,or,I for the type of computation
Prob   P[X=x]   what is x?        P=        1-P=
Cumul  P[X≤x]   what is x? -1     P= .159   1-P= .841
Range  P[a≤X≤b] what is a?        b?          P=
Invrs  x for P[X≤x]=P:   what is P?         X=
                 Plot? (y/n) Y
```



Source: W. Greene, Econometrics Toolkit, Version 3.0, Econometric Software, Inc. New York 1992, S. 62 ff.

ET: Continuous Distributions



Source: W. Greene, Econometrics Toolkit, Version 3.0, Econometric Software, Inc. New York 1992, S. 62 ff.

5 Normalverteilung als Näherungsverteilung

Die Normalverteilung dient als Näherungsverteilung für:

- **Binomialverteilung**:

  Die Approximation ist um so besser, je näher p bei 0,5 liegt.

  Faustregel: $n \cdot p \cdot (1-p) \geq 9$

  Als standardisierte Zufallsvariable erhält man hier:

  $Z^* = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot q}}$ ($n \cdot p = E(X)$ der Binomialverteilung; $n \cdot p \cdot (1-p) = Var(X)$ der Binomialverteilung; wobei $q = (1-p)$)

- **Poissonverteilung**:

  Die Approximation kann verwendet werden, wenn $\mu \to \infty$

  Faustregel: $\mu \geq 9$

  Als standardisierte Zufallsvariable erhält man hier:

  $\hat{Z} = \frac{X - \mu}{\sqrt{\mu}}$ ($\mu = E(X)$ der Poissonverteilung; $\mu = Var(X)$ der Poissonverteilung)

- **Hypergeometrische Verteilung**:

  Die Approximation kann verwendet werden, wenn

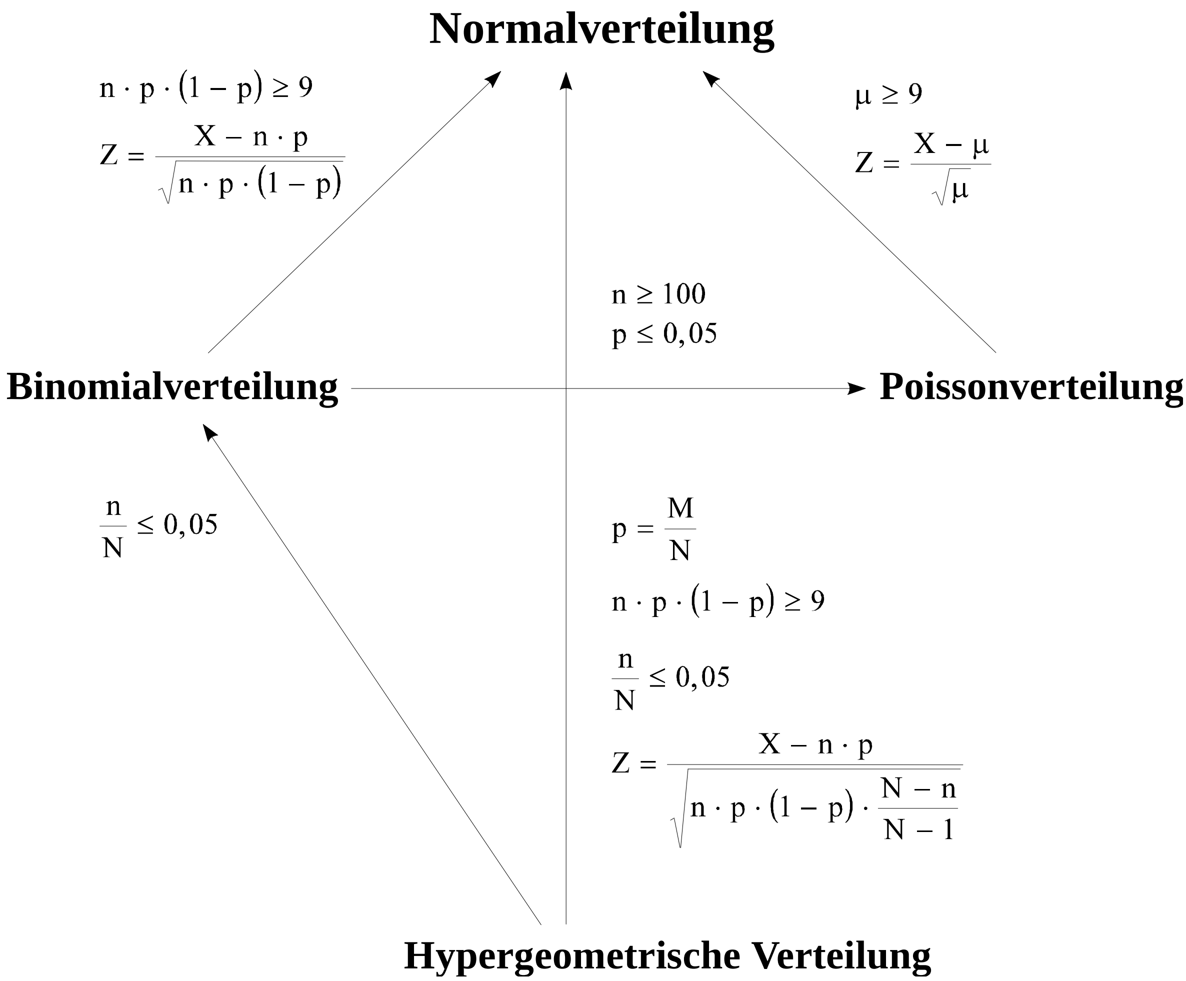
  $n \cdot p \cdot (1-p) \geq 9$ mit $p = \frac{M}{N}$ und $\frac{n}{N} \leq 0{,}05$.

Weitere in enger Beziehung zur Normalverteilung stehende Verteilungen:

- gestutzte ('tuncated') Normalverteilung
- Lognormalverteilung
- Chi[2] (χ^2) - Verteilung
- t - Verteilung
- F - Verteilung

siehe unter V.3

Übergänge zwischen den Verteilungen



Keyconcepts

Gleichverteilung
Exponentialverteilung
Gammaverteilung
Normalverteilung
Standardisierung einer normalverteilten Zufallsvariablen
Standardnormalverteilung
Lognormalverteilung
Normalverteilung als Näherungsverteilung
Übergänge zwischen den Verteilungen

V Induktive Statistik, Stichprobenfunktionen und Testverteilungen

Der Schluss auf die Grundgesamtheit: Berechnungen aus der Stichprobe und Einführung in die Testverteilungen

1 Zum Schluß von der Stichprobe auf die Grundgesamtheit

Die induktive Statistik **(auch: Inferenzstatistik, schließende Statistik)** hat das Ziel, von einer Auswahl (Stichprobe) auf die Allgemeinheit (Grundgesamtheit, Population) zu schließen.

Beispiele:

Aus einer Fertigung von CD-ROM-Laufwerken werden zufällig einige zur Qualitätskontrolle herausgeholt. Was kann mit dem Fehleranteil dieser Auswahl über die Gesamtproduktion gesagt werden?

Ein weiterer Bereich ist die empirische Überprüfung von allgemeinen Hypothesen an Stichprobenergebnissen:

Eine Untersuchung brachte folgendes Ergebnis zu Tage: 30 von Frauen geschriebene Statistikklausuren wiesen einen um zwei Punkte höheren Durchschnitt als von 40 Männern geschriebene Statistikklausuren auf.

Kann daraus geschlossen werden, daß Frauen generell bessere Statistikklausuren schreiben als Männer?

Wichtige Begriffe:

Population, Grundgesamtheit	=	alle potentiell untersuchbaren Einheiten
Stichprobe	=	Teilmenge der Untersuchungseinheiten
Zufallsstichprobe	=	zufällige Auswahl aus der Grundgesamtheit
Geschichtete Stichprobe	=	Grundgesamtheit wird in homogene Gruppen (Schichten, 'strata') unterteilt und dann Stichprobe aus jeder Schicht gezogen
Klumpenstichprobe	=	Grundgesamtheit wird in kleinere Klumpen ('cluster') aufgeteilt. Jede Beoabachtung von zufällig ausgewählten Klumpen wird betrachtet (z. B. Wohnungsumfrage einer Stadt: Aufteilung der Stadt in Wohnbezirke, zufällige Auswahl der Wohnbezirke, Untersuchung aller Wohnungen in den ausgewählten Bezirken).
proportional geschichtete oder stratifizierte Stichprobe	=	Merkmalsverteilung in der einzelnen Schicht ('strata') $\hat{=}$ Verteilung in der Population (z. B. Konsumverhalten: Anteile Geschlecht, Alter, Haushaltsgröße etc. wie in der Grundgesamtheit)

Ziel der induktiven Statistik: Schluß von der Stichprobe auf die Gesamtheit zu ermöglichen bzw. aus einer Stichprobe Erkentnisse über die Gundgesamtheit zu gewinnen

2 Stichprobenfunktionen

Die n Werte einer Stichprobe $x = (x_1,....,x_n)$ seien durch Zufallsexperimente gewonnen worden. Jeder Stichprobenwert x_i kann dann als Realisation einer Zufallsvariablen X_i (i= 1,...,n) aufgefaßt werden:

Zufallsstichprobe = Realisation der n-dimensionalen Zufallsvariablen $X = (X_1,...,X_n)$

Zur Durchführung eines statistischen Verfahrens wird nun oft nicht die Stichprobe selbst, sondern ein daraus berechneter Funktionswert (z.B. Mittelwert, Streuung,…) verwendet

$$w_n = g_n(x_1, ..., x_n);$$

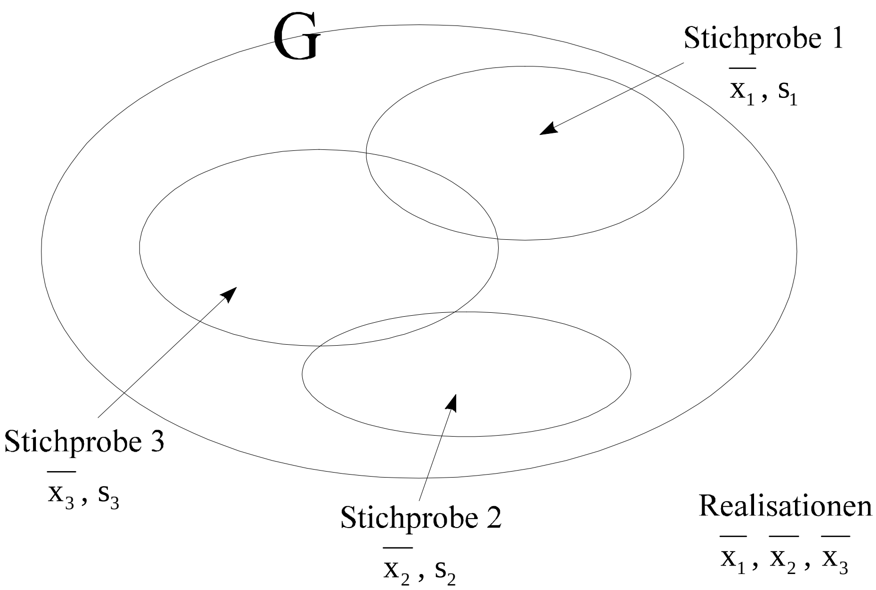
der Funktionswert ist Realisation einer Zufallsvariablen

$$W_n = g_n(X_1, ..., X_n).$$

Die Stichprobenfunktion (W_n) dient als

- Schätzfunktion (zur Schätzung eines Parameters)
- Testfunktion (zur Durchführung eines Tests)

Aus der Grundgesamtheit kann man verschiedene Stichproben mit ihren Stichprobenkennwerten ermitteln:



Stichprobenfunktionen (Stichprobenkennwerte) sind z. B. der Mittelwert ($\overline{x}$) und die Streuung (s).

Ein Beispiel für eine Stichprobenfunktion ist $\overline{X} = g_n(X_1, X_2, X_3,...)$. Die Stichprobenfunktion ist eine Funktion von Zufallsvariablen, die die Information einer Stichprobe zusammenfaßt.

Stichprobenfunktionen haben (da sie Zufallsvariablen sind) bestimmte Verteilungen (Stichprobenverteilungen).

Wichtige Stichprobenfunktionen:

- **arithmetisches Mittel der Stichprobe**

$\overline{\mathrm{X}}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{X}_{\mathrm{i}}$ (großes X, da Zufallsvariable)

Für den **Erwartungswert** von $\overline{\mathrm{X}}$ gilt:

$$E(\bar{X})=E\left[\frac{1}{n}(X_1+X_2+...+X_n)\right]=\frac{1}{n}\cdot\left[E\left(X_1\right)+E\left(X_2\right)+\ldots+E\left(X_n\right)\right]$$

wegen $E(X_1)=\mathrm{E}(\mathrm{X}_2)=...=\mu$ *ist*

$$E\left(\bar{X}\right)=\frac{1}{n}\cdot n\cdot\mu=\mu$$

Für die **Varianz** von ($\overline{\mathrm{X}}$) gilt:

$$Var(\bar{X})=Var\left[\frac{1}{n}\left(X_1+X_2+.....+X_n\right)\right]$$

$$=\frac{1}{n^2}\left[Var\left(X_1\right)+Var\left(X_2\right)+....+Var\left(X_n\right)\right]$$

wegen $Var\left(X_1\right)=Var\left(X_2\right)=...=\sigma^2$ *ist*

$$Var\left(\bar{X}\right)=\frac{1}{n^2}\cdot n\cdot\sigma^2=\frac{\sigma^2}{n}$$

$$\sigma_{\bar{X}}^2=\frac{\sigma^2}{n}$$

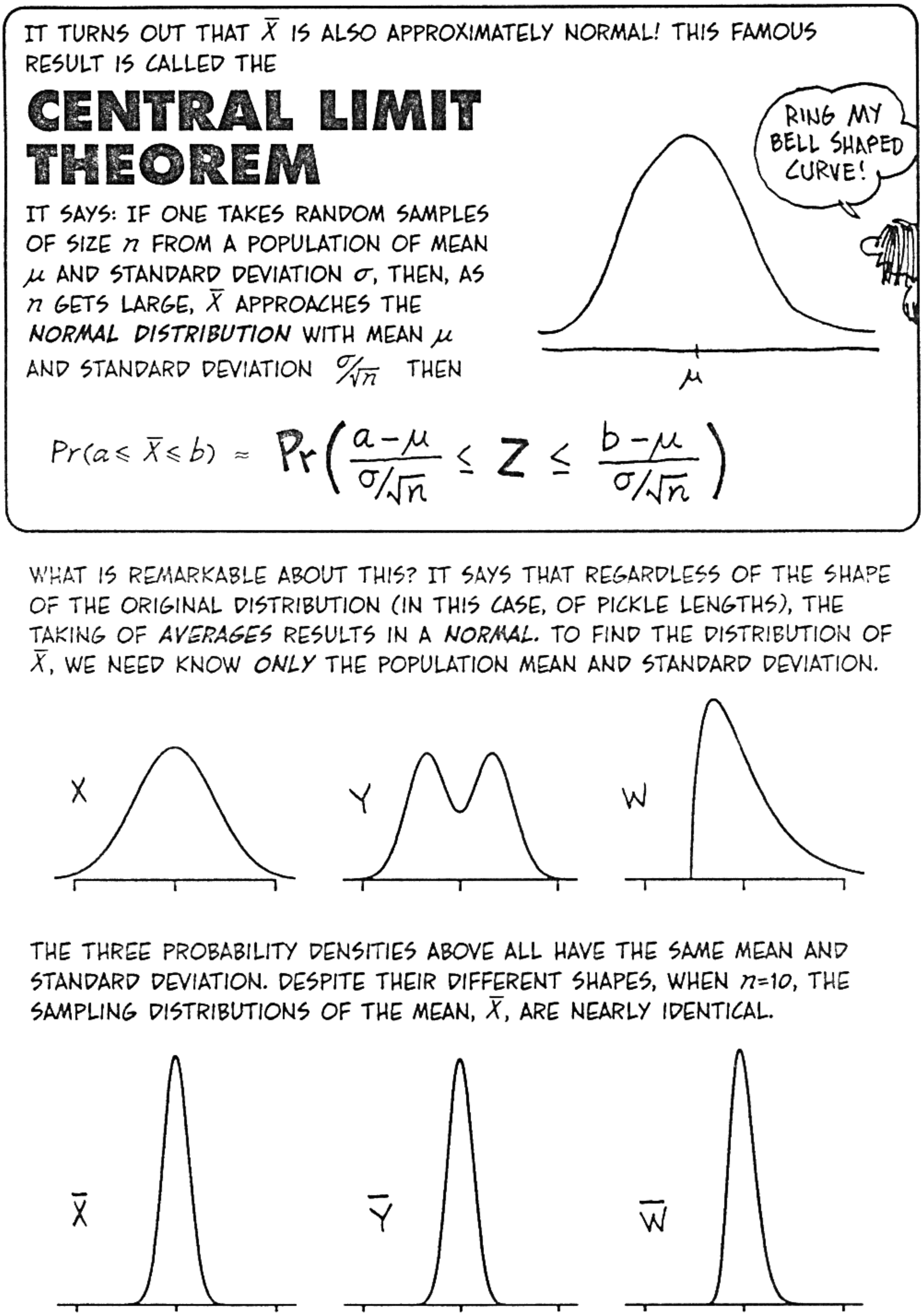
Frage: Welche **Verteilungsform** hat die Zufallsvariable $\overline{\mathrm{X}}$?

Wenn alle X_i normalverteilt mit μ und σ^2 sind, dann ist die Summe $\mathrm{X}_1+\mathrm{X}_2+\mathrm{X}_3+\ldots+\mathrm{X}_\mathrm{n}$ (unabhängige Zufallsvariablen) ebenfalls normalverteilt; damit ist auch $\overline{\mathrm{X}}$ normalverteilt. Falls die Verteilung der X_i unbekannt ist, dann läßt sich aufgrund des **zentralen Grenzwertsatzes** folgende Aussage machen: Die Verteilung von $\overline{\mathrm{X}}$ von n unabhängigen identisch verteilten Zufallsvariablen X_i strebt mit wachsendem Stichprobenumfang n gegen eine Normalverteilung mit $\mathrm{E}(\overline{\mathrm{X}}) = \mu$ und $\mathrm{Var}(\overline{\mathrm{X}}) = \frac{\sigma^2}{\mathrm{n}}$ (Faustregel: $\mathrm{n}>30$) (vgl. Abbildung).

Die standardisierte Zufallsvariable

$$\mathrm{Z}=\frac{\overline{\mathrm{X}}-\mu}{\frac{\sigma}{\sqrt{\mathrm{n}}}}=\frac{\overline{\mathrm{X}}-\mu}{\sigma}\cdot\sqrt{\mathrm{n}}$$

ist unter diesen Bedingungen approximativ standardnormalverteilt.



Gonick, Smith 1993

- **Varianz der Stichprobe**

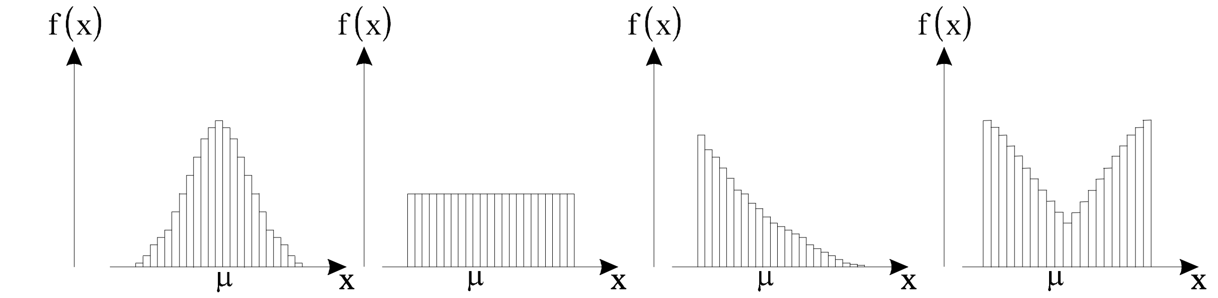
  Die Stichprobenvarianz

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

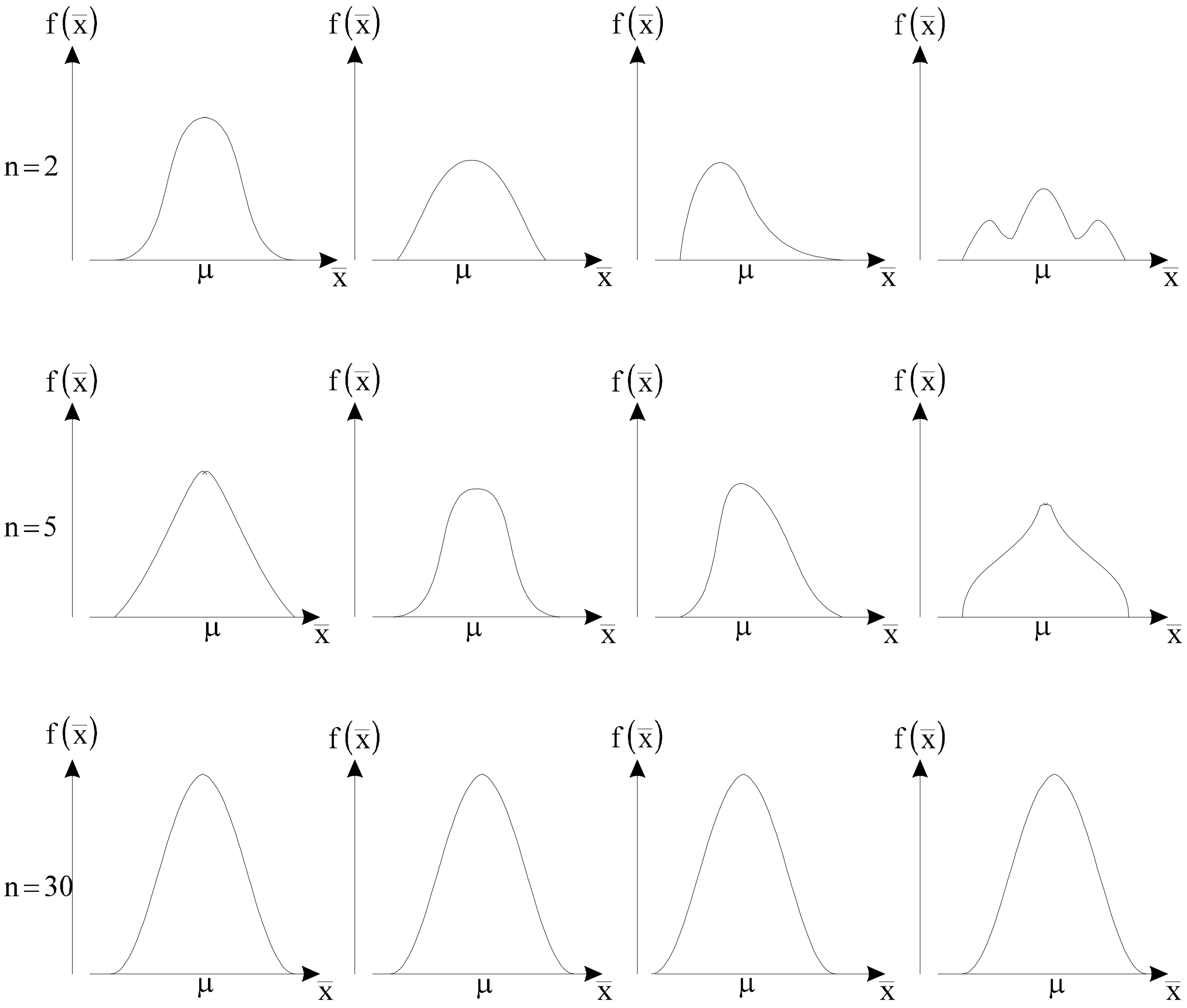
  ist als Funktion von Zufallsvariablen ebenfalls eine Zufallsvariable.

Beispiele zum zentralen Grenzwertsatz:

Verteilung der Grundgesamtheit



Verteilung des Stichprobenmittelwertes $\overline{X}$



3 Testverteilungen

Die Chi-(χ^2)-Quadrat, t-(Student)Verteilung und F-Verteilung sind wichtige Testverteilungen für die noch folgenden Schätzansätze (Inferenzmethoden). Sie finden ihre Anwendung bei Parametertests (Kapitel VIII) und Verteilungstests (Kapitel IX).

Alle Testverteilungen hängen von der Anzahl der Freiheitsgrade (df = degrees of freedom) ab. Die Freiheitsgrade geben die Anzahl der noch freiwählbaren Informationen nach Berücksichtigung gegebener Parameter an.

3.1 Chi (χ^2)-Quadratverteilung

Die Formel geht auf den Astronomen HELMERT (1875) zurück. Die Namensgebung erfolgte durch PEARSON (1900). Die χ^2-Verteilung entspricht folgendem stochastischen Modell:

Sind $Z_1, Z_2, \ldots, Z_\nu$ unabhängig standardnormalverteilte Zufallsvariablen, so hat die Quadratsumme

$$\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_\nu^2$$

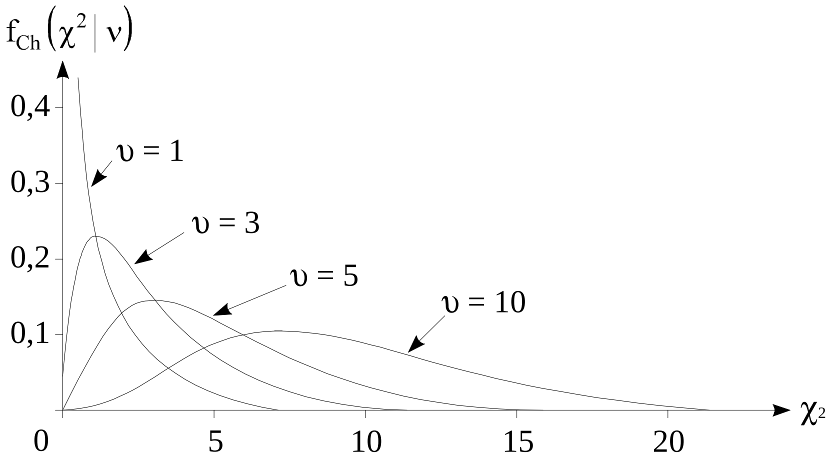
eine χ^2-Verteilung mit ν Freiheitsgraden, (df = 'degrees of freedom'): $\chi^2 \sim \chi^2(\nu)$.

Dichtefunktion der χ^2-Verteilung

$$f_{\chi^2}(z \mid \nu) = C(\nu) \cdot z^{\frac{\nu}{2}-1} \cdot e^{-\frac{z}{2}} \qquad , z > 0$$

$$mit \;\; C(\nu) = \left[2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right) \right]^{-1}$$

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt$$



Dichtefunktion der χ^2-Verteilung für ν = 1, 3, 5, 10 Freiheitsgrade

Die χ^2-Verteilung liegt in Abhängigkeit von ν tabelliert vor.

Erwartungswert

$$E\left(\chi^2\right)=\nu$$

Varianz

$$Var\left(\chi^2\right)=2\nu$$

Aus dem zentralen Grenzwertsatz folgt:

$$\frac{\chi^2-\nu}{\sqrt{2\nu}} \underset{\nu\to\infty}{\to} N(0,1)$$

Aus der **Stichprobenvarianz** entwickelt man eine **χ^2-verteilte Zufallsvariabl**e

$$Z_i=\frac{X_i-\mu}{\sigma} \quad (Z_i \text{ standardnormalverteilt})$$

$$\sum_{i=1}^{n} Z_i^2=\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2 \ ist\ \chi^2-\text{verteilt mit n Freiheitgraden}$$

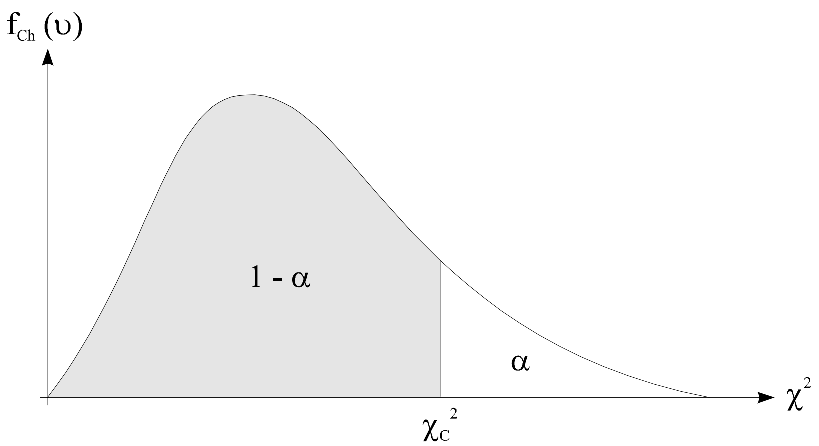
Wird μ durch $\overline{X}$ ersetzt, so verliert man einen Freiheitsgrad, d.h.,

$$\chi^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2=\frac{n\cdot S^2}{\sigma^2} \qquad \left(mit\ S^2=\frac{1}{n}\sum_i (X_i-\overline{X})^2\right)$$

ist χ^2-verteilt mit $(n-1)$ Freiheitsgraden ($\overline{X}$ ist Zufallsvariable).

Tabellenwerte:

Zu den angegebenen Freiheitsgraden ν sind die Werte χ^2 tabelliert, bei denen $P\left(0<\chi^2<\chi_c^2\right)=1-\alpha$ die im Tabellenkopf angegebenen Werte erreicht.



Z.B. ν = 25: $\chi^2=34{,}38$, also ist die Wahrscheinlichkeit $P\left(0<\chi^2\le 34{,}38\right)=0{,}9$.

3.2 Studentverteilung (t-Verteilung)

Von GOSSET 1908 unter dem Pseudonym 'Student' veröffentlicht.

Sind zwei Zufallsvariablen Z und U voneinander unabhängig verteilt mit $Z \sim N(0,1)$ und $U \sim \chi^2(\nu)$, so ist die Zufallsvariable T

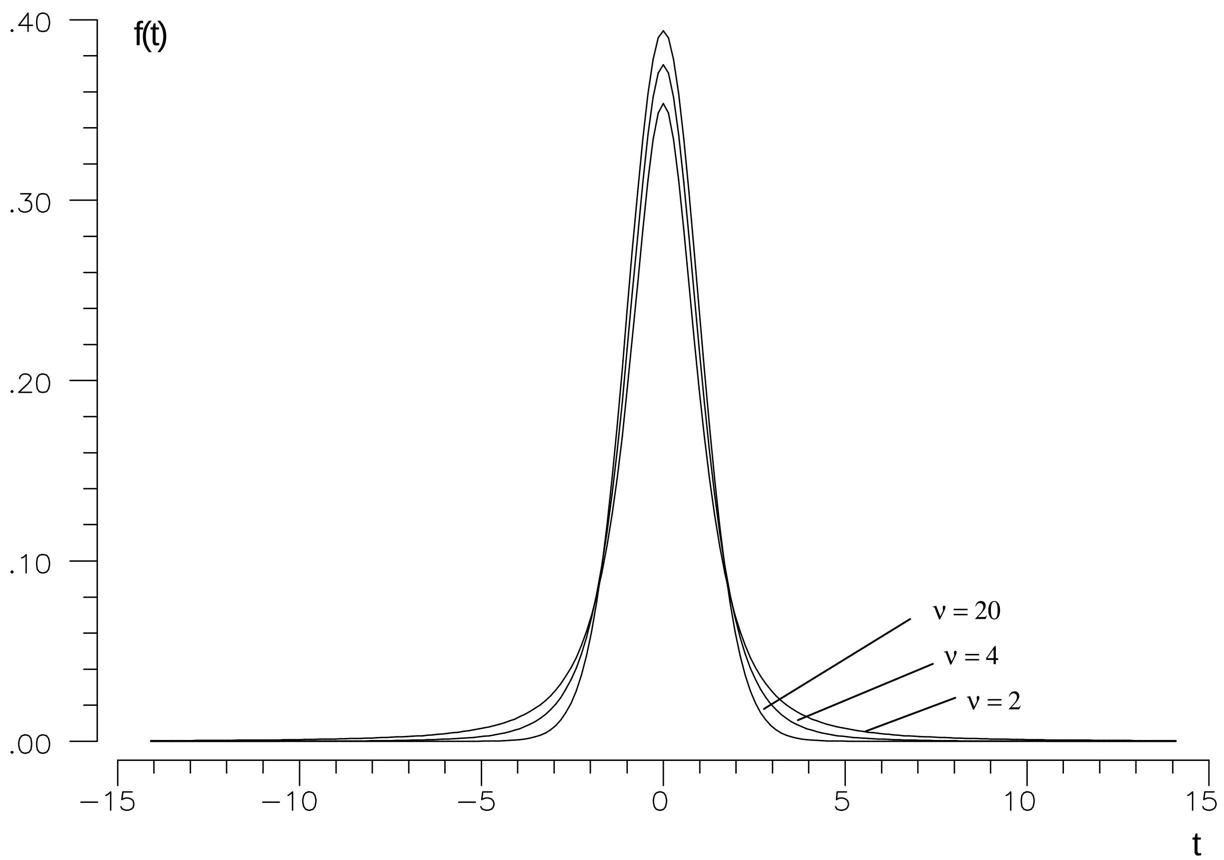
$$T = \frac{Z}{\sqrt{\frac{U}{\nu}}} \qquad t-\text{verteilt mit } \nu \text{ Freiheitsgraden.}$$

Dichtefunktion der t-Verteilung

$$f_t(z|\ \nu) = C(\nu) \cdot \frac{1}{\left(1+\frac{z^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad -\infty \le z \le \infty$$

$$mit\ C(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu \cdot \pi} \cdot \Gamma\left(\frac{\nu}{2}\right)}$$

Die t-Verteilung ist in Abhängigkeit vom Parameter ν tabelliert. Sie ist wichtig für Signifikanztests der Parameter in ökonometrischen Modellen (vgl. VIII. 9).



Dichtefunktionen der t-Verteilung mit ν = 2, 4, 20 Freiheitgraden

Ab ν ≥ 30 kann die t-Verteilung gut durch die Normalverteilung angenähert werden.

Erwartungswert

E(T) = 0

Varianz

$$\text{Var}(T) = \frac{\nu}{\nu - 2} \qquad \text{für } \nu > 2$$

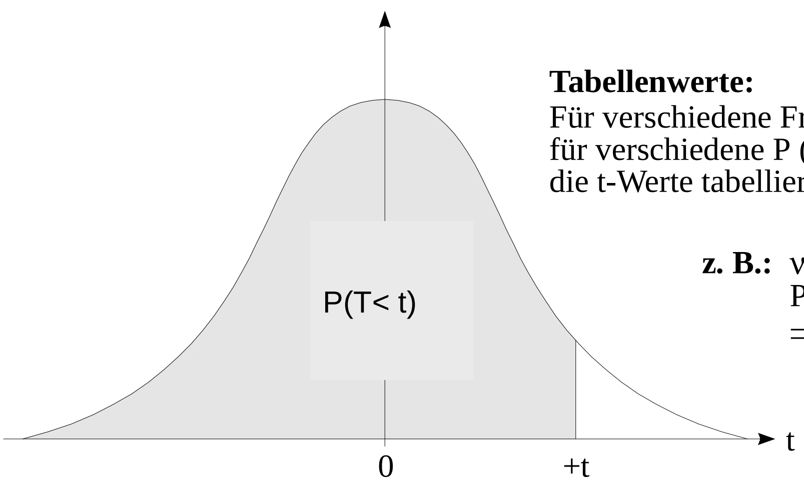
Die Stichprobenfunktion

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n-1}}} = \frac{\bar{X}-\mu}{\sigma} \cdot \sqrt{n} \cdot \frac{\sigma}{\frac{\sqrt{S^2 \cdot n}}{\sqrt{n-1}}} \qquad \left(mit\ S^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2 \right)$$

$$= \frac{\frac{\frac{\bar{X}-\mu}{\sigma}}{\sqrt{n}}}{\frac{\frac{\sqrt{S^2 \cdot n}}{\sqrt{n-1}}}{\sigma}} = \frac{\frac{\frac{\bar{X}-\mu}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{\frac{S^2 \cdot n}{\sigma^2}}{n-1}}} = \frac{\frac{\frac{\bar{X}-\mu}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{\chi^2}{n-1}}} \qquad da\ \chi^2 = \sum_i \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{n \cdot S^2}{\sigma^2}$$

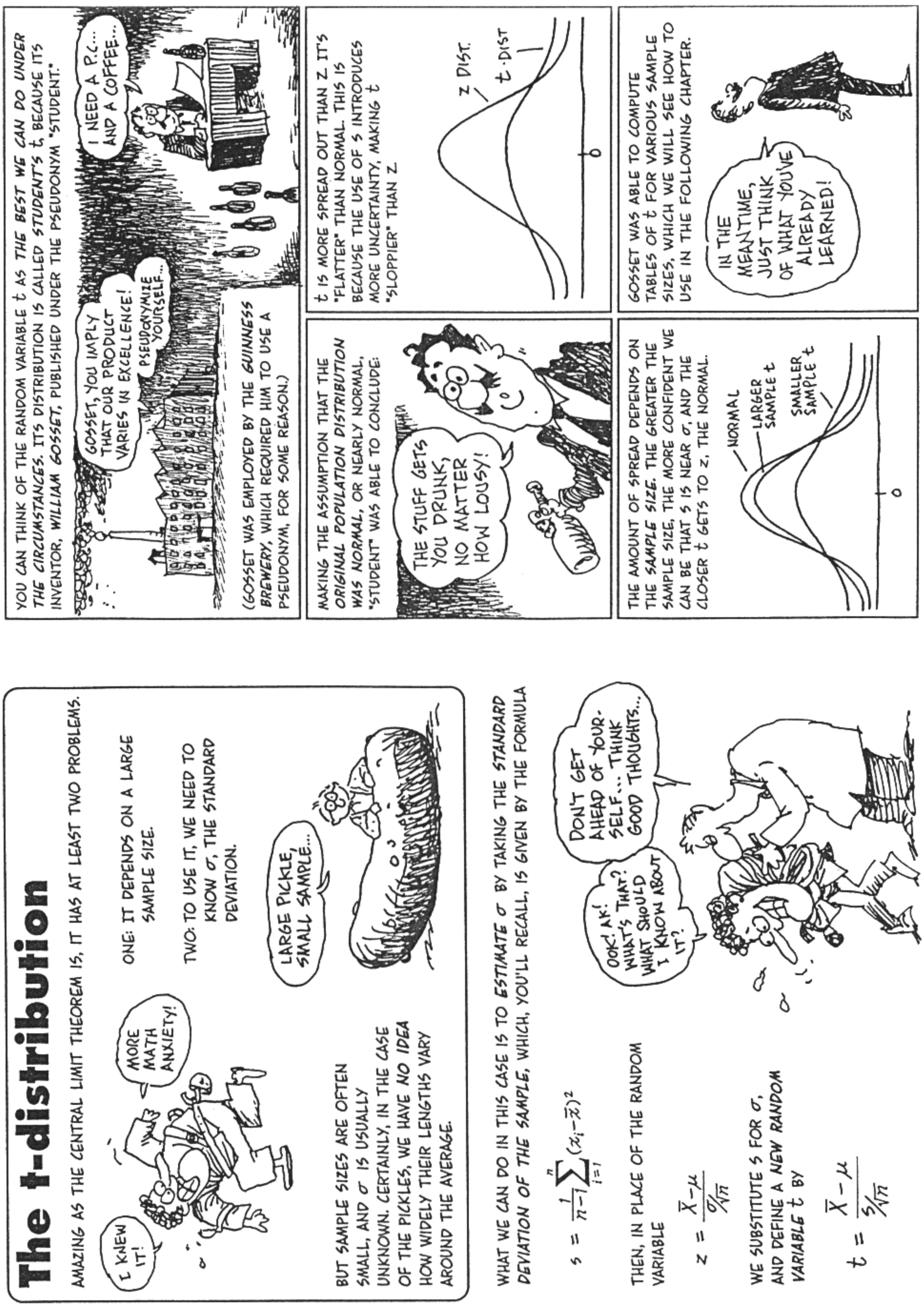
ist t-verteilt mit ν = n – 1 Freiheitsgraden, da

$$\frac{\bar{X}-\mu}{\sigma / \sqrt{n}} \sim \text{N}(0,1) \text{ und } \frac{S^2 \cdot n}{\sigma^2} \sim \chi^2(n-1).$$



Tabellenwerte:
Für verschiedene Freiheitsgrade ν und für verschiedene P (T < t) sind die t-Werte tabelliert.

z. B.: ν = 13
P (T < + t) = 99,5 %
⇒ t = 3,012



Gonick, Smith 1993

3.3 F-Verteilung

Nach FISHER gilt:
Der Quotient aus zwei voneinander unabhängigen χ^2-verteilten und durch ihre Freiheitsgrade dividierten Zufallsvariablen U_1 und U_2 ist F-verteilt:

$$F = \frac{\frac{U_1}{\nu_1}}{\frac{U_2}{\nu_2}} = \frac{U_1}{U_2} \cdot \frac{\nu_2}{\nu_1}$$

Dichtefunktion der F-Verteilung

$$f_F(z|\nu_1, \nu_2) = C(\nu_1, \nu_2) \cdot \frac{z^{\frac{\nu_1}{2}-1}}{(\nu_2 + \nu_1 z)^{\frac{\nu_1+\nu_2}{2}}}, \qquad z > 0$$

$$\text{mit } C(\nu_1, \nu_2) = \frac{\nu_1^{(\nu_1/2)} \cdot \nu_2^{(\nu_2/2)}}{\frac{\Gamma\left(\frac{\nu_1}{2}\right) \cdot \Gamma\left(\frac{\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2} + \frac{\nu_2}{2}\right)}}$$

Erwartungswert

$$E(F) = \frac{\nu_2}{\nu_2 - 2} \qquad \text{für } \nu_2 \geq 2$$

Varianz

$$Var(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \qquad \text{für } \nu_2 > 4$$

Tabellenwerte:

In Abhängigkeit von ν_1 und ν_2 sind die Werte F tabelliert, bei denen die Verteilungsfunktion den Wert 0,90 bzw. 0,995 annimmt:

z.B. $\nu_1 = 4$, $\nu_2 = 8$; $F^{0,99} = 7{,}01$.

Für zwei voneinander unabhängige Stichproben mit Umfang n_1 bzw. n_2 aus derselben Grundgesamtheit, also $\sigma_1^2 = \sigma_2^2$, ergibt sich:

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_{1i} - \overline{X}_1)^2 \qquad S_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (X_{2i} - \overline{X}_2)^2$$

$$\frac{S_1^2 \cdot n_1}{\sigma^2} \sim \chi^2(n_1 - 1) \qquad \frac{S_2^2 \cdot n_2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

Stichprobenfunktion

$$F = \frac{\frac{S_1^2 \cdot n_1}{\sigma^2}}{\frac{S_2^2 \cdot n_2}{\sigma^2}} \cdot \frac{n_2 - 1}{n_1 - 1}$$

$$= \frac{S_1^2 \cdot n_1}{S_2^2 \cdot n_2} \cdot \frac{n_2 - 1}{n_1 - 1} = \frac{\frac{\chi_1^2}{n_1 - 1}}{\frac{\chi_2^2}{n_2 - 1}}$$

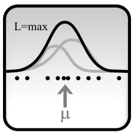
ist F-verteilt mit $n_1 - 1$ und $n_2 - 1$ Freiheitsgraden.

Die F-Verteilung ist wichtig für Signifikanztests der Parameter in ökonometrischen Modellen (vgl. VIII.9)

Keyconcepts

Induktive Statistik
Stichprobenfunktion
Zentraler Grenzwertsatz
Testverteilungen
t-Verteilung
χ^2-Verteilung
F-Verteilung

VI Punktschätzung



Schätzung einzelner Werte der Grundgesamtheit aus den Ergebnissen der Stichprobe

Die Parameter der Grundgesamtheit sind meist unbekannt. Daher versucht man eine Annäherung durch die Stichprobenparameter (Schluß: Stichprobe → Grundgesamtheit).

Punktschätzung

Aufgrund der Stichprobenergebnisse (mit Zufallsprinzip) werden unbekannte Parameter der Grundgesamtheit geschätzt. Der geschätzte Parameter wird hierbei exakt angegeben, allerdings wird der „wahre" Parameter aus der Grundgesamtheit nur mit einer sehr geringen Wahrscheinlichkeit auch wirklich genau dem geschätzten Wert entsprechen. Das Ergebnis einer solchen Schätzung könnte beispielsweise lauten: „Das durchschnittliche Einkommen in der Grundgesamtheit liegt bei 42.352,35 € (mit einer Wahrscheinlichkeit von $12 * 10^{-20}$)." Die hohe „Genauigkeit" der Angabe bis auf zwei Nachkommastellen wird also erkauft durch eine recht geringe Wahrscheinlichkeit, dass der Wert nicht etwa 42.352,36 € (oder 42.352,34…) ist.

Intervallschätzung

Es werden Vertrauensbereiche (Genauigkeit und Sicherheit) der Schätzergebnisse berechnet. Hierbei wird für den gesuchten Parameter ein Bereich angegeben, in dem er mit einer relativ hohen Wahrscheinlichkeit liegen wird. Eine typische Aussage, die sich aus einer Intervallschätzung ableiten ließe, wäre, um das obige Beispiel wieder aufzunehmen: „Das durchschnittliche Einkommen in der Grundgesamtheit liegt mit 95%er Wahrscheinlichkeit zwischen 41.595,36 € und 44.231,56 €." Die höhere Ungenauigkeit bei Angabe des Wertes führt hierbei zu einer höheren Wahrscheinlichkeit, dass der wahre Wert der Grundgesamtheit tatsächlich im Schätzbereich liegt.

Testverfahren

Es wird gefragt, ob Hypothesen über Eigenschaften der Grundgesamtheit zutreffen. Beispielsweise könnte die Hypothese, dass das durchschnittliche Einkommen in der Grundgesamtheit unter 40.000 € liegt, mit Hilfe von Testverfahren verworfen oder „gesichert" werden.

1 Grundlagen der Punktschätzung

Wahrer aber unbekannter Parameter der Grundgesamtheit: θ (Theta)
(z.B. μ, σ, p, ...)

Der Schätzwert für θ ist eine Funktion der Stichprobenergebnisse (Stichprobenfunktion) und damit wie diese vom Zufall abhängig:

$$\hat{\theta} = g_n(x_1, x_2, \ldots, x_n).$$

Dieser Schätzwert kann als Realisation einer speziellen Stichprobenfunktion, die man auch als **Schätzfunktion** bezeichnet, interpretiert werden:

$$\hat{\theta} = g_n(X_1, X_2, \ldots, X_n).$$

Beispiele:

Schätzfunktion für den Mittelwert μ

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_i X_i$$

mit Schätzwert z.B.

$$\hat{\theta} = \hat{\mu} = g_4(12,13,15,18) = \frac{1}{4}(12+13+15+18) = 14{,}5,$$

wobei man $x_1 = 12$, $x_2 = 13$, $x_3 = 15$ und $x_4 = 18$ als Realisation oder als Stichprobenergebnis bezeichnet.

Schätzfunktion für eine Grundwahrscheinlichkeit p

$$\hat{p} = \frac{X}{n} \; (relative\,Häufigkeit)$$

mit Schätzwert z.B.:

$$\hat{\theta} = \hat{p} = g_{60}(12), n = 60$$

$$\hat{p} = \frac{12}{60} = 0{,}2$$

2 Eigenschaften von Schätzfunktionen

Ein Grundproblem bei Punktschätzungen liegt in der Auswahl der Schätzfunktion. Aus der Menge der zu Verfügung stehenden Schätzfunktionen muss diejenige ausgewählt werden, die den unbekannten Parameter der Grundgesamtheit am „besten“ schätzt.

Man muß also geeignete Kriterien zur Beurteilung der Güte von Schätzfunktionen entwickeln: Wir werden feststellen, daß z.B. Erwartungstreue und Effizienz wünschenswerte Eigenschaften von Schätzfunktionen sind.

2.1 Erwartungstreue

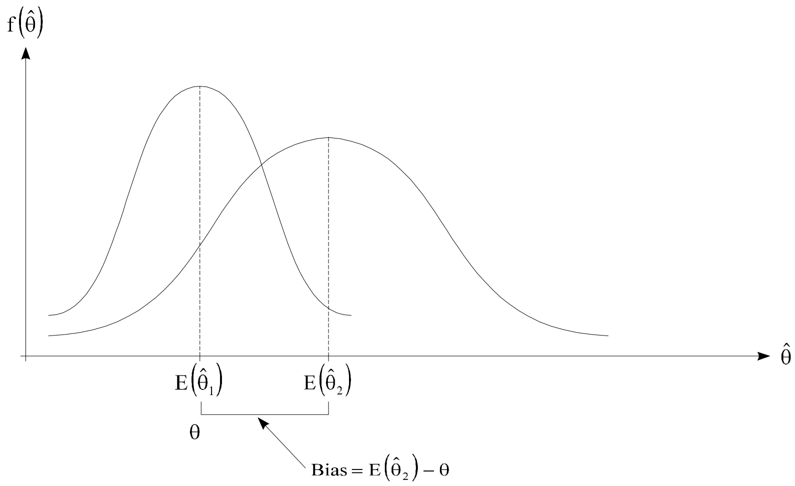
Eine Schätzfunktion $\hat{\theta}=g_n(X_1,X_2,...,X_n)$ für den unbekannten Parameter θ heißt erwartungstreu, wenn

$E[\hat{\theta}]=\theta$, die Schätzung also im Mittel den gesuchten wahren Wert trifft.

Als Verzerrung ('bias') einer Schätzfunktion bezeichnet man dann die Differenz:

$E[\hat{\theta}]-\theta.$

Falls für diese Differenz $E[\hat{\theta}]-\theta=0$ gilt, dann heißt $\hat{\theta}$ erwartungstreu.



θ_1 ist erwartungstreu, da $E[\hat{\theta}_1]=\theta$, d.h., Bias = 0.

Beispiele:

a) Ist die Schätzfunktion $\hat{p}=\frac{X}{n}$ für den unbekannten Parameter p einer binomialverteilten Zufallsvariablen X erwartungstreu?

$$E(\hat{p})=E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)=\frac{1}{n}\cdot n\cdot p=p$$

$\Rightarrow E\left(\hat{\theta}=\frac{X}{n}\right)=\theta=p,$ also ist $\hat{p}$ erwartungstreu.

b) Schätzfunktion für den Mittelwert μ: $\overline{X}=\frac{1}{n}\sum_i X_i$

$$E(\overline{X})=E\left[\frac{1}{n}\sum_i X_i\right]=\frac{1}{n}\left[E(X_1)+E(X_2)+...+E(X_n)\right]$$

Da X_i aus einer Grundgesamtheit stammt, gilt: $E(X_1)=E(X_2)=...=\mu$

$\Rightarrow E(\overline{X})=\frac{1}{n}\cdot n\cdot\mu=\mu \Rightarrow E\left(\hat{\theta}=\overline{X}=\frac{1}{n}\sum_i X_i\right)=\theta=\mu,$ also ist $\overline{X}$ erwartungstreu

c) Klassische lineare Regression: Erklärungsmodell:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_K x_{Ki} + \varepsilon_i$$

$\Rightarrow$ Schätzfunktion für β: $b_{OLS} = \left(X'X\right)^{-1} X'y$

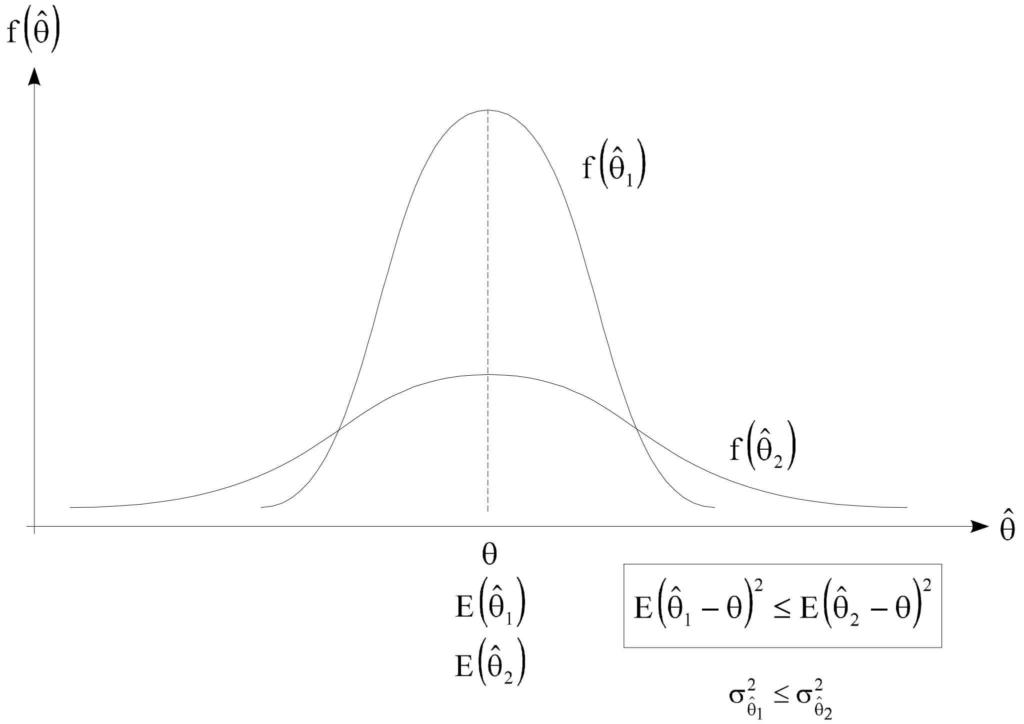
Man kann $E\left(\hat{\theta} = b_{OLS}\right) = \theta = \beta$,

also zeigen, daß b_{OLS} erwartungstreu ist.

(siehe meine Vorlesung: Regressionsanalyse – Einführung in die Ökonometrie).

2.2 Effizienz (Minimale Varianz)

Es können verschiedene erwartungstreue Schätzfunktionen mit der Eigenschaft $E(\hat{\theta}) = \theta$ existieren, die sich aber in ihrer Varianz unterscheiden, d.h., die Abweichungen vom Mittelwert können für verschiedene erwartungstreue Schätzfunktionen verschieden groß sein.



(Relative) Effizienz

Ist die Streuung (Varianz) einer erwartungstreuen Schätzfunktion $\hat{\theta}_1$ kleiner als die Streuung einer anderen erwartungstreuen Schätzfunktion $\hat{\theta}_2$, dann heißt $\hat{\theta}_1$ effizient in bezug auf die Schätzfunktion $\hat{\theta}_2$ (**relative Effizienz**).

(Absolute) Effizienz

Ist die Streuung (Varianz) der Schätzfunktion $\hat{\theta}_1$ minimal, d.h. kleiner als die Streuung aller anderen erwartungstreuen Schätzfunktionen , dann heißt $\hat{\theta}_1$ **absolut effizient**.

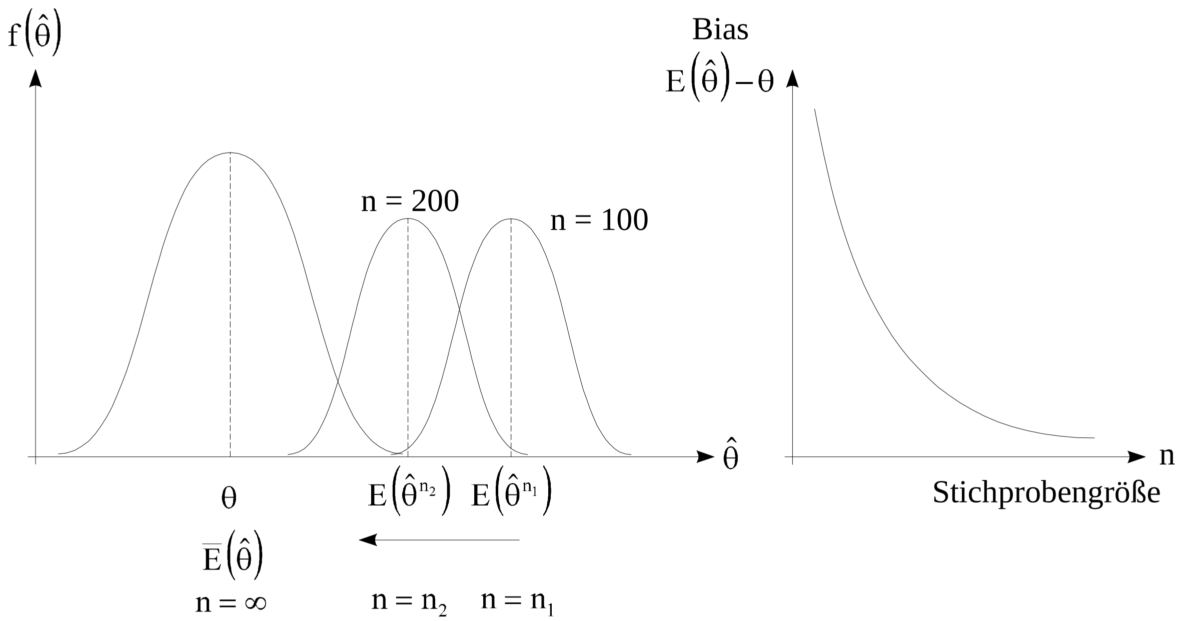
2.3 Asymptotische Erwartungstreue und Konsistenz

Asymptotische Erwartungstreue und **Konsistenz** sind Schätzeigenschaften für große Stichproben.

Asymptotische Erwartungstreue

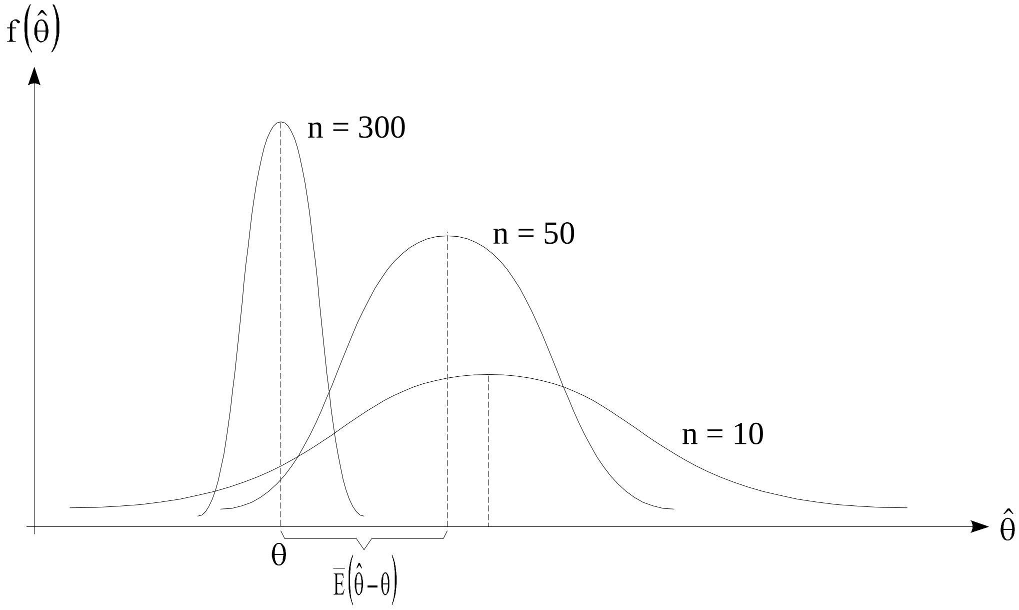
Die Erwartungswerte von immer größer werdenden Stichproben weichen immer weniger von θ ab:

$\overline{E}(\hat{\theta}) = \lim_{n\to\infty} E(\hat{\theta}^{(n)}) = \theta,$ wobei n die Stichprobengröße darstellt.



Konsistenz

Die Verzerrung und Varianz geht mit wachsendem Stichprobenumfang n gegen Null.



$$\text{plim}\,\hat{\theta} = \theta$$

$$\lim_{n\to\infty} prob\left\{\left|\hat{\theta}^n - \theta\right| \geq \delta\right\} = 0$$

3 Schätzmethoden

Die Schätzmethoden dienen zur Bestimmung geeigneter Schätzfunktionen zur Schätzung eines Parameters θ.

3.1 Methode der Momente (K. PEARSON)

Die Parameter der Stichprobe werden als Schätzer für die entsprechenden Parameter der Grundgesamtheit verwendet.

Die Methode der Momente wird insbesondere dann angewendet, wenn die Verteilung der Grundgesamtheit nicht bekannt ist.

Die mit der Momentenmethode abgeleiteten Schätzfunktionen besitzen im allgemeinen nur wenige der wünschenswerten Schätzeigenschaften.

Beispiele:

- Schätzfunktion für μ der Poissonverteilung: $\hat{\mu} = \overline{X}$
- Schätzfunktion für μ und σ^2 der Normalverteilung: $\hat{\mu} = \overline{X}$ und $\hat{\sigma}^2 = s^2$

3.2 Methode der kleinsten Quadrate (MKQ/OLS)

MKQ: Methode der kleinsten Quadrate (OLS: Ordinary least Squares)

MKQ: Aus der Minimierung aller quadrierten Abweichungen zwischen beobachteten Werten und Werten auf der zu findenden Ausgleichsgerade werden die Parameter der Ausgleichsgeraden (R^n: Hyperebene) berechnet. (Zur Methode: vgl. z. B. Merz 1993, Statistik I – Deskription)

Im **klassischen linearen Regressionsmodell** (Classical Linear Regression, CLR) wird MKQ im Rahmen eines **stochastischen Modells** mit wiederholten Beobachtungen in einer Stichprobe verwendet.

Stochastisches Modell:

$$y_i = \underbrace{\underbrace{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_K x_{Ki}}_{\mu_i}}_{\text{systematischer Einfluß } f(x_i)} + \underbrace{\underbrace{\varepsilon_i}_{\varepsilon_i}}_{\text{zufälliger Einfluß}}$$

$\varepsilon_i =$ Fehler mit bestimmter Verteilung: $N(\mu_i, \sigma^2)$

$\beta_k =$ zu schätzender, unbekannter Parameter der Grundgesamtheit

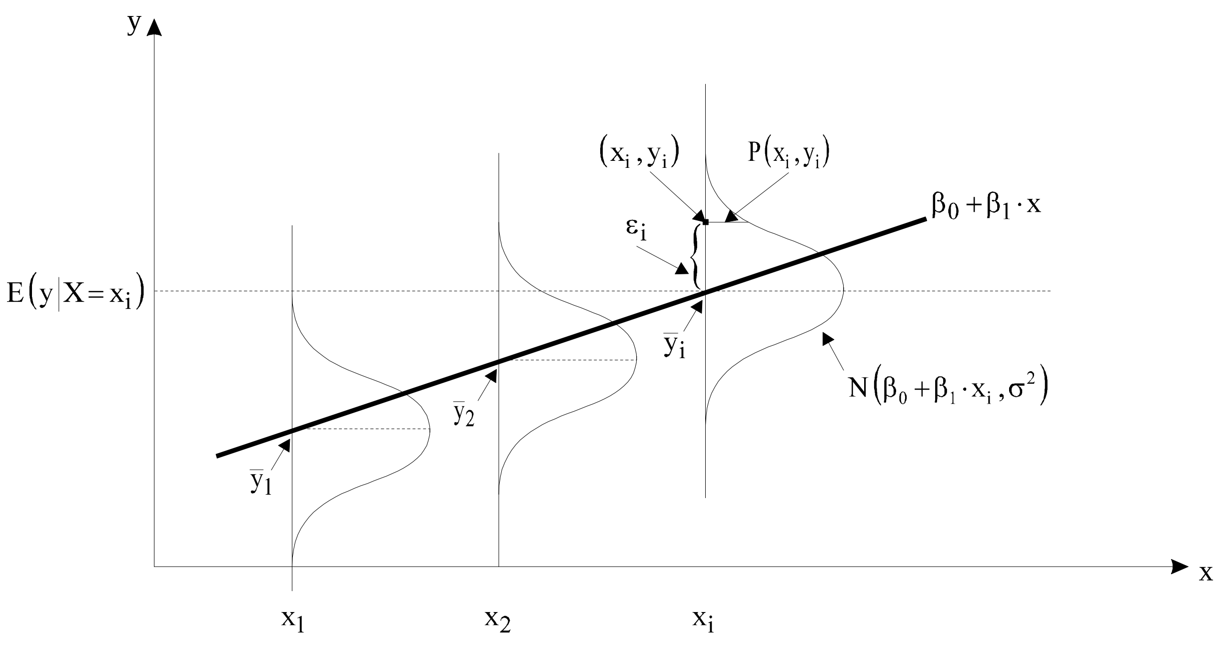
$x_k =$ erklärende Größen

$y_i =$ zu erklärende Größe

Zu jedem x_i gibt es mehrere Beobachtungen y, bzw. es werden m Stichproben bei gleichen x_i gezogen ('fixed in repeated samples'): $\rightarrow$ Wahrscheinlichkeitsverteilung der y

Zufallsvariablen (Stichprobe der Größe n mit ℓ unterschiedlichen X-Werten):

X	Y				Mittelwert $\bar{y}_i$
x_1	y_{11}	y_{12}	...	y_{1m}	$\bar{y}_1$
x_2	y_{21}	y_{22}	...	y_{2m}	$\bar{y}_2$
.	.	.	...	.	.
.	.	.	...	.	.
x_ℓ	$y_{\ell 1}$	$y_{\ell 2}$	...	$y_{\ell m}$	$\bar{y}_l$



$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$Ey_i = E[\beta_0 + \beta_1 x_i + \varepsilon_i]$$
$$= \beta_0 + \beta_1 x_i + E\varepsilon_i (= 0)$$
$$= \beta_0 + \beta_1 x_i$$

Schätzung der Regressionskoeffizienten aus $(x_1, y_1), (x_2, y_2)...$:

$$y = \begin{cases} \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + \varepsilon \; Grundgesamtheit \\ b_0 + b_1 x_1 + ... + b_K x_K + e \;\; Stichprobe \end{cases}$$

$$\hat{y} = Xb \, mit \, \hat{y}_i = b_0 + b_1 x_{1i} + ... + b_K x_{Ki} \;\; als \; Schätzer \;\; für \; Ey(Ausgleichsgerade)$$

MKQ/OLS: Minimierung der Fehlerquadratsumme:

$$\sum_i (y_i - \hat{y}_i)^2 = \min!$$

$$\sum_i \left(y_i - \left(b_0 + b_1 x_{1i} + ... + b_K x_{Ki} \right) \right)^2 = \sum_i (e_i)^2 = \min!$$

Die notwendigen Bedingungen für ein Minimum sind die zu findenden Werte b_K, die die ersten Ableitungen der Fehlerquadratsumme nach den $ß_K$ Null setzen (Normalgleichungs-system):

$$X'Xb = X'y$$

Lösung des K+1-Gleichungssystems bei existierender Inversen $(X'X)^{-1}_{(K+1,K+1)}$ ergibt

$$\hat{\beta} = b^{OLS} = b = (X'X)^{-1} X'y$$

b(= b^{OLS}) ist **BLUE**: **B**est **L**inear **U**nbiased **E**stimator, d.h., b^{OLS} ist linear, erwartungstreu und effizient.

Beispiele:———————————————————————————

a) Sei y_i die Körpergröße des Kindes i und x_i die Körpergröße der Mutter dieses Kindes. CLR-Ansatz:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

Stochastisches Modell: Zu jeder Mutter i mit der Körpergröße x_i gibt es ein oder mehrere Kinder mit jeweiliger Körpergröße. Mehrere Mütter können die gleiche Körpergröße haben (≙ mehrere Stichproben) und jeweils ein oder mehrere Kinder ('fixed in repeated samples'). Das bedeutet, es gibt zu jeder Beobachtung x_i (Muttergröße) eine Verteilung der Kinder-größen y_i. Im klassischen linearen Regressionsmodell wird für diese Verteilung jeweils die Normalverteilung angenommen.

Aus den Beobachtungen der Stichprobe ist dann ein Schätzer für die unbekannten Parameter β_0 und β_1 der Grundgesamtheit z.B. mit MKQ/OLS zu finden.

ET-Befehle:

```
? --------------------------------
? ET hkid: ols by matrix algebra
? --------------------------------
```

```
?
? read data
?      y : hkid
?      x0: one
?      x1: hmother
read; file=hkid.dat;nvar=2;names=1$
list; hkid, hmother$
?
? create X'X (=XSX) and X'y (XSy)
? ------------------------------
namelist; X=one,hmother$
matrix; XSX=xdot(X)$
matrix; XSy=xdot(X,hkid)$
?
? create invers of (X'X)
? ------------------------------
calc; DET = XSX(1,1)*XSX(2,2) - XSX(1,2)*XSX(2,1)$
matrix; A = XSX$
calc; a12 = -1*XSX(1,2)
    ; a21 = -1*XSX(2,1)$
matrix; A(1,1) = XSX(2,2)
      ; A(1,2) = a12
      ; A(2,1) = a21
      ; A(2,2) = XSX(1,1)$
matrix; C = 1/DET*A$




matrix; XSXINV=ginv(XSX)$

?? compute ols? ------------------------------
matrix; bols=XSXINV | XSy$
?
? regression by et
? ------------------------------
regres; dep=hkid; ind=one,hmother$
```

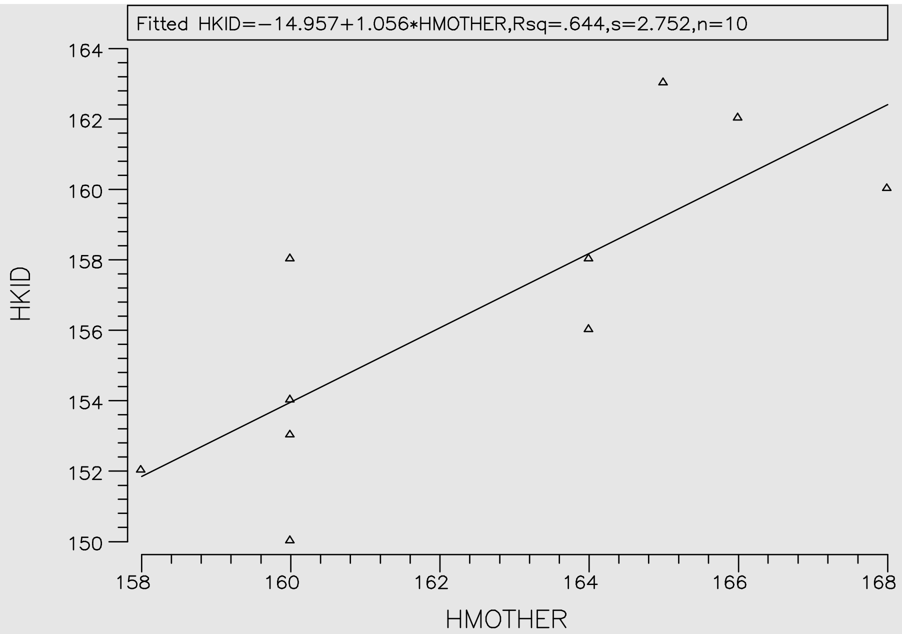
ET-Ergebnis:

```
ET: CLR, Stochastische Regression (OLS)
=======================================

DATA LISTING (Current sample)
Observation      HKID        HMOTHER
      1        154.00      160.00
      2        152.00      158.00
      3        156.00      164.00
      4        160.00      168.00
      5        162.00      166.00
      6        163.00      165.00
      7        158.00      160.00
      8        150.00      160.00
      9        158.00      164.00
     10        153.00      160.00

     1. Matrix -> XSX=XDOT(X)

      <<<< XSX      >>>>   COLUMN
                1             2
 ROW   1   10.0000       1625.00
 ROW   2   1625.00       264161.

     1. Matrix -> XSY=XDOT(X,HKID)
```

```
      <<<< XSY       >>>>   COLUMN
              1
 ROW   1   1566.00
 ROW   2   254579.

>>> DET      =     985.0000     <<<

     1. Matrix -> A=XSX

      <<<< A         >>>>   COLUMN
              1             2
 ROW   1   10.0000       1625.00
 ROW   2   1625.00       264161.

>>> A12      =    -1625.000     <<<

>>> A21      =    -1625.000     <<<

     1. Matrix -> A(1,1)=XSX(2,2)

     2. Matrix -> A(1,2)=A12

     3. Matrix -> A(2,1)=A21

     4. Matrix -> A(2,2)=XSX(1,1)

     1. Matrix -> C=1/DET*A

      <<<< C         >>>>   COLUMN
              1             2
 ROW   1   268.184      -1.64975
 ROW   2  -1.64975       .101523E-01

     1. Matrix -> XSXINV=GINV(XSX)

      <<<< XSXINV    >>>>   COLUMN
              1             2
 ROW   1   268.184      -1.64975
 ROW   2  -1.64975       .101523E-01

     1. Matrix -> BOLS=XSXINV|XSY

      <<<< BOLS      >>>>   COLUMN
              1
 ROW   1  -14.9859
 ROW   2   1.05581
```

d.h., OLS über ET-Matrix-Algebra ergibt: $\beta_0 = -14{,}9859$ und $\beta_1 = 1{,}05581$.

```
==============================================================================
Ordinary    Least Squares
Dependent Variable        HKID         Number of Observations              10
Mean of Dep. Variable     156.6000     Std. Dev. of Dep. Var.        4.351245
Std. Error of Regr.         2.7521     Sum of Squared Residuals   60.5929
R - squared                 .64441     Adjusted R - squared            .59996
F( 1,   8)                 14.4977     Prob. Value for F               .00518
==============================================================================
Variable  Coefficient  Std. Error  t-ratio Prob|t|>x   Mean of X  Std.Dev.of X
------------------------------------------------------------------------------
Constant  -14.9736       45.07       -.332   .74825
HMOTHER    1.05584       .2773       3.808   .00518    162.50000       3.30824
```



b) $\text{Nachfrage} = f\left(\text{Einkommen, Preise, sozioökonomische Charakteristika}\right) + \varepsilon_i$ oder

$\text{Investitionen} = f\left(\text{Zinssatz, Betriebsgröße, }\ldots\right) + \varepsilon_i$

Die OLS-Methode und das klassische lineare Regressionsmodell werden ausführlich in meiner Vorlesung: „Regressionsanalyse – Einführung in die Ökonometrie“ behandelt.

3.3 Maximum Likelihood Methode (R.A. FISHER)

Die Schätzfunktion $\hat{\theta}$ für den unbekannten Parameter θ wird hier so bestimmt, daß die Wahrscheinlichkeit ('Likelihood') dafür, daß die beobachtete Stichprobe aus der angenommenen Grundgesamtheit stammt, maximiert wird (ML-Prinzip; R. A. FISHER (1890–1962)).

Beispiel:

Es existiere ein Lagerbestand von N gleich 10 000 'Schokoladenriegel'. Es wird angenommen, daß die Ausschußquote 10 % oder 20 % beträgt.

Man zieht eine Stichprobe vom Umfang n = 10, um eine Entscheidung über den Ausschußanteil zu fällen.

Die Zahl der fehlerhaften Schokoladenriegel in der Stichprobe ist approximativ binomialverteilt.
Damit erhält man folgende Binomialverteilungen für die Stichprobe:

x	0	1	2	3	4
$f_B(x \mid 10; 0{,}1)$	0,349	0,387	0,194	0,057	0,011
$f_B(x \mid 10; 0{,}2)$	0,107	0,268	0,302	0,201	0,088

Maximum Likelihood (ML) Prinzip:
Man entscheidet sich für die Grundgesamtheit, bei der das beobachtete Stichprobenergebnis die größere Wahrscheinlichkeit besitzt.Wenn beispielsweise für die Stichprobe $x = 1$ Ausschuß gilt, dann stammt die Stichprobe mit größerer Likelihood (38,7 % zu 26,8 %) aus einer Grundgesamtheit mit p = 0,1, d.h. 10 % Ausschußanteil.

Insgesamt gilt hier für x fehlerhafte Schokoladenriegel:

$x \leq 1$: Entscheidung für $\hat{p} = 0{,}1$

$x > 1$: Entscheidung für $\hat{p} = 0{,}2$

Allgemeine ML-Vorgehensweise

Das Merkmal X besitzt die Wahrscheinlichkeits- bzw. Dichtefunktion $f(X|\theta)$. Man besitze ferner eine Zufallsstichprobe $(x_1,\ldots,x_n)$.

Die Wahrscheinlichkeit ('Likelihood'), diese Stichprobe zu erhalten, beträgt unter der Bedingung der stochastischen Unabhängigkeit der Zufallsvariablen $X_1,\ldots,X_n$:

$$L(\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot \ldots \cdot f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Oftmals logarithmiert man die Likelihoodfunktion L(θ), bevor man sie maximiert. Die logarithmische Transformation verschiebt die Extremwerte nicht, d.h., L(θ) besitzt gerade dort ein Maximum, wo auch ln L(θ) ein Maximum besitzt. Die Logarithmierung vereinfacht aber die Maximierung, da eine Summe sich leichter als ein Produkt differenzieren läßt

$$\ln L(\theta) = \ln \prod_{i=1}^{n} f(x_i|\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta).$$

Der Schätzwert $\hat{\theta}$ für θ ist der Wert, der die Likelihoodfunktion maximiert: $L(\hat{\theta}) = \max$. Dazu bildet man – wie bei jedem Optimierungsproblem – die erste Ableitung von $L(\theta)$ bzw. $\ln L(\theta)$ und setzt diese gleich Null:

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{bzw.} \quad \frac{\partial \ln L}{\partial \theta} = 0 \quad \Rightarrow \hat{\theta}$$

Beispiele:

1) ML-Schätzer für den Parameter θ = μ von poissonverteilten Zufallsvariablen $X_1,\ldots,X_n$:

Likelihoodfunktion:

$$L(\mu|x_1,\ldots,x_n) = \prod_{i=1}^{n} \left(\frac{\mu^{x_i}}{x_i!} e^{-\mu} \right)$$

$$\begin{aligned}
\ln L(\mu|x_1,\ldots,x_n) &= \sum_{i=1}^{n} \ln\left(\frac{\mu^{x_i}}{x_i!} \cdot e^{-\mu} \right) \\
&= \sum_i \left[x_i \cdot \ln(\mu) - \ln(x_i!) - \mu \right] \\
&= \sum_{i=1}^{n} x_i \cdot \ln(\mu) - \sum_{i=1}^{n} \ln(x_i!) - \sum_{i=1}^{n} \mu \\
&= \ln(\mu) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!) - n \cdot \mu \\
&= n \cdot \bar{x} \cdot \ln(\mu) - \sum_{i=1}^{n} \ln(x_i!) - n \cdot \mu
\end{aligned}$$

Notwendige Bedingung (1. Ableitung Null setzen):

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\mu} \cdot n \cdot \bar{x} - n \overset{!}{=} 0$$

$$\Rightarrow \frac{1}{\hat{\mu}} \cdot n \cdot \bar{x} = n \qquad \Rightarrow n \cdot \bar{x} = n \cdot \hat{\mu} \qquad \Rightarrow \hat{\mu} = \bar{x}$$

Das heißt, das arithmetische Mittel der Stichprobe $\bar{x}$ ist der ML-Schätzer für μ, den unbekannten Poisson-Parameter.

2) Die Lebensdauer X der Computer Chips von CHIO CHONG genüge einer Verteilung mit der Dichte

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Der Chiptest von fünf Teilen ergebe die folgenden Lebensdauern (in Betriebsstunden):

110, 100, 90, 140, 60

a) Bestimmen Sie den ML-Schätzwert von θ allgemein!

$$L = \prod_{i=1}^{n} f(x_i, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}}$$

$$\begin{aligned} \ln L &= \sum_i \ln\left[\frac{1}{\theta} \cdot e^{-\frac{x_i}{\theta}}\right] = \sum_i \left[\ln\left(\frac{1}{\theta}\right) - \frac{x_i}{\theta} \cdot \underbrace{\ln e}_{=1}\right] \\ &= n \cdot \ln\left(\frac{1}{\theta}\right) - \sum_i x_i \cdot \theta^{-1} \\ &= n \cdot \left[\underbrace{\ln 1}_{=0} - \ln \theta\right] - \sum_i x_i \cdot \theta^{-1} \\ &= -n \cdot \ln \theta - \theta^{-1} \cdot \sum_i x_i \end{aligned}$$

Notwendige Bedingung :

$$\frac{\partial \ln L}{\partial \theta} = -n \cdot \frac{1}{\theta} - (-1) \cdot \theta^{-2} \cdot \sum_i x_i \overset{!}{=} 0$$

$$\frac{n}{\hat{\theta}} = \hat{\theta}^{-2} \cdot \sum_i x_i$$

$$\frac{\hat{\theta}^2 \cdot n}{\hat{\theta}} = \sum_i x_i$$

$$\hat{\theta}^{opt} = \frac{1}{n} \sum_i x_i$$

Das arithmetische Mittel ist der ML-Schätzer $\hat{\theta}^{opt}$. Der ML-Schätzwert $\hat{\theta}^{opt}$ ist der Parameter, der das Stichprobenergebnis $x_1,...,x_n$ mit größter Wahrscheinlichkeit liefert.

b) Wie lautet der ML-Schätzer für diese Stichprobe?

$$\hat{\theta}=\frac{1}{n}\sum_i x_i$$

$$=\frac{1}{5}\cdot(110+100+90+140+60)=\frac{1}{5}\cdot(500)=100$$

$$f(x_i,\theta=100)=\begin{cases}\frac{1}{100}\cdot e^{-\frac{x_i}{100}} & für\, x>0\\ 0 & sonst\end{cases}$$

$$Sei\, \tilde{\theta}=\frac{1}{2n}\cdot\sum_i x_i:$$

$$f(x_i,\tilde{\theta}=50)=\begin{cases}\frac{1}{50}\cdot e^{-\frac{x_i}{50}} & \text{für } x>0\\ 0 & \text{sonst}\end{cases}$$

x_i	$f(x_i, \hat{\theta}=100)$	$f(x_i, \tilde{\theta}=50)$
60	0,00549	0,00602
90	0,00407	0,00331
100	0,00368	0,00271
110	0,00333	0,00222
140	0,00247	0,00122

$$L(\hat{\theta})=\prod_{i=1}^{n} f(x_i,\hat{\theta}=100)=6,76325\cdot 10^{-13}$$

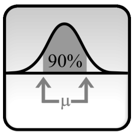
$$L(\tilde{\theta})=\prod_{i=1}^{n} f(x_i,\tilde{\theta}=50)=1,46254\cdot 10^{-13}$$

Damit maximiert gerade $\hat{\theta}=100$ gegenüber jedem anderen Schätzer (wie bspw. $\tilde{\theta}=50$) die Likelihood.

Keyconcepts

Punktschätzung
Unbekannter „wahrer“ Parameter
Erwartungstreue
Effizienz
Asymptotische Erwartungstreue
Konsistenz
Methode der kleinsten Quadrate (MKQ), Ordinary Least Squares (OLS)
Maximum Likelihood

VII Intervallschätzung



Schätzung von Wertebereichen (Intervallen), in denen ein Wert der Grundgesamtheit mit hoher Wahrscheinlichkeit liegt

1 Konfidenzintervall

Bei der behandelten Punktschätzung wird für einen unbekannten Parameter θ, wie z. B. ein Mittelwert, ein Anteilswert oder eine Standardabweichung, als Schätzwert die Realisierung einer Zufallsvariablen, also ein bestimmter Zahlenwert angegeben.
Auch wenn die dazugehörige Schätzfunktion sehr gute Schätzeigenschaften besitzt, wird der Schätzwert im allgemeinen vom unbekannten Parameter der Verteilung der Grundgesamtheit mehr oder weniger stark abweichen.

Ziel der Intervallschätzung ist es, ein Intervall anzugeben, das 'in den meisten Fällen' den unbekannten Parameter θ tatsächlich enthält.
Die Intervallschätzung liefert also ein Intervall, in dem der zu schätzende unbekannte Parameter θ mit einer bestimmten Wahrscheinlichkeit $(1 - \alpha)$ erwartet wird.

Diese Intervalle bezeichnet man als **Konfidenz- oder Vertrauensintervalle.**

Aufgrund der Stichprobenergebnisse werden die Intervallgrenzen bestimmt. Die Intervallgrenzen sind die Stichprobenfunktionen $\hat{\theta}_u$ und $\hat{\theta}_o$. Allgemein gilt dann:

$$P(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1 - \alpha .$$

Die **Irrtumswahrscheinlichkeit** α bezeichnet die Wahrscheinlichkeit, mit der das Schätzverfahren ein Intervall liefert, das den Parameter *nicht* enthält (zugelassene Irrtumwahrscheinlichkeit).

Die **Konfidenzwahrscheinlichkeit (Vertrauens- oder Sicherheitswahrscheinlichkeit)** $1-\alpha$ bezeichnet die Wahrscheinlichkeit, mit der die Schätzung ein Intervall liefert, das den unbekannten Parameter θ tatsächlich enthält. $1-\alpha$ wird auch als Konfidenzniveau bezeichnet.

2 Konfidenzintervall für das arithmetische Mittel bei normalverteilter Grundgesamtheit

Bei normalverteilter Grundgesamtheit (bzw. über den zentralen Grenzwertsatz bei genügend großen Stichproben) ist das arithmetische Mittel, die Zufallsvariable $\overline{X}$, ebenfalls normalverteilt mit:

$$E\left(\overline{X}\right) = \mu$$

$$Var\left(\overline{X}\right) = \frac{Var\left(X\right)}{n} = \frac{\sigma^2}{n} = \sigma_{\bar{x}}^2$$

Gonick, Smith 1993

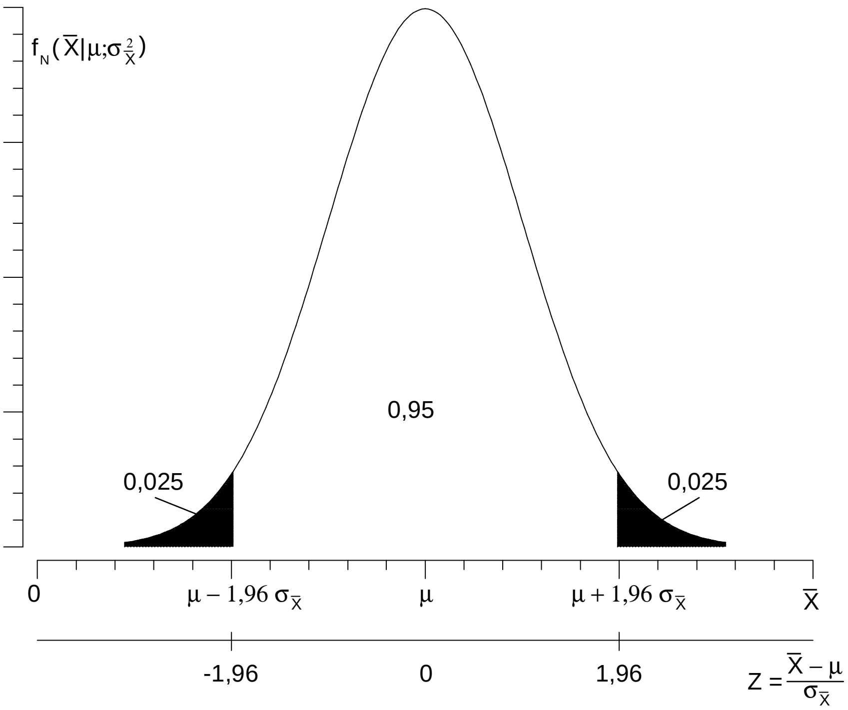


Abb. Konfidenzintervall für $\bar{X}$ mit $\alpha = 0{,}05$

Zum Konfidenzniveau von 95 % ergibt sich für $\bar{X}$ das symmetrische Konfidenzintervall

$$P\left(\mu - 1{,}96 \cdot \sigma_{\bar{x}} \le \bar{X} \le \mu + 1{,}96 \cdot \sigma_{\bar{x}}\right) = 1 - \alpha = 0{,}95.$$

Ist allgemein z_α der Wert der standardnormalverteilten Zufallsvariablen bei dem die Verteilungsfunktion den Wert α annimmt, und setzen wir anstelle von 95 % allgemein $1-\alpha$, so ergibt sich statt 1,96 bzw. –1,96 allgemein $z_{1-\alpha/2}$ bzw. $z_{\alpha/2}$ mit

$$P(\mu + z_{\alpha/2} \cdot \sigma_{\bar{X}} \le \bar{X} \le \mu + z_{1-\alpha/2} \cdot \sigma_{\bar{X}}) = 1 - \alpha.$$

Wegen der Symmetrie der Normalverteilung gilt $z_{\alpha/2} = -z_{1-\alpha/2}$ und damit:

$$P(\mu - z_{1-\alpha/2} \cdot \sigma_{\bar{X}} \le \bar{X} \le \mu + z_{1-\alpha/2} \cdot \sigma_{\bar{X}}) = 1 - \alpha$$

In diesem Intervall liegt $\bar{X}$ mit der Wahrscheinlichkeit $1-\alpha$.

Die Konstruktion des Konfidenzintervalls für $\bar{X}$ zu einem gegebenen α veranschaulichen die folgenden Abbildungen:

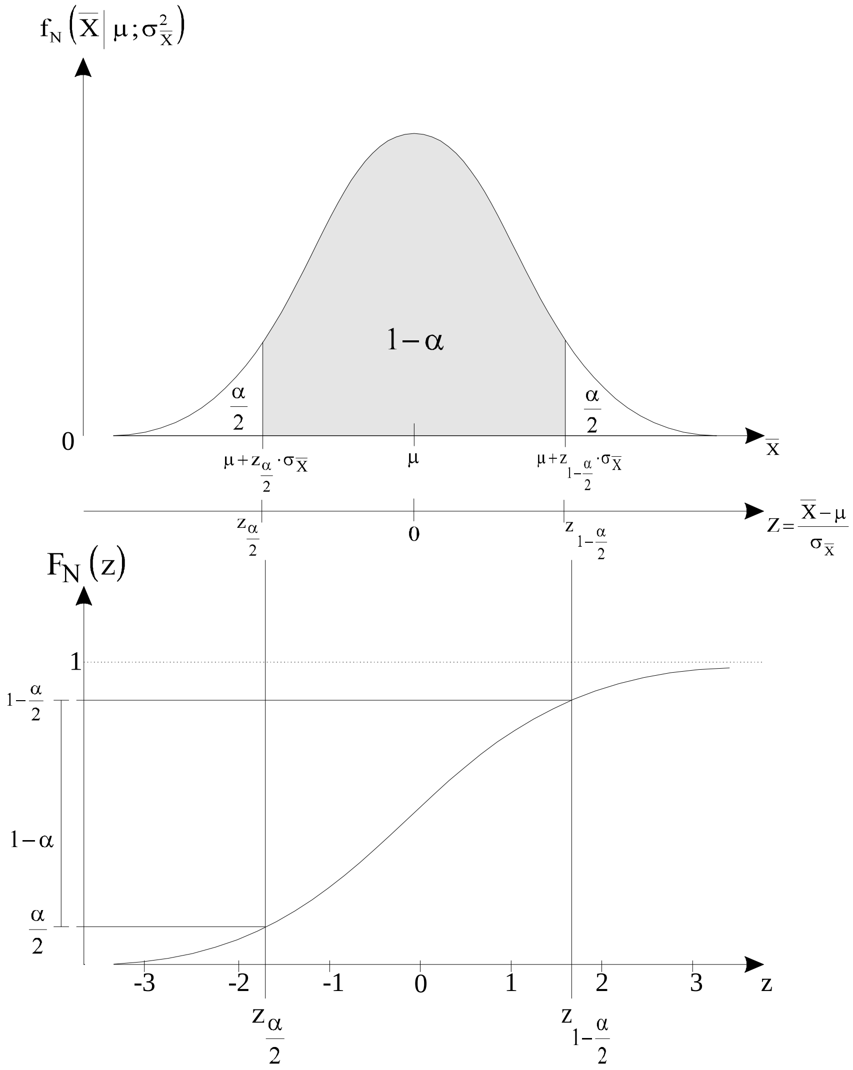


Abb. Konstruktion des Konfidenzintervalls für $\overline{X}$ bei gegebenem α

2.1 Konfidenzintervall für μ bei bekannter Varianz σ^2 der normalverteilten Grundgesamtheit

Gesucht wird nun das Konfidenzintervall für das unbekannte arithmetische Mittel μ der Grundgesamtheit bei bekannter Varianz.

Dazu wird das Konfidenzintervall des Stichprobenmittelwertes nur umgerechnet. Aus der obigen Intervallbeziehung

$$P\left(\mu - z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \leq \overline{X} \leq \mu + z_{1-\alpha/2} \cdot \sigma_{\bar{x}}\right) = 1-\alpha$$

wird durch Subtraktion von μ innerhalb der Klammer und Multiplikation mit -1 in der Klammer:

$$P\left(z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \geq -\overline{X} + \mu \geq -z_{1-\alpha/2} \cdot \sigma_{\bar{x}}\right) = 1-\alpha\,.$$

Die Addition von $\overline{X}$ in der Klammer liefert:

$$P\left(\overline{X} + z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \geq \mu \geq \overline{X} - z_{1-\alpha/2} \cdot \sigma_{\bar{x}}\right) = 1-\alpha$$

bzw.

$$P\left(\bar{X} - z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \cdot \sigma_{\bar{x}}\right) = 1 - \alpha .$$

Das Intervall $\left[\bar{X} - z_{1-\alpha/2} \cdot \sigma_{\bar{X}}, \bar{X} + z_{1-\alpha/2} \cdot \sigma_{\bar{X}}\right]$ ist eine Zufallsvariable, weil $\bar{X}$ eine Zufallsvariable ist und damit auch die Grenzen des Intervalls Zufallsvariablen sind. Es überdeckt den unbekannten wahren Parameter $\theta = \mu$ mit der Wahrscheinlichkeit $1 - \alpha$.

Bei mehrmaligen, häufigen Wiederholungen der Stichprobenerhebungen werden $(1-\alpha)\%$ der Intervalle den Parameter μ überdecken.

Prinzipiell kann hier für die Konfidenzintervallberechnung die dazu notwendige Mittelwertstandardabweichung $\sigma_{\bar{X}}$ aus der Varianz bzw. Standardabweichung σ der Zufallsvariablen X aus

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

berechnet werden (siehe oben).

Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0,05$:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

(n = Stichprobenumfang, N = Umfang der Grundgesamtheit).

2.2 Konfidenzintervall für μ bei unbekannter Varianz σ^2 der normalverteilten Grundgesamtheit

Ausgehend von einer normalverteilten Grundgesamtheit, ist die Zufallsvariable

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$$

standardnormalverteilt.

Wie kann nun die unbekannte Varianz des arithmetischen Mittels $Var\left(\bar{X}\right) = \sigma_{\bar{X}}^2$ aus der Stichprobe geschätzt werden?

Aus dem Kapitel V.2 Stichprobenfunktionen wissen wir, daß

$$Var\left(\bar{X}\right) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} .$$

Würden wir alleine die Varianz der Zufallsvariablen X

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

verwenden, hätten wir **keine** erwartungstreue Schätzfunktion für σ^2 (es wäre $E\left(S^2\right) = \frac{n-1}{n} \sigma^2$).

Erst mit entsprechender (Besselscher) Korrektur $n/(n-1)$ erhält man eine **erwartungstreue Schätzung für die Varianz von X der Grundgesamtheit**

$$\hat{\sigma}^2 = \frac{n}{n-1} S^2$$

und weiter den **erwartungstreuen Schätzer für die gesuchte Varianz des arithmetischen Mittels**

$$\hat{\sigma}_{\overline{X}}^2 = \frac{\hat{\sigma}^2}{n} = \frac{S^2}{n} \cdot \frac{n}{n-1} = \frac{S^2}{n-1} \qquad \text{bzw.}$$

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n-1}}.$$

Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0,05$:

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n-1}} \sqrt{\frac{N-n}{N}}.$$

Durch Einsetzen des erwartungstreuen Schätzers $\hat{\sigma}_{\overline{X}}$ in Z erhält man die neue Zufallsvariable

$$T = \frac{\overline{X} - \mu}{\hat{\sigma}_{\overline{X}}},$$

die mit **ν = n – 1 Freiheitsgraden studentverteilt** ist.

Diese Zufallsvariable T wurde bereits bei der Behandlung der t-Verteilung als Beispiel herangezogen.

Werden mit $t_{\alpha/2,\nu}$ bzw. $t_{1-\alpha/2,\nu}$ die Punkte bezeichnet, bei denen die Verteilungsfunktion der t-Verteilung mit ν Freiheitsgraden die Werte α/2 bzw. 1 – α/2 annimmt, so ergibt sich

$$P\left(t_{\alpha/2,n-1} \leq \frac{\overline{X} - \mu}{\hat{\sigma}_{\overline{X}}} \leq t_{1-\alpha/2,n-1} \right) = 1 - \alpha.$$

Da die t-Verteilung symmetrisch ist, gilt

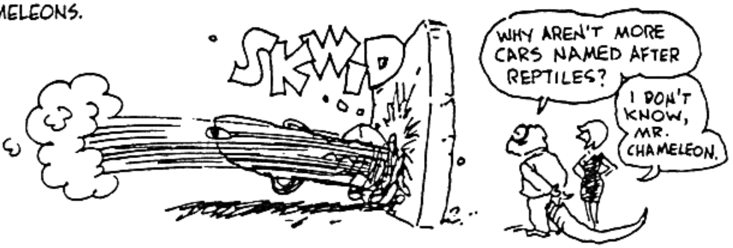
$$t_{1-\alpha/2,n-1} = -t_{\alpha/2,n-1} := t.$$

Somit ist das Konfidenzintervall

$$P\left(\overline{X} - t \cdot \hat{\sigma}_{\overline{X}} \leq \mu \leq \overline{X} + t \cdot \hat{\sigma}_{\overline{X}} \right) = 1 - \alpha.$$

Für Freiheitsgrade von ν ≥ 30 kann die t-Verteilung durch die Normalverteilung approximiert werden.

example: SUPPOSE *CHAMELEON MOTORS* HAS TO CRASH TEST ITS CARS TO DETERMINE THE AVERAGE REPAIR COST OF A 10 M.P.H. HEAD-ON COLLISION. THIS IS EXPENSIVE! THEY DECIDE TO TRY IT ON JUST *FIVE* CHAMELEONS.



THEY FIND THE DAMAGE DATA TO BE \$150, \$400, \$720, \$500, AND \$930.

THE SAMPLE MEAN:

$\bar{x} = \$540$

THE STANDARD DEVIATION:

$s = \$299$

YOU CAN CHECK s WITH A HAND CALCULATOR. IT'S

$$\sqrt{\tfrac{1}{4}\left((150-540)^2+(400-540)^2+(720-540)^2+(500-540)^2+(930-540)^2\right)}$$

SO WHERE CAN WE PLACE THE MEAN WITH 95% CONFIDENCE? WE FIND OUR CRITICAL VALUE $t_{.025}$ WITH 4 DEGREES OF FREEDOM:

	$1-\alpha$	.80	.90	.95	.99
	α	.20	.10	.05	.01
	$\alpha/2$	.10	.05	.025	.005
DEGREES OF FREEDOM	1	3.09	6.31	12.71	63.66
	2	1.89	2.92	4.30	9.92
	3	1.64	2.35	3.18	5.84
	4	1.53	2.13	2.78	4.60
	5	1.48	2.01	2.57	4.03

AND PLUG IT IN:

$$\mu = \bar{x} \pm 2.78\, s/\sqrt{n}$$

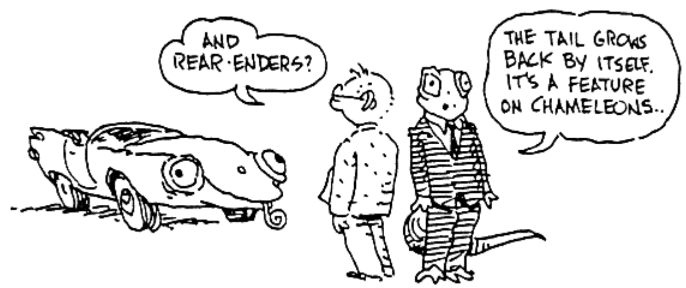
$$= 540 \pm 2.78\,(299/\sqrt{5})$$

$$= \mathbf{540 \pm 372}$$

SO THE BEST WE CAN SAY WITH 95% CONFIDENCE IS THAT THE AVERAGE DAMAGE WILL LIE BETWEEN \$168 AND \$912.

THE COMPANY CAN EITHER BE SATISFIED WITH THAT, OR DO FURTHER TESTS...

TO COMPUTE THIS CONFIDENCE INTERVAL USING STUDENT'S t, WE HAVE MADE AN ***UNSTATED ASSUMPTION:*** WE ASSUMED THAT CRASH REPAIR COSTS ARE APPROXIMATELY ***NORMALLY DISTRIBUTED***, I.E., IF WE CRASHED 1000 CHAMELEONS, THE HISTOGRAM OF REPAIR COSTS WOULD BE SYMMETRICAL AND MOUND-SHAPED. WE ***CAN NOT KNOW THIS*** FROM 5 DATA POINTS ALONE... BUT MAYBE YEARS OF EXPERIENCE WITH EARLIER MODELS PROVIDE NORMALLY DISTRIBUTED COST HISTOGRAMS FOR FRONT END REPAIRS: INFORMATION WHICH WOULD TEND TO SUPPORT OUR USE OF STUDENT'S t.



Gonick, Smith 1993

3 Konfidenzintervall für die Varianz

Die Zufallsvariable

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

ist bei normalverteilter Grundgesamtheit χ^2**-verteilt mit $\nu = n - 1$ Freiheitsgraden** (einer der n Freiheitsgrade der Stichprobe wird durch die Verwendung von $\overline{X}$ 'verbraucht').

Werden mit $\chi^2_{\alpha/2,\nu}$ bzw. $\chi^2_{1-\alpha/2,\nu}$ die Punkte bezeichnet, bei denen die Verteilungsfunktion der Chi-Quadrat-Verteilung die Werte $\alpha/2$ bzw. $1-\alpha/2$ annimmt, dann ergibt sich

$$P\left(\chi^2_{\alpha/2,n-1} \le \frac{n \cdot S^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha\,.$$

Als Konfidenzintervall für die unbekannte Varianz ergibt sich dann:

$$P\left(\frac{n \cdot S^2}{\chi^2_{1-\alpha/2,n-1}} \le \sigma^2 \le \frac{n \cdot S^2}{\chi^2_{\alpha/2,n-1}}\right) = 1-\alpha\,.$$

4 Konfidenzintervall für den Anteilswert

Der Anteilswert $\hat{p}$ einer Stichprobe aus einer dichotomen Grundgesamtheit ist bei genügend großem Stichprobenumfang **normalverteilt** (entsprechend der Approximationsbedingungen für die Binomialverteilung und die Hypergeometrische Verteilung) mit

$E(\hat{p}) = p$ und

$Var(\hat{p}) = \frac{p \cdot (1-p)}{n} = \sigma^2_{\hat{p}}$ (*Modell mit Zurücklegen*)

$Var(\hat{p}) = \frac{p \cdot (1-p)}{n} \cdot \frac{N-n}{N-1}$ (*Modell ohne Zurücklegen* und $\frac{n}{N} > 0{,}05$).

Aus der standardisierten Zufallsvariablen

$$Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

läßt sich das **Konfidenzintervall für das unbekannte p** der Grundgesamtheit ableiten:

$$P\left(\hat{p} - z \cdot \sigma_{\hat{p}} \le p \le \hat{p} + z \cdot \sigma_{\hat{p}}\right) = 1-\alpha\,.$$

Da p unbekannt ist, muß auch für $\sigma_{\hat{p}}$ ein Schätzer eingesetzt werden. Damit ergibt sich dann:

Modell mit Zurücklegen

$$P\left(\hat{p} - z \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \le p \le \hat{p} + z \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right) = 1-\alpha\,.$$

Entsprechend gilt für das

Modell ohne Zurücklegen

$$P\left(\hat{p}-z\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}\cdot\frac{N-n}{N-1}}\leq p\leq\hat{p}+z\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}\cdot\frac{N-n}{N-1}}\right)=1-\alpha\,.$$

z ist $z_{1-\alpha/2}$, also abhängig vom gewählten Signifikanzniveau α.

5 Bestimmung des notwendigen Stichprobenumfangs

Bisher wurden Fragen nach μ,σ^2,p bei einem gegebenem Konfidenzniveau $1-\alpha$ beantwortet, z. B.: μ liegt in dem Konfidenzintervall $\left[\overline{X}-z\cdot\sigma_{\overline{X}}\,;\overline{X}+z\cdot\sigma_{\overline{X}}\right]$.

Setzt man nun

$$\Delta\mu=z\cdot\sigma_{\overline{X}}$$

als Maß für die Genauigkeit, so kann man nach dem **notwendigen Stichprobenumfang** fragen:

$$\Delta\mu=z\cdot\sigma_{\overline{X}}=z\cdot\frac{\sigma}{\sqrt{n}}=\text{absoluter Fehler}$$

Daraus ermittelt man:

$$n=\frac{z^2\cdot\sigma^2}{(\Delta\mu)^2}$$

Analog bestimmt man den notwendigen Stichprobenumfang bei Vorliegen einer dichotomen Grundgesamtheit (0/1-Verteilung).

Für das **Konfidenzintervall** ergibt sich:

$$\left[\hat{p}-z\cdot\sigma_{\hat{p}}\,;\hat{p}+z\cdot\sigma_{\hat{p}}\right]$$

Aus der halben Breite des Konfidenzintervalls Δp, was identisch mit dem absoluten Fehler für das Modell mit Zurücklegen ist,

$$\Delta p=z\cdot\sigma_{\hat{p}}=z\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}},$$

lässt sich die gewünschte Stichprobengröße n errechnen:

Notwendiger Stichprobenumfang (Modell mit Zurücklegen)

$$n=\frac{z^2\cdot\hat{p}\cdot(1-\hat{p})}{(\Delta p)^2}\,.$$

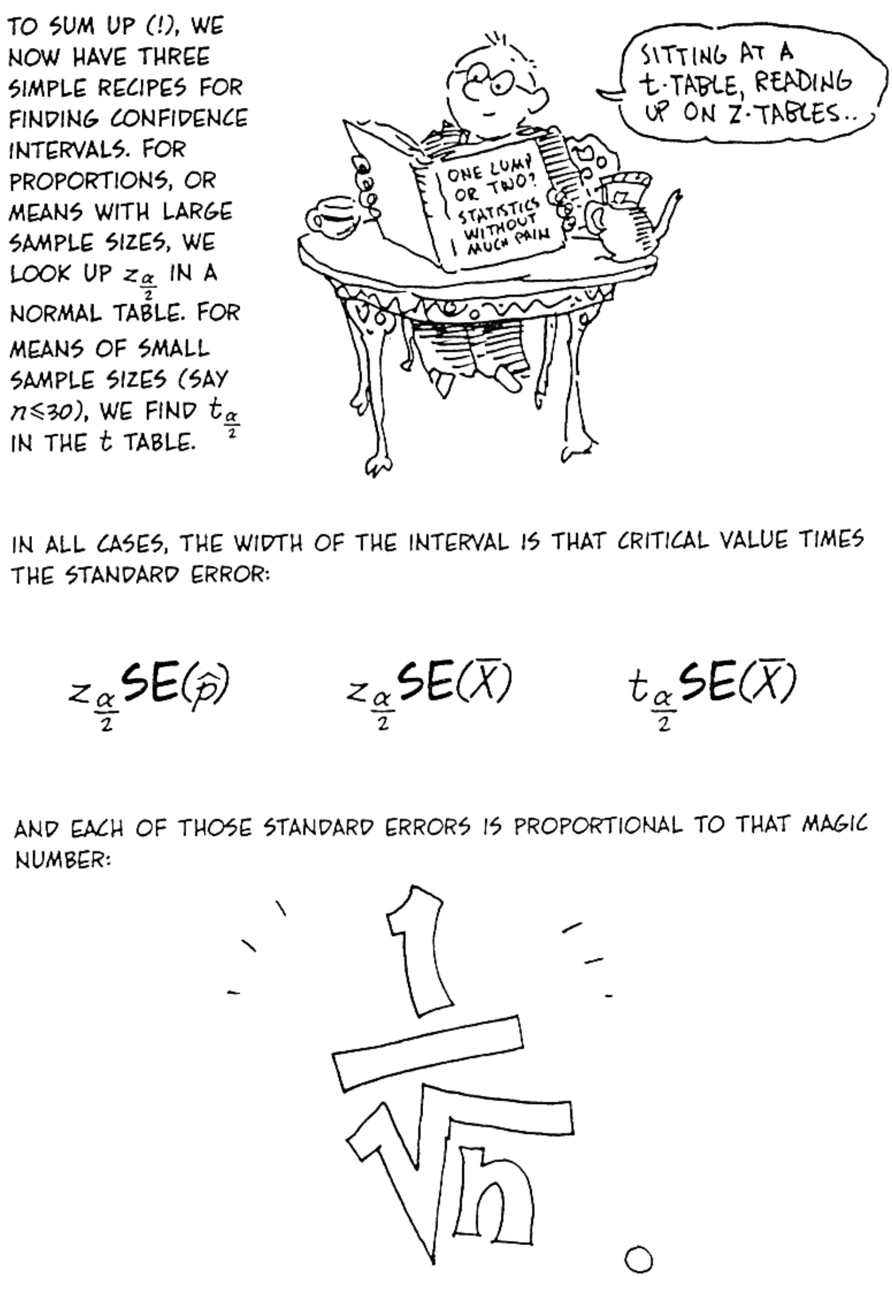
Der absolute Fehler für das Modell ohne Zurücklegen beträgt:

$$\Delta p=z\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\cdot\sqrt{\frac{N-n}{N-1}}.$$

Notwendiger Stichprobenumfang (Modell ohne Zurücklegen)

$$n=\frac{z^2\cdot N\cdot\hat{p}\cdot(1-\hat{p})}{(\Delta p)^2\cdot(N-1)+z^2\cdot\hat{p}\cdot(1-\hat{p})}\,.$$

Als Schätzwert für p kann der Stichprobenanteilswert $\hat{p}$ einer Vorstichprobe oder ein aus früheren Erhebungen bekannter Wert eingesetzt werden. Wenn man keine Vorstellung von $\hat{p}$ besitzt, sollte $\hat{p}=0{,}5$ verwendet werden.



Gonick, Smith 1993

6 Konfidenzintervall für die Differenz zweier arithmetischer Mittel

Aus zwei großen Grundgesamtheiten Nr. 1 und Nr. 2 wird je eine Stichprobe des Umfangs n_1 und n_2 entnommen. Die Stichprobenmittel $\overline{X}_1$ und $\overline{X}_2$ sind Zufallsvariablen.

Falls $\overline{X}_1$ und $\overline{X}_2$ nicht aus normalverteilten Grundgesamtheiten stammen, sind sie unter den beiden Annahmen

- **Unabhängigkeit** der beiden Stichproben
- genügend **große Stichprobenumfänge** (Faustregel: $n_1 > 30, n_2 > 30$)

dann nach dem zentralen Grenzwertsatz normalverteilt mit dem

Erwartungswert

$$E(D) = E\left(\overline{X}_1 - \overline{X}_2\right) = \mu_1 - \mu_2$$

und der Varianz

$$Var(D) = Var\left(\overline{X}_1 - \overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Beweis zur Varianz:

$$\begin{aligned} Var\left(\overline{X}_1 - \overline{X}_2\right) &= E\left(\left\{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)\right\}^2\right) \\ &= E\left(\left\{\left(\overline{X}_1 - \mu_1\right) - \left(\overline{X}_2 - \mu_2\right)\right\}^2\right) \\ &= E\left[\left(\overline{X}_1 - \mu_1\right)^2\right] + E\left[\left(\overline{X}_2 - \mu_2\right)^2\right] - 2 \cdot E\underbrace{\left[\left(\overline{X}_1 - \mu_1\right) \cdot \left(\overline{X}_2 - \mu_2\right)\right]}_{\substack{=0 \text{ bei Unabhängigkeit} \\ \text{der Stichproben}}} \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma_D^2 \qquad \text{q.e.d.} \end{aligned}$$

Für die normalverteilte Zufallsvariable D gilt dann:

$$P\left(E(D) - z \cdot \sigma_D \leq D \leq E(D) + z \cdot \sigma_D\right) = 1 - \alpha$$

bzw. ausführlicher:

$$P\left(\left(\overline{X}_1 - \overline{X}_2\right) - z \cdot \sigma_D \leq \mu_1 - \mu_2 \leq \left(\overline{X}_1 - \overline{X}_2\right) + z \cdot \sigma_D\right) = 1 - \alpha .$$

Sind σ_1^2 und σ_2^2 unbekannt, dann muß σ_D^2 geschätzt werden.

Falls $\sigma_1^2 = \sigma_2^2$, dann ergibt sich:

$$S^2 = \frac{1}{n_1 + n_2}(n_1 \cdot S_1^2 + n_2 \cdot S_2^2) \text{ bzw.}$$

$$\hat{\sigma}^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2} \cdot S^2 = \frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2}$$

$$\hat{\sigma}_D^2 = \hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= \hat{\sigma}^2 \left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right)$$

$$= \frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right)$$

Als Konfidenzintervall erhält man:

$$P\left(\left(\bar{X}_1 - \bar{X}_2 \right) - t \cdot \hat{\sigma}_D \le \mu_1 - \mu_2 \le \left(\bar{X}_1 - \bar{X}_2 \right) + t \cdot \hat{\sigma}_D \right) = 1 - \alpha$$

mit $t = t_{1-\alpha/2, n_1 + n_2 - 2}$.

Beispiele:

- Heilungswahrscheinlichkeit zweier verschiedener Medikamente;
- Ausschußwahrscheinlichkeit zweier Maschinen;
- mittlere Erträge bei Verwendung zweier (natürlicher!) Düngemittel.

7 Konfidenzintervall für die Differenz zweier Anteilswerte

Die Differenz zweier voneinander unabhängigen Stichprobenanteilswerten ist bei genügend großen Stichprobenumfängen normalverteilt mit

$$E(D) = p_1 - p_2$$

$$Var(D) = Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}.$$

Damit erhält man als **Konfidenzintervall**:

$$P\left(\left(\hat{p}_1 - \hat{p}_2 \right) - z \cdot \sigma_D \le p_1 - p_2 \le \left(\hat{p}_1 - \hat{p}_2 \right) + z \cdot \sigma_D \right) = 1 - \alpha .$$

Als Schätzfunktion für σ_D wird verwendet

$$\hat{\sigma}_D^2 = \frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}.$$

Voraussetzung für die Normalverteilung ist hier:

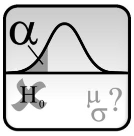
$$n_1 \cdot p_1 \cdot (1 - p_1) \ge 9$$

$$n_2 \cdot p_2 \cdot (1 - p_2) \ge 9.$$

Keyconcepts

Vertrauensbereich/Konfidenzbereich
Konfidenzwahrscheinlichkeit
Irrtumswahrscheinlichkeit
Konfidenzintervalle für arithmetisches Mittel, Varianz und Anteilswerte
Notwendiger Stichprobenumfang
Konfidenzintervalle für die Differenz zweier arithmetischer Mittel und zweier Anteilswerte

VIII Parametertests



Überprüfung von Vermutungen über die Grundgesamtheit mit Hilfe der Stichprobe

Schätztheorie: Für unbekannte (Verteilungs-)Parameter einer Grundgesamtheit (GG) wird versucht, mithilfe von Stichprobenergebnissen einen numerischen Wert zu schätzen (**Punktschätzung**), bzw. ein Intervall anzugeben, das den wahren Wert mit einer vorgegebenen Sicherheitswahrscheinlichkeit $(1-\alpha)$ enthält (**Intervallschätzung**).

Testtheorie: Hier geht es um die Frage, ob Annahmen (**Hypothesen**) für eine Grundgesamtheit durch ein Stichprobenergebnis mit einer vorgegebenen **Irrtumswahrscheinlichkeit α abgelehnt werden müssen oder nicht**.

- **Parametertests**:

  Überprüfung von Hypothesen über unbekannte Parameter einer Grundgesamtheit (üblicherweise als θ (Theta) bezeichnet).

- **Verteilungstests**:

  Überprüfung von Hypothesen über die unbekannte Verteilungsform einer Grundgesamtheit.

1 Methodische Grundlagen der Testtheorie

1.1 Prinzip und Aufbau eines statistischen Tests

Das Prinzip eines statistischen Testes soll anhand der folgenden Geschichte erläutert werden, die von R. A. FISHER stammt (Hochstädter 1991, S. 573):

Bei einer Gesellschaft behauptet eine Dame X: Setze man ihr eine Tasse Tee vor, der etwas Milch beigegeben wurde, so könne sie im allgemeinen einwandfrei schmecken, ob zuerst Tee oder ob zuerst Milch eingegossen worden sei. Wie prüft man diese Behauptung?

Sicher nicht so: Zwei äußerlich völlig gleichartige Tassen vorsetzen, wobei in die erste zuerst Milch und dann Tee (Reihenfolge MT) und in die zweite zuerst Tee und dann Milch (TM) eingegossen wurde. Würde man jetzt die Dame wählen lassen, so hätte sie offenbar eine Chance von 50 %, die richtige Antwort zu geben, auch wenn ihre Behauptung falsch ist.

Besser ist folgendes Vorgehen: Acht äußerlich gleiche Tassen werden der Dame vorgesetzt. Davon sind vier in der Reihenfolge TM, die anderen vier in der Reihenfolge MT gefüllt worden. Die Tassen werden zufällig über den Tisch verteilt und danach die Dame herbeigerufen. Man teilt ihr mit, daß von den Tassen je vier vom Typ TM bzw. MT seien. Ihre Aufgabe bestehe darin, die vier TM Tassen herauszufinden. Jetzt ist die Wahrscheinlichkeit, ohne die Sonderbegabung die richtigen vier Tassen zu finden, sehr gering geworden. Aus acht Tassen kann man nämlich auf

$\binom{8}{4} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2} = 70$ verschiedene Möglichkeiten vier Tassen auswählen.

Von diesen 70 Kombinationen ist aber nur eine einzige die Richtige. Die Wahrscheinlichkeit, ohne Sonderbegabung, also zufällig, die richtige Kombination zu treffen, beträgt daher nach Laplace 1/70=0,0143 oder etwa 1,4 %, sie ist also sehr gering. Wählt die Dame nun wirklich die richtigen vier Tassen aus, so wird man die Nullhypothese – H_0: die Dame besitzt diese Sonderbegabung nicht – verwerfen und ihr diese besondere Fähigkeit zuerkennen. Dabei nimmt man eine Irrtumswahrscheinlichkeit von 1,4 % in Kauf. Natürlich könnte man diese Irrtumswahrscheinlichkeit dadurch verringern, daß man die Anzahl der Tassen erhöhen würde, z.B. auf 12, wobei wieder die Hälfte nach TM bzw. MT gefüllt wäre. In diesem Fall würde die Irrtumswahrscheinlichkeit auf $\alpha \approx 0{,}1$ % sinken.

Prinzip des statistischen Tests:

Aufstellen einer Nullhypothese H_0, die verworfen wird, wenn ein Ergebnis beobachtet wird, das bei Gültigkeit dieser Nullhypothese unwahrscheinlich ist. Desweiteren wird eine Alternativhypothese H_A formuliert, die alle möglichen Ergebnisse umfasst, die nicht von H_0 abgedeckt werden. Ein Beispiel: Eine Nullhypothese „Der unbekannte Parameter ist positv oder Null.“ ($H_0 : \theta \geq 0$) würde als entsprechende Gegenhypothese die Aussage „Der unbekannte Parameter ist negativ.“ ($H_A : \theta < 0$) nach sich ziehen.

Aufbau eines statistischen Tests in 5 Schritten:

1. Aufstellen der Parametermenge Θ, der Nullhypothese H_0, der Alternativhypothese H_A; Festlegung des Signifikanzniveaus.

↓

2. Festlegung einer geeigneten Prüfgröße und Bestimmung der Testverteilung unter H_0.

↓

3. Bestimmung des kritischen Bereichs.

↓

4. Berechnung der Prüfgröße.

↓

5. Entscheidung und Interpretation.

Zur Entscheidung und Interpretation: Fällt das Stichprobenergebnis in den kritischen Bereich mit Prob(Ereignis|H_0) $\leq \alpha$, dann wird H_0 abgelehnt (α entspricht dem Anteil, mit dem im Durchschnitt H_0 abgelehnt wird). Die Schritte 2 und 4 können auch zusammengefasst werden.

1.2 Grundbegriffe der Testtheorie

Wir unterscheiden wir zwischen einfachen und zusammengesetzten Hypothesen. Dabei heißt eine Hypothese einfach, wenn sie genau ein Element enthält und zusammengesetzt, wenn sie mehr als ein Element enthält.

$H_0: \theta = \theta_0$ einfache Nullhypothese
$H_A: \theta \neq \theta_0$ zusammengesetzte Alternativhypothese
$H_0: \theta \geq$ oder $< \theta_0$ zusammengesetzte Alternativhypothese

Einführendes Beispiel:

CHIO CHONG produziert Computerchips. Als Produzent eines Massenartikels behauptet CHIO CHONG, daß der Ausschußanteil einer Lieferung höchstens 10 % betrage.

Ausgangshypothese, Nullhypothese: $H_0: p \leq 10\%$

Alternativhypothese, Gegenhypothese: $H_A(H_1): p > 10\%$

Je nach Problemstellung werden Null- und Alternativhypothese wie folgt bestimmt:

1. Fall: Die Parametermenge Θ (= Menge aller möglichen Parameter) **hat nur zwei Elemente**: $\Theta = \{\theta_1, \theta_2\}$

Beispiel:

Zwei Maschinen produzieren dasselbe Gut mit unterschiedlichen Ausschußanteilen von 2 % bzw. 5 %. Mit einer Stichprobe soll entschieden werden, von welcher Maschine eine Lieferung stammt.

$\Theta = \{2\%; 5\%\}$

H_0: Lieferung von der Maschine mit 2 % Ausschuß,

H_A: Lieferung von der Maschine mit 5 % Ausschuß.

Allgemein:

H_0: Der wahre (aber unbekannte) Parameter θ ist θ_1.

H_A: Der wahre (aber unbekannte) Parameter θ ist nicht θ_1, sondern θ_2.

Formal:

$H_0: \theta = \theta_1$

$H_A: \theta = \theta_2$

Man spricht hier von einer **einfachen Null- bzw. Alternativhypothese**.

2. Fall: Die Parametermenge Θ hat mehr als zwei Elemente

z. B. $\Theta = \mathrm{IR}$, $\Theta = \mathrm{IR}_0^+$, $\Theta = [0,1]$, ...

Generell muß nun zwischen zweiseitigen und einseitigen Fragestellungen unterschieden werden.

a) Zweiseitige Fragestellung ('zweiseitiger Test'):

H_0: Der wahre Parameter θ ist θ_0.

H_A: Der wahre Parameter θ ist nicht θ_0.

Formal:

H_0: $\theta = \theta_0$ (einfache Nullhypothese)

H_A: $\theta \neq \theta_0$ (zusammengesetzte Alternativhypothese).

Beispiel:

Die Fa. Bossmann hat eine Maschine, die Apfelwein in Flaschen abfüllt. Als sie neu war, betrug die durchschnittliche Menge Apfelwein in einer Flasche 1,0 Liter mit einem Standardfehler von 0,001 Liter. Nach einem Jahr wird überprüft, ob die Durchschnittsfüllmenge noch 1,0 Liter beträgt. Dazu hat die Firma eine Stichprobe von n = 30 Flaschen aus der laufenden Produktion entnommen.

$\Theta = R_0^+$

H_0: $\theta = 1{,}000\ell$

H_A: $\theta \neq 1{,}000\ell$

---

b) Einseitige Fragestellung ('einseitiger Test'):

$H_0: \theta \leq \theta_0$ (zusammengesetzte Nullhypothese)

$H_A: \theta > \theta_0$ (zusammengesetzte Alternative)

bzw.

$H_0: \theta \geq \theta_0$

$H_A: \theta < \theta_0$

Beispiel:

CHIO CHONG mit $\Theta = [0,1]$ und $H_0: \theta \leq 0{,}1$ sowie $H_A: \theta > 0{,}1$.

1.3 Fehlermöglichkeiten bei statistischen Tests

Die Entscheidungen bei einem statistischen Test basieren auf stochastischen Ereignissen, d. h., es besteht das **Risiko einer Fehlentscheidung**. Es ist daher wichtig, die **Wahrscheinlichkeiten solcher Fehlentscheidungen** (für den langfristigen Durchschnitt) näher zu betrachten.

Aufgrund einer Stichprobe wird H_0 entweder abgelehnt (d. h., H_A ist statistisch gesichert) oder H_0 beibehalten. In beiden Fällen kann die Entscheidung richtig oder falsch sein. Damit ergeben sich die in der folgenden Übersicht dargestellten vier Möglichkeiten:

In Wirklichkeit gilt	Entscheidung	
	H_0 ablehnen	H_0 beibehalten
H_0 ist richtig	**Fehler 1. Art (α)**	kein Fehler
H_0 ist falsch	kein Fehler	**Fehler 2. Art (β)**

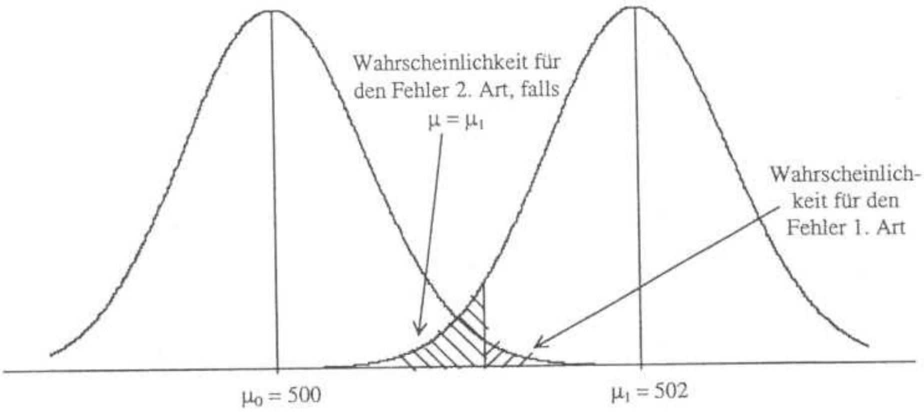
Fehler 1. Art:

Man spricht von einem Fehler 1. Art, wenn H_0 abgelehnt wird, obwohl H_0 in Wirklichkeit richtig ist. Die Wahrscheinlichkeit für diesen Fehler soll kleiner oder gleich einem vorgegebenen Signifikanzniveau α sein. Bei einem Test wird die Wahrscheinlichkeit für einen Fehler 1. Art vom Prüfenden festgelegt.

Fehler 2. Art:

Man spricht von einem Fehler 2. Art, wenn H_0 beibehalten wird, obwohl H_A in Wirklichkeit richtig ist. Die Wahrscheinlichkeit für diesen Fehler wird mit β bezeichnet.

Die Größe des β-Fehlers hängt dabei von dem Abstand zwischen der „wahren" und der unter der Nullhypothese angenommenen Verteilung ab, wie die folgende Grafik zeigt. Unter der Nullhypothese wurde hierbei eine Verteilung um den vermuteten Mittelwert $\mu_0 = 500$ angenommen. Der wahre Mittelwert liegt jedoch bei $\mu_1 = 502$. Der β-Fehler ergibt sich nun als die linke Schnittfläche der beiden Verteilungen.



Da nun in der Praxis der wahre Parameter unbekannt ist (diesen versucht man ja gerade über den Test zu ermitteln), ist der β-Fehler, der ja vom wahren Parameter abhängt, in der Regel unbekannt.

Es lässt sich allerdings zeigen, daß der folgende Zusammenhang gilt:

> Je kleiner die Wahrscheinlichkeit für einen Fehler 1. Art ist, desto größer ist die Wahrscheinlichkeit für einen Fehler 2. Art.

Grafik: Eine Verkleinerung des Fehlers 1. Art würde bedeuten, dass die Grenze der rechten schraffierten Fläche nach rechts verschoben wird. Da der Fehler 2. Art (die linke schraffierte Fläche) direkt an dieser Grenze anschließt, vergrößert sich der β-Fehler entsprechend.

Es ist festzuhalten, daß die Wahrscheinlichkeit, H_0 abzulehnen, obwohl H_0 richtig ist, maximal α beträgt (Fehler 1. Art). Diese Wahrscheinlichkeit kann durch den Untersuchenden festgelegt werden und sollte möglichst klein sein, so dass die Entscheidung für eine Ablehnung von H_0 ziemlich sicher ist.

Die Wahrscheinlichkeit, H_0 anzunehmen, obwohl H_0 falsch ist, hat aber einen unbekannten, möglicherweise sehr großen Wert. Eine Entscheidung für H_0 ist daher nicht annähernd so sicher wie eine Entscheidung für H_A.[5] Praktiker der empirischen Wirtschafts- und Sozialfor-

---

5 Bamberg und Baur (1991) betonen, daß, wenn H_0 nicht abgelehnt werden kann, dies nicht bedeutet, daß H_0 bestätigt ist. Vielmehr reichen in diesem Fall die Beobachtungsdaten nicht zu einer Ablehnung von H_0 aus (sozusagen eine Stimmenenthaltung oder ein Freispruch aus Mangel an Beweisen). Vgl. hierzu Bamberg und Baur 1991, S. 182.

schung wählen daher diejenige Hypothese als Gegenhypothese H_A, die sie „bestätigen“ oder „statistisch untermauern“ wollen.

Generelles Ziel eines Testverfahrens ist also: Ein vorgegebenes Signifikanzniveau α (Irrtumswahrscheinlichkeit) ist einzuhalten, wobei β möglichst klein sein sollte.

1.4 Testentscheidung: Kritischer Wert und p-value

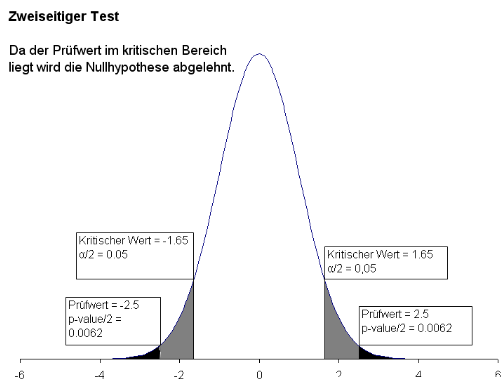
Das Signifikanzniveau α gibt an, welche Irrtumswahrscheinlichkeit höchstens zugelassen wird, einen Fehler 1. Art zu begehen. Bei einem Signifikanzniveau von $\alpha = 0{,}05$ darf die Wahrscheinlichkeit H_0 abzulehnen, obwohl diese richtig ist, höchstens 5% betragen.

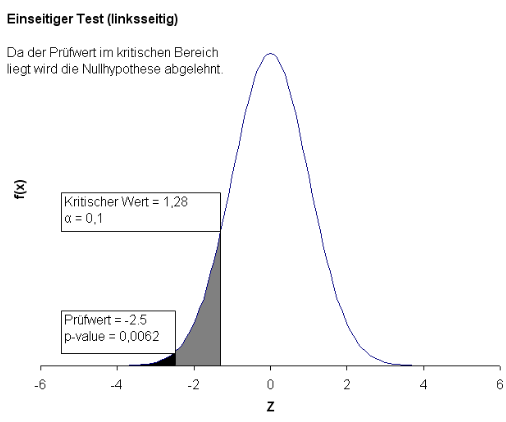
Aus dem Signifikanzniveau α lässt sich mit Hilfe einer Testverteilung aus einer Tabelle ein **kritischer Testwert** ermitteln. Ein Vergleich des kritischen Testwerts mit der Prüfgröße aus der Stichprobe führt zur Testentscheidung. Liegt die Prüfgröße im kritischen Bereich, so wird H_0 abgelehnt. Liegt die Prüfgröße im Konfidenzbereich (außerhalb des kritischen Bereichs), so kann H_0 nicht abgelehnt werden.

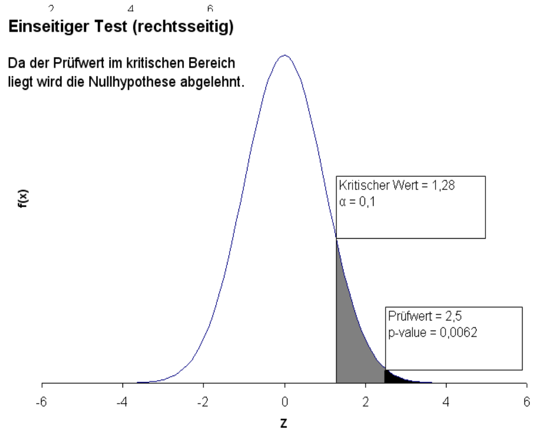
Die Testentscheidung kann aber auch über den probability-value (**p-value**) getroffen werden. Der p-value gibt für einen Prüfwert der Stichprobe die Wahrscheinlichkeit an, den Fehler 1. Art zu begehen. Ist der p-value kleiner als die zugelassene Irrtumswahrscheinlichkeit α, kann H_0 abgelehnt werden. Für einen p-value größer als die zugelassene Irrtumswahrscheinlichkeit muss H_0 beibehalten werden.

Testentscheidung	**Kritischer Wert**	**p-value**
H_0 **nicht ablehnen**	$\lvert$Prüfgröße$\rvert$ < $\lvert$Kritischer Wert$\rvert$	p-value > Signifikanzniveau α
H_0 **ablehnen**	$\lvert$Prüfgröße$\rvert$ > $\lvert$Kritischer Wert$\rvert$	p-value < Signifikanzniveau α

Grafisch lässt sich das Signifikanzniveau α interpretieren als die Fläche unterhalb der jeweiligen Wahrscheinlichkeits(Dichte-)funktion ab diesem kritischen Wert.
Der p-value wird durch die Fläche ab dem Prüfwert dargestellt. Für eine Standardnormalverteilung ergeben sich beispielsweise folgende Bilder:







1.5 Beurteilungskriterien für statistische Tests: Gütefunktion und Operationscharakteristik

Statistische Tests sollen zwei Kriterien erfüllen:

1. Die Wahrscheinlichkeit, den Fehler 1. Art (die Nullhypothese wird fälschlicherweise abgelehnt) zu begehen, darf höchstens α betragen.
2. Die Wahrscheinlichkeit den Fehler 2. Art (die Nullhypothese wird fälschlicherweise beibehalten) zu begehen, soll unter Geltung der 1. Bedingung möglichst gering sein.

Grundsätzlich kann der Fehler 1. Art nur dann begangen werden, wenn der unbekannte, wahre Parameter θ im Bereich der Nullhypothese (Θ_0) liegt; der Fehler 2. Art nur, wenn θ außerhalb dieses Bereichs liegt. Die Wahrscheinlichkeit einen dieser beiden Fehler zu begehen hängt also vom Wert des wahren Parameters θ ab.

Um zu beurteilen, inwieweit ein Parametertest den o.g. Anforderungen genügt, wird die **Gütefunktion (**Teststärke, power des Tests) sowie die **Operationscharakteristik** herangezogen. Die Wahrscheinlichkeiten für den Fehler 1. und 2. Art können mit Hilfe dieser Funktionen berechnet werden.

Liegt der wahre Wert im **Bereich der Nullhypothese** ($\theta \in \Theta_0$), gibt die Gütefunktion die Wahrscheinlichkeit für einen Fehler 1. Art an:

$G(\theta) = P("H_A"|H_0) \leq \alpha$ für alle $\theta \in \Theta_0$.

Gütefunktion: Die Gütefunktion gibt die Wahrscheinlichkeit einer Ablehnung der Nullhypothese (in Abhängigkeit von möglichen Parameterwerten θ) an: $G(\theta) = P(\text{Prüfgröße im Ablehnungsbereich von } \mathrm{H}_0 | \theta) = P("H_A"|\theta)$.

Liegt der wahre Wert im **Bereich der Alternativhypothese** ($\theta \in \Theta_A$**),** dann gibt die **Operationscharakteristik** die Wahrscheinlichkeit für einen Fehler 2. Art an:

$OC(\theta) = 1 - G(\theta) = P("H_0"|H_A) = \beta$ für alle $\theta \in \Theta_\mathrm{A}$.

Operationscharakteristik (OC): Die Operationscharakteristik ist die Wahrscheinlichkeit der Nichtablehnung der Nullhypothese, wiederum in Abhängigkeit von den möglichen Parameterwerten θ:

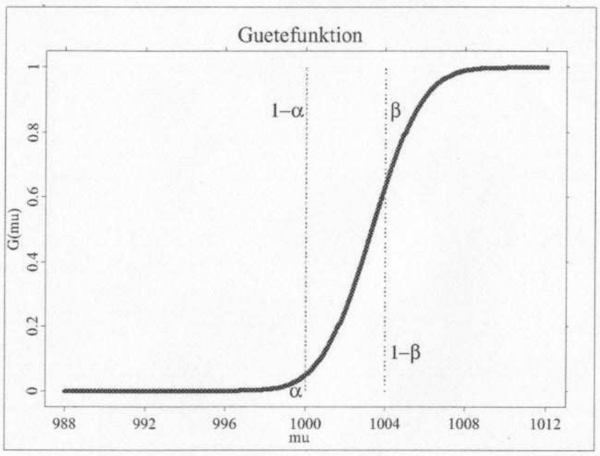
$OC(\theta) = 1 - G(\theta) = P(\text{Prüfgröße nicht im Ablehnungsbereich von } \mathrm{H}_0 | \theta) = P("H_0"|\theta)$.

Aus Gütefunktion und Operationscharakteristik wird deutlich: Liegt der wahre Parameter θ im Bereich der Nullhypothese ($\theta \in \Theta_A$**),** so ist die Wahrscheinlichkeit für den Fehler 1. Art höchstens α. Liegt θ im Bereich der Alternativhypothese, so nimmt die Wahrscheinlichkeit für den Fehler 2. Art mit zunehmenden Abstand von θ zur H_0 ab.

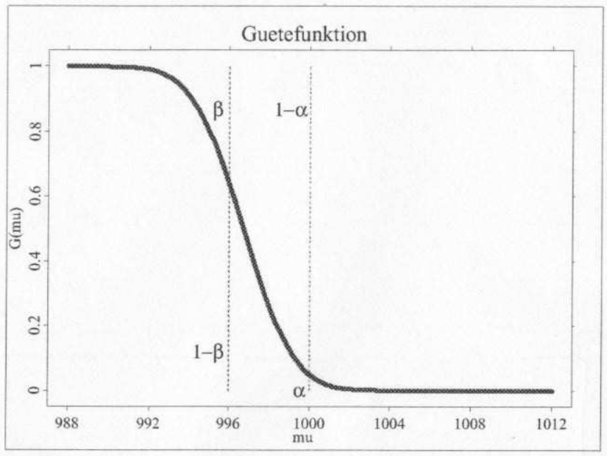
Beispiele:

Nachstehend findet sich die grafische Darstellung der Gütefunktion für einen rechtsseitigen, linksseitigen sowie einen zweiseitigen Parametertest. Hierbei ist die Wahrscheinlichkeit den Fehler 1. Art zu begehen (y-Achse) für verschiedene Werte von θ (x-Achse) abgetragen. Die Wahrscheinlichkeit für den Fehler 2. Art ergibt sich dann aus $1 - G(\theta)$.

Test rechtsseitig $H_0: \mu \leq 1000; H_A > 1000$



Test linksseitig $H_0: \mu \geq 1000; H_A: \mu < 1000$:



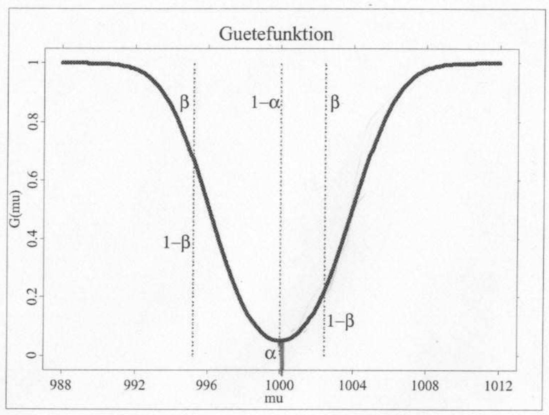
Rechtsseitiger Test $H_0: \mu \leq 1000; H_A > 1000$

Ist der wahre Wert θ kleiner als 1000, liegt die Wahrscheinlichkeit den Fehler 1. Art zu begehen unter α. Beträgt θ=1000, so ist die Wahrscheinlichkeit für diesen Fehler genau α. Die Wahrscheinlichkeit für den Fehler 1. Art lässt sich also als Ordinatenwert (Strecke) unterhalb der Gütefunktion ablesen. Nimmt θ jedoch einen Wert über 1000 an, besteht die Gefahr einen Fehler 2. Art zu begehen. Die Wahrscheinlichkeit für den Fehler 2. Art ergibt sich aus $1-G(\theta)$. Dies ist grafisch als Strecke oberhalb der Gütefunktion abzulesen. Es wird deutlich, dass die Wahrscheinlichkeit für den Fehler 2. Art sinkt, wenn θ zunimmt.

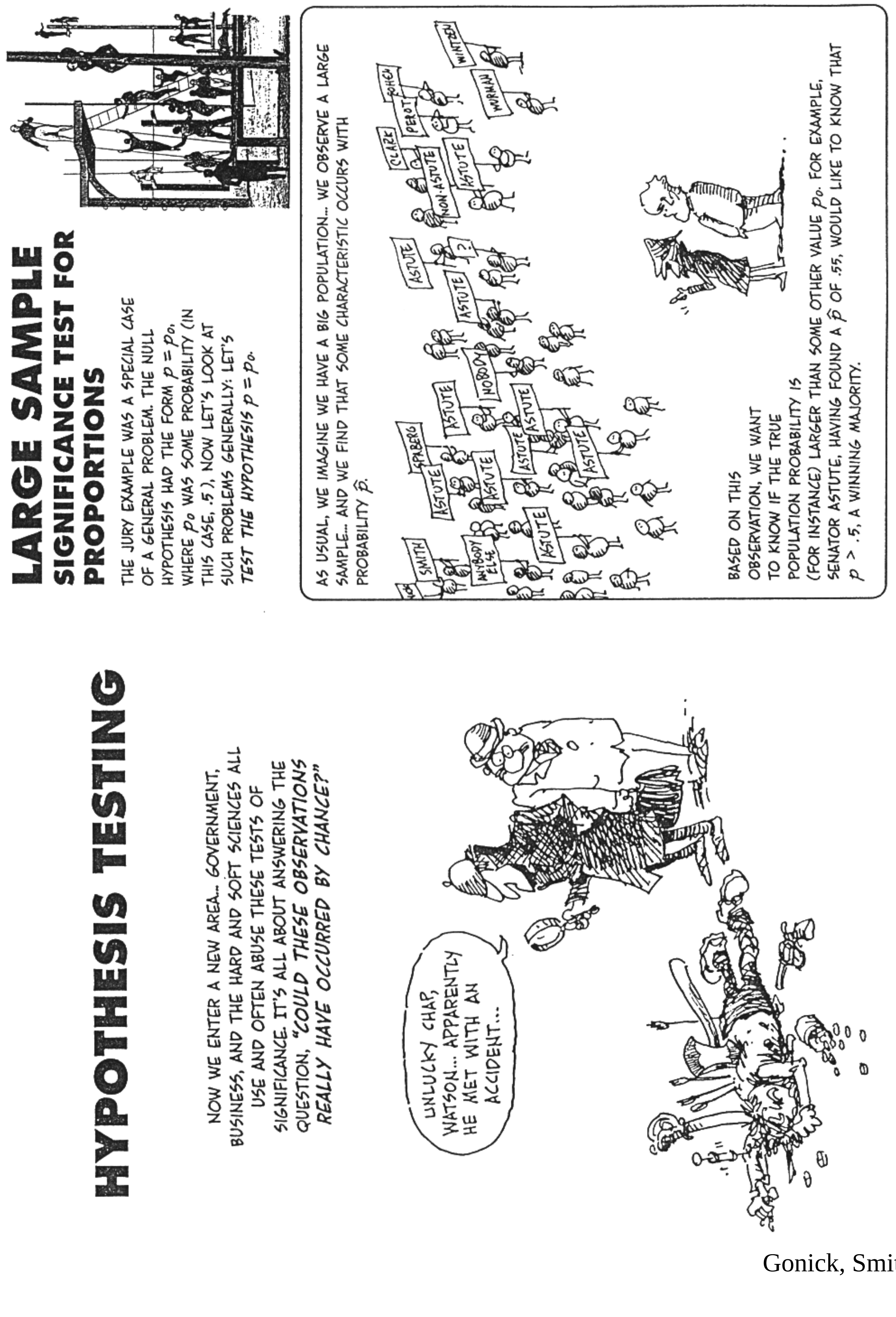
Linksseitiger Test $H_0: \mu \geq 1000; H_A: \mu < 1000$

Bei einer linksseitigen Fragestellung ist die Argumentation genau umgekehrt. Nimmt θ Werte über 1000 an, liegt die Wahrscheinlichkeit für den Fehler 1. Art unter α. Für Werte kleiner als 1000 besteht die Wahrscheinlichkeit für den Fehler 2. Art. Je niedriger die Werte von θ sind, desto geringer ist die Wahrscheinlichkeit einen Fehler 2. Art zu begehen.

Zweiseitiger Test $H_0: \mu = 1000; H_A: \mu \neq 1000$:



In diesem Fall beträgt die Wahrscheinlichkeit einen Fehler 1. Art zu begehen genau α, wenn der wahre Wert 1000 ist. Ist θ ungleich 1000 ist es nur noch möglich einen Fehler 2. Art zu begehen. Wie bei den anderen Tests wird der Fehler 1. Art unterhalb der Gütefunktion und der Fehler 2. Art oberhalb der Gütefunktion abgelesen.



Gonick, Smith 1993

Step 1.

THE *NULL* HYPOTHESIS IS

$H_0 : p = p_0$

THE ***ALTERNATE*** HYPOTHESIS DEPENDS ON THE DIRECTION OF THE EFFECT WE ARE LOOKING FOR. IN SENATOR ASTUTE'S CASE,

$H_a : p > p_0$

BUT IN OTHER CASES, THE ALTERNATE HYPOTHESIS MIGHT WELL BE

$H_a : p < p_0$

OR

$H_a : p \neq p_0$

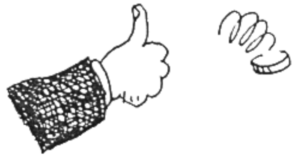
FOR EXAMPLE, IN THE JURY SELECTION EXAMPLE, THE ALTERNATIVE HYPOTHESIS WAS

$H_a : p < 0.5$

AND AT OTHER TIMES, WE ARE INTERESTED IN KNOWING THAT p IS DIFFERENT FROM SOME VALUE p_0. FOR INSTANCE, IN TESTING FOR A FAIR COIN, WE HAVE AN ALTERNATE HYPOTHESIS OF

$H_a : p \neq 0.5$

BUT HAVE NO *A PRIORI* OPINION ABOUT WHETHER HEADS OR TAILS WILL COME UP MORE OFTEN.



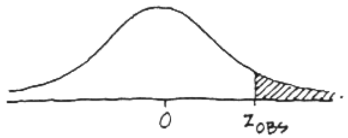
Step 2. THE TEST STATISTIC IS

$$z_{OBS} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}}$$

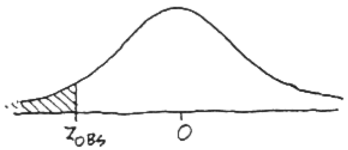
WHICH MEASURES HOW FAR p DEVIATES FROM p_0. UNDER THE NULL HYPOTHESIS, z_{OBS} HAS THE STANDARD NORMAL DISTRIBUTION.

Step 3. THE P-VALUE DEPENDS ON WHICH ALTERNATE HYPOTHESIS IS RELEVANT:

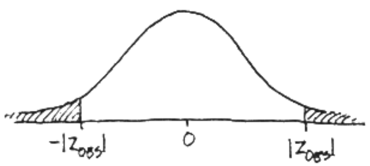
a) "RIGHT-HANDED" $H_a : p > p_0$ USES P-VALUE $Pr(Z > z_{OBS})$



b) "LEFT-HANDED" $H_a : p < p_0$ USES P-VALUE $Pr(Z < z_{OBS})$



c) "TWO-SIDED" $H_a : p \neq p_0$ USES P-VALUE $Pr(|Z| > |z_{OBS}|)$



IN THE CASE OF SENATOR ASTUTE:

1) THE HYPOTHESES ARE

$H_0 : p = .5$

$H_a : p > .5$

2) HIS TEST STATISTIC IS

$$z_{OBS} = \frac{.55 - .50}{\sqrt{(.5)(.5)}/\sqrt{1000}} = 3.16$$

3) HIS P-VALUE IS

$Pr(Z > z_{OBS}) = Pr(z \geq 3.16) = .0008$

(FROM THE NORMAL TABLE).

4) ASTUTE, BEING FAIRLY CONSERVATIVE, TAKES A SIGNIFICANCE LEVEL α OF .01 AND OBSERVES THAT

$Pr(Z > z_{OBS}) = .0008 < \alpha$

THE SENATOR THUS REJECTS THE NULL HYPOTHESIS, AND HE (AND HIS BACKERS) NOW FEEL CERTAIN HE'S IN THE LEAD.

Gonick, Smith 1993

2 Einstichprobentest für den Anteilswert

Unter einem Einstichprobentest versteht man einen statistischen Test auf der Grundlage des Ergebnisses *einer* Stichprobe.

2.1 Einfache Hypothese und einfache Alternative

Beispiel zur Testentwicklung:

Ein Schraubenproduzent betreibt zwei Maschinen, deren Ausschußanteile 20 % bzw. 50 % betragen. In einer Stichprobe vom Umfang n = 20 wurden x = 6 defekte Schrauben festgestellt.

Frage: Wird damit die Behauptung, die Lieferung stamme von der Maschine mit p = 0,2 bei einem Signifikanzniveau von $\alpha = 0{,}05$ widerlegt?

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Prüfgröße Parametermenge: $\Theta = \{0{,}2; 0{,}5\}$

Nullhypothese: $H_0: p = 0{,}2$

Alternativhypothese: $H_A: p = 0{,}5$

Signifikanzniveau: $\alpha = 0{,}05$

2. Schritt: Testverteilung

Die Stichprobe ist aus einer dichotomen Grundgesamtheit. Die Anzahl der defekten Schrauben in der Stichprobe ist (n/N klein) binomialverteilt[6] mit n = 20 und p = 0,2 oder p = 0,5.

Die Prüfgröße X = 'Anzahl der defekten Schrauben' ist für kleines n/N unter H_0 binomialverteilt mit n = 20 und p = 0,2:

$X \sim B(20; 0{,}2)$

3. Schritt: Kritischer Bereich

Sind in der Stichprobe wenig defekte Schrauben, so ist $H_0: p = 0{,}2$ plausibel. Werden dagegen viele defekte Schrauben gezählt, so ist H_0 zugunsten der Alternative $H_A: p = 0{,}5$ abzulehnen.

Genügen x = 6 defekte Schrauben schon, um sich für H_A zu entscheiden? Wo ist die Grenze zwischen der Entscheidung für H_A und der Beibehaltung von H_0 zu ziehen?

Gesucht wird hier also ein kritischer Wert x_c, der für folgende Regel eingesetzt werden soll:

Lehne H_0 ab, falls $x \geq x_c$;

Lehne H_0 nicht ab, falls $x < x_c$.

---

6 Bei dem beschriebenen Experiment handelt es sich um ein Ziehen ohne Zurücklegen. Voraussetzung für die Anwendung der Binomialverteiung ist aber die unabhängige Wiederholung eines Zufallsexperiments (Ziehen mit Zurücklegen). Näherungsweise gilt dies aber auch noch beim Ziehen ohne Zurücklegen, wenn der Auswahlsatz n/N vernachlässigbar klein ist, d.h. N weit größer als n ist.

Da als Signifikanzniveau $\alpha = 0{,}05$ gewählt wurde, muß gelten:

$$P\left(X \geq x_c \middle| p=0{,}2\right) \leq 0{,}05$$

Zusätzlich sollte x_c so klein sein, daß α möglichst ausgeschöpft wird.

X ist unter H_0 B(20;0,2)-verteilt, d.h.

$$\sum_{k=x_c}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \leq \alpha$$

im Beispiel:

$$\sum_{k=x_c}^{\infty} \binom{20}{k} \cdot 0{,}2^k \cdot 0{,}8^{20-k} \leq 0{,}05$$

$$\Leftrightarrow \sum_{k=0}^{x_c-1} \binom{20}{k} \cdot 0{,}2^k \cdot 0{,}8^{20-k} \geq 0{,}95 \quad (*)$$

Gesucht wird also das kleinste x_C, das (*) erfüllt.

Aus der Tabelle ermittelt man:

$$\sum_{k=0}^{6} \binom{20}{k} \cdot 0{,}2^k \cdot 0{,}8^{20-k} = F(X \leq 6 \mid p = 0{,}2) = 0{,}9133 < 0{,}95$$

$$\sum_{k=0}^{7} \binom{20}{k} \cdot 0{,}2^k \cdot 0{,}8^{20-k} = F(X \leq 7 \mid p = 0{,}2) = 0{,}9679 \geq 0{,}95$$

Also $7 = x_c - 1 \Rightarrow x_c = 8$

Annahmebereich: {0, 1,..., 7}

Kritischer Bereich: {8, 9, 10,...}

4. Schritt: Wert der Prüfgröße

x = 6 defekte Schrauben sind in der Stichprobe.

5. Schritt: Entscheidung und Interpretation

x = 6 liegt nicht im kritischen Bereich, x = 6 liegt also im Annahmebereich:
H_0: p = 0,02 wird beibehalten.

2.2 Zweiseitige Fragestellung

Beispiel zur Testentwicklung:

Der Gummibärchenhersteller HARRY PO will überprüfen, ob der Anteil der roten Bärchen 25 % beträgt. Er entnimmt 50 Bärchen aus der laufenden Produktion und zählt zehn rote Bärchen.
Ist damit die Hypothese, daß durchschnittlich ein Viertel der Bärchen rot sind, bei einem Signifikanzniveau von $\alpha = 0{,}05$ widerlegt?

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	$\Theta=[0,1]$
Nullhypothese:	$H_0: p=0,25$
Alternativhypothese:	$H_A: p\neq 0,25$
Signifikanzniveau:	$\alpha=0,05$

2. Schritt: Prüfgröße, Testverteilung

Als Schätzfunktion kann hier $\hat{p}=\frac{X}{n}$ verwendet werden. Es ergibt sich $\hat{p}=\frac{10}{50}=0,20$.

Unter H_0 ist X ~ B(50; 0,25), da n/N klein ist.

Da $n\cdot p\cdot(1-p)=50\cdot 0,25\cdot(1-0,25)=9,375>9$ ist, kann für $\hat{p}$ unter H_0 auch approximativ eine Normalverteilung angenommen werden.[7]

$$E(\hat{p})=p=0,25 \text{ und}$$

$$Var(\hat{p}\,|\,p)=\frac{p\cdot(1-p)}{n}=\frac{0,25\cdot 0,75}{50}=0,00375$$

Testverteilung: Normalverteilung

Als standardnormalverteilte Prüfgröße kann hier dann verwendet werden:

$$Z=\frac{\hat{p}-0,25}{\sqrt{0,00375}} \sim N(0,\,1)$$

3. Schritt: Kritischer Bereich

Die Nullhypothese H_0: p = 0,25 wird verworfen, wenn der Anteil der gezählten roten Bärchen in der Stichprobe von p = 0,25 'weit' nach oben oder unten abweicht. Kritischer Bereich für die Prüfgröße Z:

$$(-\infty, z_u]\cup[z_o, +\infty).$$

Bei einer zweiseitigen Fragestellung erscheint es sinnvoll, die Irrtumswahrscheinlichkeit α symmetrisch in $2\cdot\alpha/2$ aufzuteilen:

wähle z_u so, daß $P(Z\leq z_u\,|\,p=0,25)=\frac{\alpha}{2}$;

wähle z_o so, daß $P(Z\geq z_o\,|\,p=0,25)=\frac{\alpha}{2}$.

Da Z näherungsweise standardnormalverteilt ist, sind $z_u=z_{\alpha/2}$ und $z_o=z_{1-\alpha/2}$ aus der Tabelle zu entnehmen:

$$\frac{\alpha}{2}=0,025 \Rightarrow z_u=-1,96,\ z_o=+1,96$$

Aus der Tabelle ergibt sich:

7 Vgl. hierzu auch den Abschnitt IV.5 „Normalverteilung als Näherungsverteilung“.

$F_2(Z) = 0{,}95 \, und \, z = F_2^{-1}(0{,}95)$ bzw.
$F_Z(Z) = 0{,}975 \Rightarrow z = 1{,}96.$

4. Schritt: Wert der Prüfgröße

Für die Prüfgröße ergibt sich aus der Stichprobe:

$$z = \frac{\hat{p} - E(\hat{p})}{\sqrt{Var(\hat{p}|p)}} = \frac{0{,}20 - 0{,}25}{\sqrt{0{,}00375}} = \frac{-0{,}05}{0{,}06124} = -0{,}82$$

5. Schritt: Entscheidung, Interpretation

Da $-1{,}96 < z = -0{,}82 < +1{,}96$ ist, kann $H_0: p = 0{,}25$ nicht verworfen werden.

Die Hypothese, daß im Durchschnitt ein Viertel der Bärchen rot sind, kann bei $\alpha = 0{,}05$ nicht widerlegt werden.

Alternativ kann man auch eine Entscheidung über die kritischen Anteilswerte treffen:

$$z_u = \frac{p_{cu} - p_0}{\sigma_{\hat{p}}} \Rightarrow p_{cu} = p_0 + z_u \sigma_{\hat{p}};$$

$$z_o = \frac{p_{co} - p_0}{\sigma_{\hat{p}}} \Rightarrow p_{co} = p_0 + z_o \sigma_{\hat{p}},$$

wobei $\sigma_{\hat{p}} = \sqrt{Var(\hat{p}|p)}$ und $p_0 = E(\hat{p})$.

Für das Beispiel ergibt sich hier konkret:

$$p_{cu} = 0{,}25 - 1{,}96 \cdot \sqrt{0{,}00375} = 0{,}25 - 0{,}12 = 0{,}13;$$

$$p_{co} = 0{,}25 + 1{,}96 \cdot \sqrt{0{,}00375} = 0{,}25 + 0{,}12 = 0{,}37$$

Da $0{,}13 < \hat{p} = 0{,}20 < 0{,}37$ ist, wird H_0 beibehalten, ein Anteil der roten Bärchen von 25 % kann nicht widerlegt werden (bei $\alpha = 0{,}05$).

2.3 Einseitige Fragestellung

Beispiel zur Testentwicklung:

Wir betrachten hier erneut das Unternehmen CHIO CHONG aus dem einleitenden Beispiel. Als Produzent eines Massenartikels hatte CHIO CHONG behauptet, daß der Ausschußanteil einer Lieferung höchstens 10 % betrage. Wir wollen hier die Behauptung '$p \leq 0{,}10$' überprüfen.
Eine Stichprobe ohne Zurücklegen hat bei einem Umfang von n = 100 x = 13 defekte Chips zu Tage gebracht. Kann die Behauptung bei einem Signifikanzniveau von $\alpha = 0{,}05$ widerlegt werden?

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = [0,1]$

Nullhypothese: $H_0: p \leq 0{,}10$

Alternativhypothese: $H_A: p > 0{,}10$

Signikanzniveau: $\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Unter H_0 ist die Anzahl X der defekten Chips binomialverteilt $(B(100;0{,}10))$.
Wiederum wollen wir hier annehmen, daß n/N klein genug ist, um eine Binomialverteilung für gerechtfertigt zu halten.

Die Schätzfunktion $\hat{p} = \frac{X}{n}$ ist dann approximativ normalverteilt, wenn $n \cdot p \cdot (1-p) \geq 9$ ist. Offensichtlich ist diese Ungleichung erfüllt, wenn $p = 0{,}10$ gilt, aber nicht für $0 \leq p < 0{,}10$. Man benötigt aber eine Prüfgröße, für die die Testverteilung unter der gesamten Nullhypothese angegeben werden kann.

Muß man dieses Testproblem deshalb mit der Binomialverteilung lösen? Wir werden im dritten Schritt sehen, daß dies nicht notwendig ist.

3. Schritt: Kritischer Bereich

Die Nullhypothese $H_0: p \leq 0{,}10$ wird verworfen, wenn sich sehr viele defekte Chips in der Stichprobe befinden.

Kritischer Bereich: $[k_c, \infty$

Die Grenze k_c ist so zu wählen, daß die Wahrscheinlichkeit, H_0 zu verwerfen, obwohl sie richtig ist, für jedes $p \in [0; 0{,}10]$ höchstens α ist. Gesucht wird also das kleinste k_c, für das gerade noch gilt:[8]

$$P(X \geq k_c \mid p) \leq \alpha \text{ für alle } 0 \leq p \leq 0{,}1.$$

Anschaulich ist klar, daß $P(X \geq k_c \mid p)$ mit wachsendem p größer wird, d.h., daß die Wahrscheinlichkeit also (unter H_0) für p = 0,10 maximal ist. Dies bedeutet folglich:

Aus $P(X \geq k_c \mid p = 0{,}10) \leq \alpha$ folgt $P(X \geq k_c \mid 0 \leq p \leq 0{,}10) \leq \alpha$,

weil $P(X \geq k_c \mid 0 \leq p \leq 0{,}10) \leq P(X \geq k_c \mid p = 0{,}10) \leq \alpha$ ist.

Wenn man für p = 0,10 ein k_c gefunden hat, das das Signifikanzniveau α einhält, so wird mit k_c das Signifikanzniveau α bei jedem $0 \leq p \leq 0{,}10$ auch eingehalten. Damit genügt es also, ein k_c in Abhängigkeit von p = 0,10 zu konstruieren.

Damit können wir die Binomialverteilung hier durch die Normalverteilung approximieren:

$$E(\hat{p} \mid p = 0{,}10) = E\left(\frac{X}{n} \mid p = 0{,}10\right) = 0{,}10;$$

$$Var(\hat{p} \mid p = 0{,}10) = \sigma_{\hat{p}}^2 = \frac{p \cdot (1-p)}{n} = \frac{0{,}1 \cdot 0{,}9}{100} = 0{,}0009.$$

Für p = 0,10 ist $\hat{p} \sim N(0{,}1; 0{,}03^2)$-verteilt.

Prüfgröße: $Z = \frac{\hat{p} - E(\hat{p})}{\sqrt{Var(\hat{p} \mid p)}} = \frac{\hat{p} - 0{,}1}{0{,}03}$, wobei Z ~ N(0, 1)-verteilt ist.

---

8 Die Ableitung findet sich in Hujer 1991, S. 186 ff.

Als Entscheidungsregel ergibt sich dann:

H_0 sollte verworfen werden, falls die Stichprobe ein $z > z_{1-\alpha}$ liefert. Hier ergibt sich:

$$z_{1-\alpha} = z_{0,95} = 1{,}645 \left(Aus: z = F_z^{-1}(z) = 1{,}645 \right)$$

Kritischer Bereich: $[1,645; \infty$

4. Schritt: Wert der Prüfgröße

$$\hat{p} = \frac{x}{n} = \frac{13}{100} = 0{,}13$$

$$z = \frac{\hat{p} - 0{,}1}{0{,}03} = \frac{0{,}13 - 0{,}10}{0{,}03} = \frac{0{,}03}{0{,}03} = 1$$

5. Schritt: Entscheidung, Interpretation

$z = 1 < 1{,}645$, d.h., die Stichprobe liefert hier einen Wert für z, der nicht in den kritischen Bereich fällt. $H_o: p \leq 0{,}10$ kann also nicht verworfen werden.

Alternativ kann man hier auch einen Wert für den kritischen Anteilswert p_c bestimmen:

$$z_c = \frac{p_c - p_0}{\sigma_{\hat{p}}}$$

mit $p_0 = E(\hat{p}|\, p)$ und $\sigma_{\hat{p}} = \sqrt{Var(\hat{p}|\, p)}$.

Es ergibt sich:

$$\begin{aligned} p_c &= p_0 + z_c \cdot \sigma_{\hat{p}} \\ &= 0{,}10 + 1{,}645 \cdot 0{,}03 \\ &= 0{,}14935. \end{aligned}$$

Bei 13 defekten Chips $(n = 100)$ kann H_0 nicht verworfen werden. Erst ab 15 defekten Chips ist die Lieferung zu beanstanden.

3 Einstichprobentest für das arithmetische Mittel bei normalverteilter Grundgesamtheit

Die Testgrößen im letzten Kapitel bauen auf binomialverteilten Zufallsvariablen auf. Bei bekanntem Stichprobenumfang n ist die Binomialverteilung nur von dem Parameter p abhängig.

Die Normalverteilung ist aber von zwei Parametern abhängig: μ und σ^2. Man unterscheidet bei Tests für das arithmetische Mittel zwei Fälle:

1. Die Varianz σ^2 ist bekannt, d.h., es bleibt ein unbekannter Parameter, nämlich das zu testende μ.
2. Die Varianz σ^2 ist unbekannt. Hier gibt es eine Prüfvariable, die unabhängig von σ^2 ist.

3.1 Einstichprobentest für μ bei bekanntem σ^2

Die folgenden Tests sollen an einem **Apfelweinbeispiel** entwickelt werden:

Bei der letzten Apfelwein-Kampagne hat eine Abfüllmaschine mit μ = 1,0 l und σ = 0,001 l gearbeitet. Jetzt soll überprüft werden, ob die durchschnittliche Füllmenge noch bei 1,0 l liegt. Dazu werden n = 30 Flaschen aus der laufenden Produktion entnommen und nachgemessen. Die Stichprobe liefert ein $\bar{x}$ = 0,9995 l. Hat sich die durchschnittliche Füllmenge signifikant bei α = 0,05 verändert?

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = [0,\infty$

Nullhypothese: $H_0 : \mu = 1,0$

Alternativhypothese: $H_A : \mu \neq 1,0$

Signifikanzniveau: $\alpha = 0,05$

2. Schritt: Prüfgröße, Testverteilung

Die Zufallsvariable $\overline{X}$ ist normalverteilt. Unter H_0 besitzt $\overline{X}$ einen Mittelwert $\mu_0 = 1,0$. Wenn man annimmt, daß sich der Standardfehler $\sigma = 0,001$ nicht geändert hat, ist $\overline{X}$ unter H_0 folglich $N\left(1,0;0,001^2\right)$-verteilt.
Standardisieren liefert damit die standardnormalverteilte Prüfgröße

$$Z = \frac{\overline{X} - \mu_0}{\sqrt{Var\left(\overline{X}\right)}} = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$$

Wie bei der Konfidenzintervallberechnung ergibt sich die Mittelwertvarianz $\sigma^2_{\overline{X}}$ aus der Varianz σ^2 der Zufallsvariablen X aus

$$Var\left(\overline{X}\right) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \qquad \text{bzw.} \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}.$$

Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0,05$:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

3. Schritt: Kritischer Bereich

Für den zweiseitigen Test entnimmt man der Tabelle der Standardnormalverteilung die kritischen Werte $(\alpha = 0,05)$:

$$z_c^o = 1,96 \qquad z_c^u = -1,96$$

Der kritische Bereich hat hier die Form:

$$(-\infty; -1,96] \cup [1,96; \infty).$$

4. Schritt: Wert der Prüfgröße

Mit den Werten der Stichprobe (mit Zurücklegen) erhält man:

$$z = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{0{,}9995 - 1{,}000}{0{,}001}\sqrt{30} = -\frac{0{,}0005}{0{,}001}\sqrt{30} = -2{,}74.$$

5. Schritt: Entscheidung, Interpretation

$$z = -2{,}74 < -1{,}96 = z_c^u.$$

Damit wird H_0 abgelehnt.

Mit einer Irrtumswahrscheinlichkeit von 5 % hat sich die durchschnittliche Füllmenge des feinen 'Stöffchens' also verändert.

Auch hier kann man einen einseitigen Test durchführen:

$$\begin{array}{lcl} H_0: \mu \geq \mu_0 & & H_0: \mu \leq \mu_0 \\ & \text{bzw.} & \\ H_A: \mu < \mu_0 & & H_A: \mu > \mu_0 \end{array}$$

Als kritischer Bereich ergibt sich hier:

$$\left(-\infty; z_\alpha\right] \qquad \left[z_{1-\alpha}; +\infty\right)$$

3.2 Einstichprobentest für μ bei unbekanntem σ^2

Wir betrachten hier weiterhin das **Apfelweinbeispiel**. Die Stichprobe soll hier als zusätzliches Ergebnis eine Standardabweichung von s = 0,00135 l geliefert haben. Die Annahme, daß σ nach wie vor bei 0,001 l liegt, wird jetzt fallengelassen.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = [0, \infty$

Nullhypothese: $H_0: \mu = 1{,}0$

Alternativhypothese: $H_A: \mu \neq 1{,}0$

Signifikanzniveau: $\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Die standardisierte Prüfgröße ist analog zur Konfidenzintervallberechnung

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\hat{Var}\left(\bar{X}\right)}} = \frac{\bar{X} - \mu_0}{\hat{\sigma}_{\bar{X}}} \text{ mit}$$

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n-1}}$$

bzw. bei Korrektur für eine Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0{,}05$:

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N}}.$$

Die Prüfgröße T ist unter H_0 t-verteilt mit $\nu = n - 1$ Freiheitsgraden.

3. Schritt: Kritischer Bereich

Für den zweiseitigen Test entnimmt man der Tabelle der t-Verteilung die kritischen Werte $(\alpha=0{,}05)$:

$$t_c^o = 2{,}045 \text{ und } t_c^u = -2{,}045 \qquad \nu = 29$$

Damit erhält man als kritischen Bereich:

$$(-\infty; -2{,}045] \cup [2{,}045; \infty)$$

4. Schritt: Wert der Prüfgröße

Mit den Werten der Stichprobe ist:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{(n-1)}} = \frac{0{,}9995 - 1{,}000}{0{,}00135/\sqrt{29}} = -\frac{0{,}0005}{0{,}000251} = -1{,}99.$$

5. Schritt: Entscheidung, Interpretation

$t_c^u = -2{,}045 < -1{,}99 = t < +2{,}045 = t_c^o$, d. h. $H_0 : \mu = 1{,}0$ wird nicht abgelehnt.

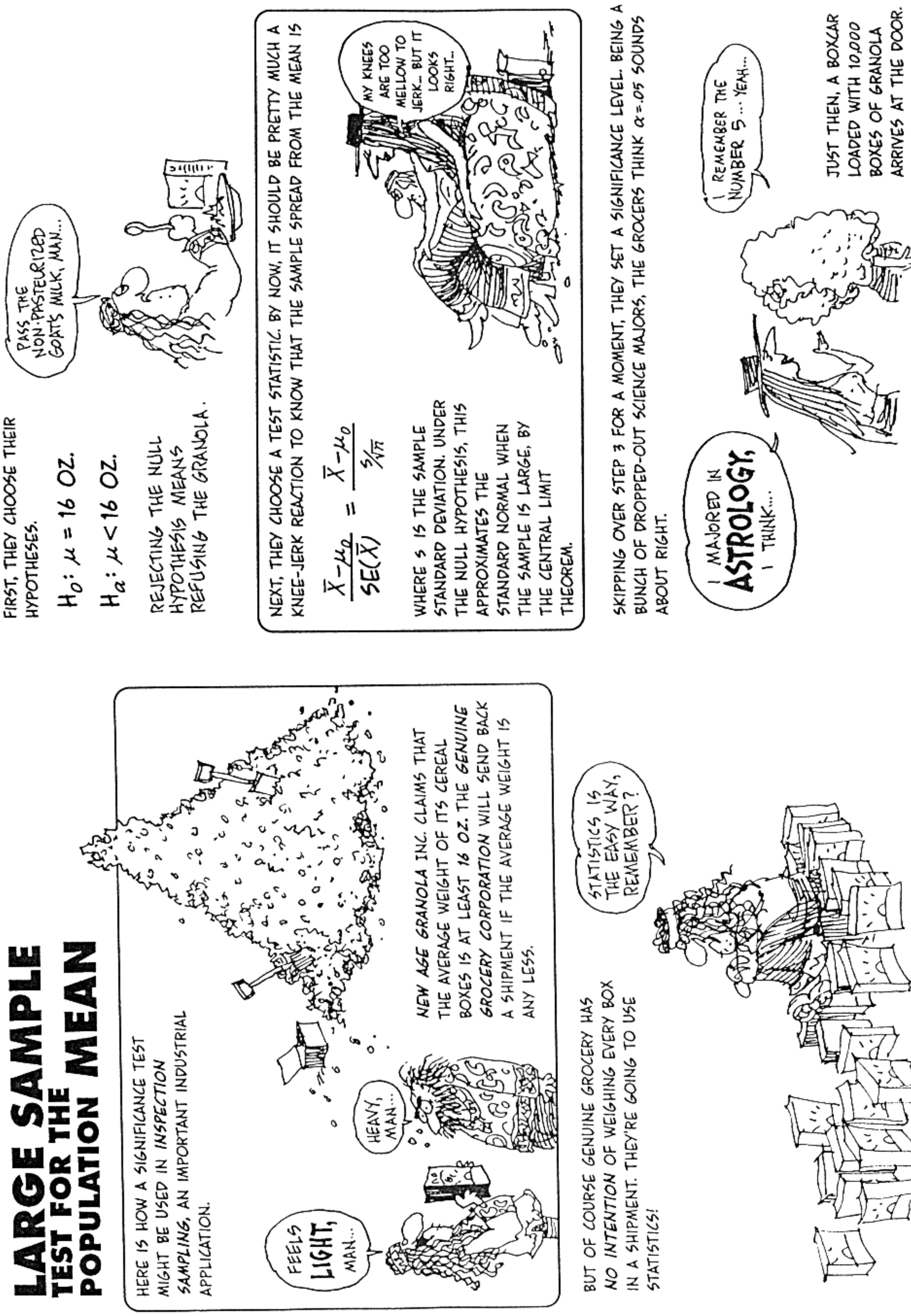
Auch hier kann man einen einseitigen Test durchführen:

$$H_0: \mu \geq \mu_0 \qquad\qquad H_0: \mu \leq \mu_0$$
$$\text{bzw.}$$
$$H_A: \mu < \mu_0 \qquad\qquad H_A: \mu > \mu_0$$

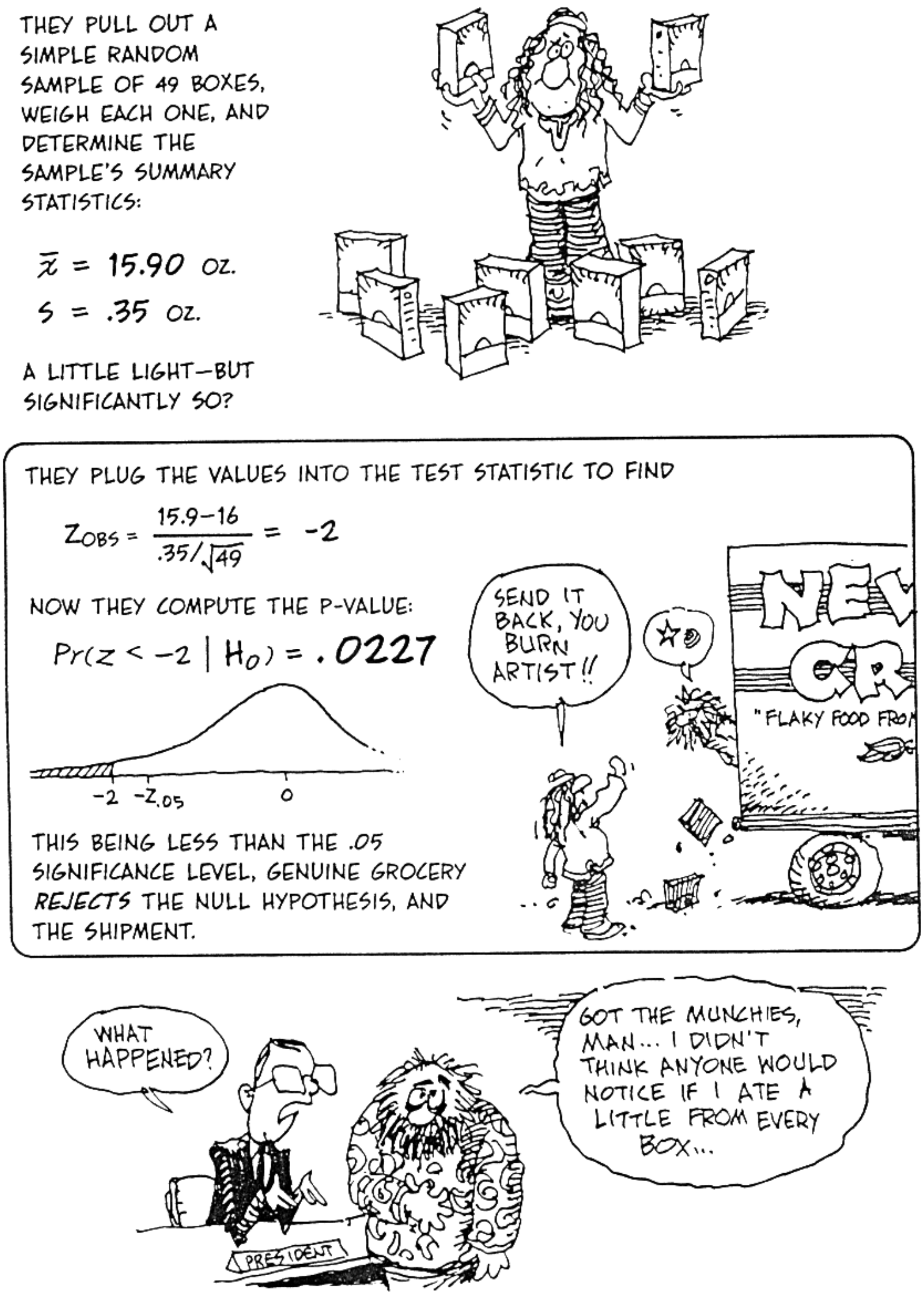
Als kritischer Bereich ergibt sich dann:

$$(-\infty; t_{\alpha;n-1}] \qquad\qquad [t_{1-\alpha;n-1}; +\infty)$$

Für n > 30 kann die t-Verteilung durch die Normalverteilung approximiert werden.



Gonick, Smith 1993



Gonick, Smith 1993

4 Einstichprobentest für die Varianz bei normalverteilter Grundgesamtheit

Im letzten Kapitel wurde das arithmetische Mittel getestet; nun steht die Varianz als zweite zentrale Kenngröße im Vordergrund des Interesses.

4.1 Zweiseitige Fragestellung

Für unser **Apfelweinbeispiel** wird nun die Frage untersucht, ob sich der Standardfehler σ der Füllmenge signifikant bei $\alpha = 0,05$ verändert hat.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = \mathrm{IR}^+$

Nullhypothese: $H_0 : \sigma = 0,001$

Alternativhypothese: $H_A : \sigma \neq 0,001$

Signifikanzniveau: $\alpha = 0,05$

2. Schritt: Prüfgröße, Testverteilung

Die Prüfgröße

$$\chi_0^2 = \frac{n \cdot S^2}{\sigma_0^2}$$

ist unter H_0 χ^2-verteilt mit $\nu = n - 1$ Freiheitsgraden.

3. Schritt: Kritischer Bereich

Für den zweiseitigen Test entnimmt man der Tabelle der χ^2-Verteilung die kritischen Werte $(\alpha = 0,05, \nu = 29)$:

$$\chi_{cu}^2 = 16,047 \quad \text{und}$$
$$\chi_{co}^2 = 45,722$$

4. Schritt: Wert der Prüfgröße

$$\chi^2 = \frac{n \cdot S^2}{\sigma_0^2} = 30 \cdot \left(\frac{0,00135}{0,001} \right)^2 = 30 \cdot (1,35)^2 = 54,675.$$

5. Schritt: Entscheidung, Interpretation

$\chi_{co}^2 = 45,722 < 54,675 = \chi^2$, d.h., H_0: $\sigma = 0,001$ wird abgelehnt: der Standardfehler der Füllmenge hat sich also signifikant verändert.

4.2 Einseitige Fragestellung

Wir wollen unser **Apfelweinbeispiel** hier weiter betrachten, aber nun überprüfen, ob sich der Standardfehler σ signifikant vergrößert hat.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = \mathrm{IR}^{+}$

Nullhypothese: $\mathrm{H}_0 : \sigma \leq 0{,}001$

Alternativhypothese: $\mathrm{H}_A : \sigma > 0{,}001$

Signifikanzniveau: $\alpha = 0{,}05$

2. und 3. Schritt: Prüfgröße, Testverteilung, kritischer Bereich

Die Prüfgröße

$$\chi^2 = \frac{n \cdot S^2}{\sigma^2}$$

ist χ^2-verteilt mit $\nu = n - 1$ Freiheitsgraden, σ^2 ist der wahre Parameter.

Als Prüfgröße wählt man hier

$$\chi_0^2 = \frac{n \cdot S^2}{\sigma_0^2},$$

denn ist $\sigma = \sigma_0 = 0{,}001$, dann ist $\chi_0^2 = \dfrac{n \cdot S^2}{\sigma_0^2}$ $\chi^2(n-1)$-verteilt, und man entnimmt der Tabelle für die χ^2-Verteilung den kritischen Wert $(\alpha = 0{,}05)$:

$$\chi_c^2(29) = 42{,}557,$$

und es gilt für jedes σ^2:

$$P\left(\chi^2 \geq \chi_c^2 \middle| \sigma^2\right) = P\left(\frac{n \cdot S^2}{\sigma^2} \geq \chi_c^2 \middle| \sigma^2\right) = \alpha.$$

Wenn $\sigma^2 \leq \sigma_0^2$ gilt, dann ist:

$$\begin{aligned} P\left(\chi_0^2 \geq \chi_c^2 \middle| \sigma^2 \leq \sigma_0^2\right) &= P\left(\frac{n \cdot S^2}{\sigma_0^2} \geq \chi_c^2 \middle| \sigma^2 \leq \sigma_0^2\right) \\ &= P\left(\frac{n \cdot S^2}{\sigma^2} \cdot \frac{\sigma^2}{\sigma_0^2} \geq \chi_c^2 \middle| \sigma^2 \leq \sigma_0^2\right) \\ &= P\left(\underbrace{\frac{n \cdot S^2}{\sigma^2}}_{\chi^2} \geq \underbrace{\chi_c^2 \cdot \underbrace{\frac{\sigma_0^2}{\sigma^2}}_{\geq 1}}_{\geq \chi_c^2} \middle| \sigma^2 \leq \sigma_0^2\right) \\ &\leq P\left(\chi^2 \geq \chi_c^2 \middle| \sigma^2 \leq \sigma_0^2\right) = \alpha \end{aligned}$$

Damit ist gezeigt, daß für die Prüfgröße $\chi_0^2 = \frac{n \cdot S^2}{\sigma_0^2}$ gilt:

$$P(\chi_0^2 \geq \chi_c^2 \mid \sigma^2) \leq \alpha \quad \text{für jedes } \sigma^2 \leq \sigma_0^2.$$

Als Ergebnis kann festgehalten werden, daß mit der Prüfgröße χ_0^2 und dem kritischen Wert $\chi_c^2(n-1)$ ein Test konstruiert ist, der für jedes σ^2 aus H_0 das Signifikanzniveau α einhält.

4. Schritt: Wert der Prüfgröße

$$\chi_0^2 = \frac{n \cdot S^2}{\sigma_0^2} = 30 \cdot \left(\frac{0{,}00135}{0{,}001} \right)^2 = 54{,}675$$

5. Schritt: Entscheidung, Interpretation

Da $\chi_0^2 = 54{,}675 > 42{,}557 = \chi_c^2$ ist, wird die Nullhypothese H_0: $\sigma \leq 0{,}001$ abgelehnt.

5 Zweistichprobentest für die Differenz zweier arithmetischer Mittel

Wir betrachten nun den Fall, daß aus zwei Grundgesamtheiten je eine Stichprobe vom Umfang n_1 bzw. n_2 gezogen wurde. Die beiden Stichproben liefern die arithmetischen Mittel $\overline{x}_1$ und $\overline{x}_2$.

Kann man nun aus der beobachteten Differenz auf die Differenz $\mu_1 - \mu_2$ in der Grundgesamtheit schließen?

Wir wollen hier zwei Modellvoraussetzungen annehmen:

- Die beiden Stichproben sind voneinander unabhängig.
- Beide Stichproben stammen aus normalverteilten Grundgesamtheiten.

Beispiel:

Mit Hilfe der Maschinen M_1 und M_2 wird Tee verpackt. Es soll nun überprüft werden, ob die Maschine M_1 mit dem gleichen durchschnittlichen Füllgewicht arbeitet wie M_2. Dazu wird je eine Stichprobe von jeder Maschine erhoben:

Stichprobe 1: $n_1 = 12, \overline{x}_1 = 130$ gr.
Stichprobe 2: $n_2 = 10, \overline{x}_2 = 127$ gr.

Die Füllgewichte der beiden Maschinen sollen annähernd normalverteilt sein. Als Signifikanzniveau wird $\alpha = 0{,}01$ festgelegt.

Zur besseren Übersicht wollen wir den Zweistichprobentest für die Differenz zweier arithmetischer Mittel an diesem Beispiel getrennt für bekannte (5.1) und unbekannte Varianzen (5.2) durchführen.

Wie auch aus dem Kapitel über Konfidenzintervalle ersichtlich wird, wird der prinzipielle Unterschied der Vorgehensweise lediglich in der Verwendung unterschiedlicher Testverteilungen liegen:

bekannte Varianz: Standardnormalverteilung
unbekannte Varianz: Student (t)-Verteilung (da Varianz geschätzt wird).

In beiden Abschnitten (5.1 mit bekannter Varianz sowie 5.2 mit unbekannter Varianz) werden jeweils zwei Fälle unterschieden:

- Fall 1: Die Standardabweichung beider Maschinen ist gleich (Varianzhomogenität)
- *Fall 2: Die Standardabweichung beider Maschinen ist ungleich (Varianzinhomogenität)*

Korrektur bei Stichproben ohne Zurücklegen und $n / N \geq 0{,}05$:
Die Varianzen für X_1 müssen mit $(N_1 - n_1)/(N_1 - 1)$ und für X_2 mit $(N_2 - n_2)/(N_2 - 1)$ multipliziert werden.

5.1 Zweistichprobentest für die Differenz zweier arithmetischer Mittel bei bekannter Varianz

Mit bekannten Varianzen der Grundgesamtheit und dem obigen Beispiel wird nun der Zweistichprobentest für die Differenz zweier arithmetischer Mittel durchgeführt.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	$\mu_1 \in \mathrm{IR}: \Theta_1 = \mathrm{IR}$ $\mu_2 \in \mathrm{IR}: \Theta_2 = \mathrm{IR}$ Also: $\Theta = \Theta_1 \times \Theta_2 = \mathrm{IR}^2$	bzw. $\mu_1 - \mu_2 \in \mathrm{IR}$ $\Theta = \mathrm{IR}$
Nullhypothese:	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 - \mu_2 = 0$
Alternativhypothese:	$H_A: \mu_1 \neq \mu_2$	$H_A: \mu_1 - \mu_2 \neq 0$
Signifikanzniveau:	$\alpha = 0{,}01$	$\alpha = 0{,}01$

2. Schritt: Prüfgröße, Testverteilung

Wir ermitteln für die Stichprobenfunktion $(\bar{X}_1 - \bar{X}_2)$ den Erwartungswert und die Varianz:

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) \text{(wegen Unabhängigkeit, vgl. VII. 6)}$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) - \underbrace{2Cov(\bar{X}_1, \bar{X}_2)}_{=0 \text{ (wegen Unabhängigkeit, vgl. VII. 6)}}$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma_D^2$$

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 := \sigma^2$ (bekannt)

$$Var(\bar{X}_1 - \bar{X}_2) = \sigma_D^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = \sigma^2 \left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right)$$

Für die standardnormalverteilte Prüfgröße Z erhalten wir mit der Nullhypothese $\mu_1 - \mu_2 = 0$:

$$Z = \frac{X-\mu}{\sigma} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sigma_D} = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_D}$$

$$= \frac{\overline{X}_1 - \overline{X}_2}{\sigma \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Z ist unter H_0 N(0,1)-verteilt.

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$ *(bekannt)*

$$Var\left(\overline{X}_1 - \overline{X}_2\right) = \sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{n_2 \cdot \sigma_1^2 + n_1 \cdot \sigma_2^2}{n_1 \cdot n_2}$$

Prüfgröße:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_D} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{n_2 \cdot \sigma_1^2 + n_1 \cdot \sigma_2^2}{n_1 \cdot n_2}}} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_2 \cdot \sigma_1^2 + n_1 \cdot \sigma_2^2}} \cdot \sqrt{n_1 \cdot n_2}$$

Z ist unter H_0 N(0,1)-verteilt.

3. Schritt: Kritischer Bereich

Die kritischen Bereiche können hier wie im Einstichprobentest konstruiert werden.

Da die Prüfgrößen Z für die Fälle 1 und 2 (Varianzhomogenität und -inhomogenität) standardnormalverteilt sind, erhalten wir für beide Fälle die folgenden

kritischen Bereiche (zweiseitige Fragestellung):

$$\left(-\infty, z_u\right] \cup \left[z_o, +\infty\right).$$

Aus der Tabelle der Standardnormalverteilung entnimmt man dann die Werte:

$$z_u = z_{\frac{\alpha}{2}} \quad \text{und} \quad z_o = z_{1-\frac{\alpha}{2}}.$$

Mit der Irrtumswahrscheinlichkeit $\alpha = 0{,}01$ ist

$$z_u = z_{0{,}005} = -2{,}5758 \quad \text{und} \quad z_o = z_{0{,}995} = 2{,}5758.$$

Kritischer Bereich:

$$\left(-\infty, -2{,}5758\right] \cup \left[2{,}5758, +\infty\right).$$

4. Schritt: Wert der Prüfgröße

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 := \sigma^2$ (bekannt)

Mit der bekannten Varianz von $\sigma^2 = 2$ bzw. $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 2{,}0$ gr. ist:

$$
\begin{aligned}
z &= \frac{\overline{x}_1 - \overline{x}_2}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \\
&= \frac{130-127}{\sqrt{2{,}0}} \cdot \sqrt{\frac{12 \cdot 10}{12+10}} \\
&= 2{,}1213 \cdot 2{,}3355 \\
&= 4{,}9543
\end{aligned}
$$

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$ *(bekannt)*

Mit den bekannten ungleichen Standardabweichungen $\sigma_1 = 2{,}3$gr. und $\sigma_2 = 1{,}9$gr.

$$
\begin{aligned}
z &= \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{n_2 \cdot \sigma_1^2 + n_1 \cdot \sigma_2^2}} \cdot \sqrt{n_1 \cdot n_2} \\
&= \frac{130-127}{\sqrt{10 \cdot (2{,}3)^2 + 12 \cdot (1{,}9)^2}} \cdot \sqrt{12 \cdot 10} \\
&= \frac{3}{9{,}8092} \cdot 10{,}9545 \\
&= 3{,}3503
\end{aligned}
$$

5. Schritt: Entscheidung, Interpretation

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 := \sigma^2$ (bekannt)

$$z = 4{,}9543 > 2{,}5758 = z_o,$$

d.h., die Hypothese $H_0\colon \mu_1 = \mu_2$ wird abgelehnt.

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$ *(bekannt)*

$$z = 3{,}3503 > 2{,}5758 = z_o,$$

d.h., die Hypothese $H_0\colon \mu_1 = \mu_2$ wird auch in diesem Fall abgelehnt.

5.2 Zweistichprobentest für die Differenz zweier arithmetischer Mittel bei unbekannter Varianz

In der Praxis sind die Varianzen σ_1^2 und σ_2^2 jedoch meist unbekannt. Häufig muß auch die Annahme $\sigma_1^2 = \sigma_2^2$ fallengelassen werden. Wie im Einstichprobentest können die Varianzen aber durch ihre Schätzer ersetzt werden, wodurch man t-verteilte Prüfgrößen erhält.

Wir betrachten weiter das **Verpackungsmaschinenbeispiel**, wobei wir die Modifikation vornehmen, daß jetzt zusätzlich die Stichprobenstandardfehler ermittelt werden, aber dafür die wahren Varianzen σ_1^2 und σ_2^2 unbekannt sein sollen:

- Stichprobe 1: $n_1 = 12$, $\overline{x}_1 = 130$ gr., $s_1 = 2{,}2$ gr.

- Stichprobe 2: $n_2 = 10$, $\overline{x}_2 = 127$ gr., $s_2 = 1{,}8$ gr.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	$\mu_1 \in IR: \Theta_1 = IR$ $\mu_2 \in IR: \Theta_2 = IR$ Also: $\Theta = \Theta_1 \times \Theta_2 = IR^2$	bzw.	$\mu_1 - \mu_2 \in IR$ $\Theta = IR$
Nullhypothese:	$H_0: \mu_1 = \mu_2$		$H_0: \mu_1 - \mu_2 = 0$
Alternativhypothese:	$H_A: \mu_1 \neq \mu_2$		$H_A: \mu_1 - \mu_2 \neq 0$
Signifikanzniveau:	$\alpha = 0{,}01$		$\alpha = 0{,}01$

2. Schritt: Prüfgröße, Testverteilung

Unterschiede zum Fall bekannter Varianzen ergeben sich jedoch für die Prüfgröße:

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (unbekannt)

Wie unter VII. 2.2 ausgeführt, kann die unbekannte Varianz erwartungstreu geschätzt werden allgemein durch

$$\hat{\sigma}^2 = \frac{n}{n-1} S^2 = \frac{n}{n-1} \frac{\sum_i \left(X_i - \overline{X}\right)^2}{n} = \frac{\sum_i \left(X_i - \overline{X}\right)^2}{n-1}$$

Mit den beiden Teilstichproben und

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1-1} S_1^2 = \frac{\sum_{i=1}^{n_1} \left(X_{1i} - \overline{X}_1\right)^2}{n_1-1} \qquad \text{wobei } S_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_{1i} - \overline{X}_1)^2$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2-1} S_2^2 = \frac{\sum_{i=1}^{n_2} \left(X_{2i} - \overline{X}_2\right)^2}{n_2-1} \qquad S_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (X_{2i} - \overline{X}_2)^2$$

ist dann der erwartungstreue Schätzer für die Gesamtvarianz

$$S^2 = \frac{1}{n_1 + n_2} \cdot \left(n_1 \cdot S_1^2 + n_2 \cdot S_2^2\right)$$

$$\hat{\sigma}^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2} \cdot \frac{\left(n_1 \cdot S_1^2 + n_2 \cdot S_2^2\right)}{n_1 + n_2}$$

$$\hat{\sigma}^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2} \cdot S^2 = \frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2}$$

$$\hat{\sigma}^2 = \frac{n_1 \cdot \frac{1}{n_1} \cdot \sum \left(X_{1i} - \overline{X}_1\right)^2 + n_2 \cdot \frac{1}{n_2} \cdot \sum \left(X_{2i} - \overline{X}_2\right)^2}{n_1 + n_2 - 2}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{(n_1 - 1) + (n_2 - 1)}$$

$$= \frac{(n_1 - 1)\cdot \hat{\sigma}_1^2 + (n_2 - 1)\cdot \hat{\sigma}_2^2}{n_1 + n_2 - 2} = \frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2}$$

Für $S_1^2 = S_2^2 = S^2$ ist

$$\hat{\sigma}^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2} \cdot S^2$$

Eingesetzt in die Varianz der Stichprobenfunktion $\left(\overline{X}_1 - \overline{X}_2\right)$

$$Var\left(\overline{X}_1 - \overline{X}_2\right) = \sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

erhalten wir den erwartungstreuen Schätzer für σ_D^2 mit

$$\hat{\sigma}_D^2 = \hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \hat{\sigma}^2\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)$$

$$= \frac{n_1 \cdot S_1^2 + n_2 \cdot S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right).$$

Die Prüfgröße erhalten wir dadurch, daß wir die unbekannte Varianz σ^2 durch ihren Schätzer $\hat{\sigma}^2$ ersetzen $(\mu_1 - \mu_2 = 0)$:

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\hat{\sigma}_D} = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\hat{\sigma}_D} = \frac{\overline{X}_1 - \overline{X}_2}{\hat{\sigma}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

$$= \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\hat{\sigma}_D^2}}.$$

T ist wegen der geschätzten Varianz t-verteilt mit $\nu = n_1 + n_2 - 2$ Freiheitsgraden.

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$, *(*$\sigma_1^2, \sigma_2^2$ *unbekannt)*

Dieses sogenannte Behrens-Fischer-Problem ist nicht exakt lösbar. Für praktische Zwecke geeignet ist ein Ansatz mit

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} S_1^2 \qquad \hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} S_2^2$$

eingesetzt als Schätzer für σ_1^2 und σ_2^2 in

$$Var\left(\overline{X}_1 - \overline{X}_2\right) = \sigma_D^2 = \frac{n_2\sigma_1^2 + n_1\sigma_2^2}{n_1 \cdot n_2}.$$

Also

$$\hat{\sigma}_D^2 = \frac{n_2\,\hat{\sigma}_1^2 + n_1\,\hat{\sigma}_2^2}{n_1 \cdot n_2} = \frac{\frac{n_2 \cdot n_1}{n_1 - 1} S_1^2 + \frac{n_1 \cdot n_2}{n_2 - 1} S_2^2}{n_1 \cdot n_2}$$

$$= \frac{1}{n_1 - 1} S_1^2 + \frac{1}{n_2 - 1} S_2^2$$

Prüfgröße:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_2 \cdot \hat{\sigma}_1^2 + n_1 \cdot \hat{\sigma}_2^2}} \cdot \sqrt{n_1 \cdot n_2}$$

Diese Prüfgröße ist aber nur annähernd t-verteilt. Die Zahl der Freiheitsgrade bestimmt man nun über

$$\nu^* = \frac{1}{\frac{w^2}{n_1 - 1} + \frac{(1-w)^2}{n_2 - 1}} \quad \text{mit}$$

$$0 < w = \frac{\hat{\sigma}_1^2 \cdot n_2}{\hat{\sigma}_1^2 \cdot n_2 + \hat{\sigma}_2^2 \cdot n_1} < 1.$$

Da ν^* in der Regel keine ganze Zahl sein wird, erhält man die Zahl der Freiheitsgrade durch Abrunden:

$$\nu = \lfloor \nu^* \rfloor.$$

3. Schritt: Kritischer Bereich

Da die Varianzen σ^2 bzw. σ_1^2 und σ_2^2 unbekannt sind, sind die Prüfgrößen t-verteilt bzw. annähernd t-verteilt.

Kritischer Bereich:

$$(-\infty; t_u] \cup [t_o; +\infty)$$

Aus der Tabelle der t-Verteilung entnimmt man die kritischen Werte:

$$t_u = t_{\alpha/2,\nu} \text{ und } t_o = t_{1-\alpha/2,\nu}.$$

Die Fälle 1 und 2 werden in der Regel nicht dieselbe Anzahl von Freiheitsgraden aufweisen. In unserem Beispiel sind die Zahlen aber so gewählt, daß $\nu = \nu^*$.

Im Beispiel gilt:

$$s_1^2 = 2{,}2^2 = 4{,}84 \quad n_1 = 12$$

$$s_2^2 = 1{,}8^2 = 3{,}24 \quad n_2 = 10$$

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (unbekannt)

$$\nu = n_1 + n_2 - 2 = 12 + 10 - 2 = 20$$

$$t_o = t_{20,0{,}995} = 2{,}845, \; t_u = -t_o = -2{,}845$$

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$, *(* σ_1^2, σ_2^2 *unbekannt)*

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} \cdot s_1^2 = \frac{12 \cdot 4,84}{11} = 5,28$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} \cdot s_2^2 = \frac{10 \cdot 3,24}{9} = 3,60$$

$$w = \frac{\hat{\sigma}_1^2 \cdot n_2}{\hat{\sigma}_1^2 \cdot n_2 + \hat{\sigma}_2^2 \cdot n_1} = \frac{5,28 \cdot 10}{5,28 \cdot 10 + 3,60 \cdot 12} = \frac{52,8}{52,8 + 43,20} = \frac{52,8}{96} = 0,55$$

$$\nu^* = \frac{1}{\frac{w^2}{n_1 - 1} + \frac{(1-w)^2}{n_2 - 1}} = \frac{1}{\frac{(0,55)^2}{11} + \frac{(0,45)^2}{9}} = \frac{1}{0,0275 + 0,0225} = \frac{1}{0,05} = 20$$

$$t_o = t_{20,0,995} = 2,845,\ t_u = -t_o = -2,845$$

4. Schritt: Wert der Prüfgröße

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (unbekannt)

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\hat{\sigma}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

$$s^2 = \frac{1}{n_1 + n_2} \cdot \left(n_1 \cdot s_1^2 + n_2 \cdot s_2^2\right)$$

$$= \frac{1}{12 + 10} \cdot (12 \cdot 4,84 + 10 \cdot 3,24)$$

$$= \frac{1}{22} \cdot 90,48 = 4,1127$$

$$\hat{\sigma} = \sqrt{\frac{n_1 + n_2}{n_1 + n_2 - 2}} \cdot s$$

$$= \sqrt{\frac{12 + 10}{12 + 10 - 2}} \cdot \sqrt{4,1127}$$

$$= 1,0488 \cdot 2,0279 = 2,1269$$

$$t = \frac{130 - 127}{2,1269} \cdot \sqrt{\frac{12 \cdot 10}{12 + 10}}$$

$$= 1,4105 \cdot 2,3355 = 3,2942$$

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$, *(* σ_1^2, σ_2^2 *unbekannt)*

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_2 \cdot \hat{\sigma}_1^2 + n_1 \cdot \hat{\sigma}_2^2}} \cdot \sqrt{n_1 \cdot n_2}$$

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} \cdot s_1^2 = \frac{12 \cdot 4,84}{11} = 5,28$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} \cdot s_2^2 = \frac{10 \cdot 3,24}{9} = 3,60$$

$$t = \frac{130 - 127}{\sqrt{10 \cdot 5,28 + 12 \cdot 3,60}} \cdot \sqrt{120} = \frac{3}{\sqrt{96}} \cdot 10,9545 = 3,3541$$

5. Schritt: Entscheidung, Interpretation

Fall 1: Varianzhomogenität, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (unbekannt)

$$t = 3,2942 > 2,845 = t_o$$

Fall 2: Varianzinhomogenität, $\sigma_1^2 \neq \sigma_2^2$, ($\sigma_1^2, \sigma_2^2$ *unbekannt)*

$$t = 3,3541 > 2,845 = t_o$$

Die Hypothese $\mu_1 = \mu_2$ wird also in beiden Fällen abgelehnt.

Auch hier kann man für beide Fälle einseitige Fragestellungen untersuchen. Wir betrachten hier den **einseitigen Test**

$$\begin{array}{lcl} H_0: \mu_1 \leq \mu_2 & & H_0: \mu_1 - \mu_2 \leq 0 \\ & \text{bzw.} & \\ H_A: \mu_1 > \mu_2 & & H_A: \mu_1 - \mu_2 > 0. \end{array}$$

Wir erhalten in unserem Beispiel:

Die Varianzen σ^2 bzw. σ_1^2 und σ_2^2 sind bekannt:

Als kritischen Wert ermitteln wir hier $(\alpha = 0,01)$:

$$z_c = z_{0,99} = 2,3263.$$

In beiden Fällen (Varianzhomogenität und -*inhomogenität*) ist $z > z_c$. Daher wird die Hypothese $\mu_1 \leq \mu_2$ verworfen.

Die Varianzen σ^2 bzw. σ_1^2 und σ_2^2 sind unbekannt:

Als kritischen Wert ermitteln wir hier $(\alpha = 0,01)$:

$$t_c = t_{20,0,99} = 2,528.$$

Auch hier ist in beiden Fällen (Varianzhomogenität und -*inhomogenität*) $t > t_c$, so daß $\mu_1 \leq \mu_2$ verworfen wird.

6 Zweistichprobentests für den Quotienten zweier Varianzen

Zwei Voraussetzungen werden beibehalten:

- Die beiden Stichproben sind unabhängig voneinander.
- Beide Stichproben stammen aus normalverteilten Grundgesamtheiten.

Zweistichprobentests für den Quotienten zweier Varianzen sollen anhand eines Beispiels entwickelt werden.

Als **Beispiel** betrachten wir hier den **Vergleich von Einkommensverteilungen**. Es soll untersucht werden, ob die Streuung der Einkommen in A ungleich oder größer ist als in B (A, B seien Länder, Betriebe usw.). Eine Stichprobe in A vom Umfang $n_1 = 21$ hat eine Standardabweichung von $s_1 = 322$ DM ergeben. In B wurde eine Stichprobe vom Umfang $n_2 = 16$ erhoben, die eine Standardabweichung von $s_2 = 288$ DM ergab. Ist die Streuung der Einkommen in A ungleich oder größer als in B, wenn $\alpha = 0,05$ gewählt wird?

Wir unterscheiden wieder nach zweiseitiger (6.1) und einseitiger Fragestellung (6.2).

6.1 Zweiseitige Fragestellung

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	$\Theta = \left(IR^+\right)^2$	bzw.	$\Theta = IR^+$
Nullhypothese:	$H_0: \sigma_1^2 = \sigma_2^2$		$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$
Alternativhypothese:	$H_A: \sigma_1^2 \neq \sigma_2^2$		$H_A: \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$
Signifikanzniveau:	$\alpha = 0,05$		$\alpha = 0,05$

2. Schritt: Prüfgröße, Testverteilung

Die Varianzen bzw. ihre (erwartungstreuen) Schätzer $\hat{\sigma}_1^2 = \dfrac{n_1}{n_1 - 1} S_1^2$ und $\hat{\sigma}_2^2 = \dfrac{n_2}{n_2 - 1} S_2^2$ sind über S_1^2 und S_2^2 voneinander unabhängig χ^2-verteilte Zufallsvariablen. Da gilt, daß der Quotient zweier voneinander unabhängigen χ^2-verteilten und durch ihre Freiheitsgrade dividierten Zufallsvariablen F-verteilt ist,[9] kann als Prüfgröße herangezogen werden:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{n_1 \cdot S_1^2 / (n_1 - 1)}{n_2 \cdot S_2^2 / (n_2 - 1)}.$$

Diese Prüfgröße ist unter H_0: $\sigma_1^2 = \sigma_2^2$ F-verteilt mit $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ Freiheitsgraden: $F \sim F(\nu_1, \nu_2)$.

3. Schritt: Kritischer Bereich

Je mehr der Quotient $\hat{\sigma}_1^2 / \hat{\sigma}_2^2$ von 1 abweicht, desto eher wird man $H_0: \sigma_1^2 / \sigma_2^2 = 1$ ablehnen. Weil $F \in IR^+$ ist, wird man H_0 ablehnen, wenn der beobachtete F-Wert nahe bei 0 liegt oder sehr groß wird.

Kritischer Bereich:

$$[0, F_u] \cup [F_o, \infty).$$

9 Vgl. hierzu auch den Abschnitt V.3.

Da es sich hier um einen zweiseitigen Test handelt, wird α symmetrisch aufgeteilt:

$$F_o = F_{1-\alpha/2;\nu_1,\nu_2} \text{ und } F_u = F_{\alpha/2;\nu_1,\nu_2}.$$

Für einige α findet man die tabellierten Werte von $F_{1-\alpha/2;\nu_1,\nu_2}$ (für Werte größer als eins), nicht jedoch für $F_{\alpha/2;\nu_1,\nu_2}$ (für Werte kleiner als eins). Diesen Wert berechnet man durch

$$F_u = \frac{1}{F_{1-\alpha/2;\nu_2,\nu_1}}.$$

Dies gilt, weil

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{\nu_1,\nu_2}$$

$$G = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{1}{F} \sim F_{\nu_2,\nu_1}$$

$$\frac{\alpha}{2} = P(F \le F_u) = P\left(\frac{1}{G} \le F_u\right) = P\left(G \ge \frac{1}{F_u}\right) = 1 - P\left(G \le \frac{1}{F_u}\right) \text{ist.}$$

Also gilt:

$$P\left(G \le \frac{1}{F_u}\right) = 1 - \frac{\alpha}{2}.$$

Da G ~ F_{ν_2,ν_1}, folgt $\frac{1}{F_u} = F_{1-\alpha/2;\nu_2,\nu_1}$. Also ist

$$F_u = \frac{1}{F_{1-\alpha/2;\nu_2,\nu_1}}.$$

Aus der Tabelle der F-Verteilung erhält man für das Beispiel:

$$F_o = F_{0,975;20,15} = 2,7559 \text{ und}$$

$$F_u = \frac{1}{F_{0,975;15,20}} = \frac{1}{2,5731} = 0,3886.$$

4. Schritt: Wert der Prüfgröße

In unserem Beispiel haben wir die Werte

$$s_1 = 322 \ , \ n_1 = 21$$
$$s_2 = 288 \ , \ n_2 = 16.$$

Die erwartungstreuen Schätzer für die Varianzen sind:

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} \cdot s_1^2 = \frac{21}{20} \cdot 322^2 = 108868,2$$

$$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} \cdot s_2^2 = \frac{16}{15} \cdot 288^2 = 88473,6$$

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{108868,2}{88473,6} = 1,23.$$

5. Schritt: Entscheidung, Interpretation

Da $F_u = 0{,}3886 \le F = 1{,}23 \le 2{,}7559 = F_o$, kann die Hypothese „Die Einkommen in A und B besitzen die gleiche Streuung“ nicht abgelehnt werden, H_0 wird also beibehalten.

6.2 Einseitige Fragestellung

Wir wollen auch hier die einseitige Fragestellung untersuchen, d. h., wir wollen überprüfen, ob $H_0: \sigma_1^2 \le \sigma_2^2$ gilt.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = \mathrm{IR}^+$

Nullhypothese: $H_0: \dfrac{\sigma_1^2}{\sigma_2^2} \le 1$

Alternativhypothese: $H_A: \dfrac{\sigma_1^2}{\sigma_2^2} > 1$

Signifikanzniveau: $\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

$$F^* = \frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2} \text{ ist } F_{\nu_1;\nu_2} - \text{verteilt.}$$

Unter der Nullhypothese der zweiseitigen Fragestellung und für $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ der Hypothese der einseitigen Fragestellung erhält man durch Kürzen:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{\nu_1;\nu_2}.$$

Mit anderen Worten: Falls $\sigma_1^2 = \sigma_2^2$, kann die Prüfgröße F* = F ohne Kenntnis der Grundgesamtheitvarianzen berechnet werden. Die Prüfgröße entspricht der des zweiseitigen Tests.

3. Schritt: Kritischer Bereich

Der kritische Bereich hat die Form:

$$[F_c, +\infty).$$

Behauptung: Der Test hält das Signifikanzniveau ein, wenn man $F_c = F_{1-\alpha;\nu_1,\nu_2}$ setzt.

Beweis:

Unter H_0 gilt:

$$\alpha = P(F^* > F_c)$$
$$= P\left(\frac{\hat{\sigma}_1^2 \cdot \sigma_2^2}{\hat{\sigma}_2^2 \cdot \sigma_1^2} > F_c\right)$$
$$= P\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > \underbrace{F_c \cdot \underbrace{\frac{\sigma_1^2}{\sigma_2^2}}_{\leq 1}}_{\leq F_c}\right)$$
$$\geq P\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > F_c\right) = P(F > F_c).$$

Hier ergibt sich:

$$F_c = F_{0,95;20,15} = 2,33$$

4. Schritt: Wert der Prüfgröße

Wir erhalten hier wie bei der zweiseitigen Fragestellung F = 1,23.

5. Schritt: Entscheidung, Interpretation

Da F = 1,23 < 2,33 = F_c ist, kann H_0 (wie bei dem zweiseitigen Test) nicht abgelehnt werden.

7 Zweistichprobentests für die Differenz zweier Anteilswerte

Wir wollen hierbei die folgenden Voraussetzungen treffen:

- Die beiden Stichproben sind voneinander unabhängig.
- Die Stichprobenumfänge n_1 und n_2 sollen so groß sein, daß die Anteilswerte als normalverteilt angesehen werden können.

Als **Beispiel** zur Entwicklung des Zweistichprobentests soll der folgende Zusammenhang untersucht werden:

In zwei Vororten von Hamburg wurde das Jahreseinkommen der Haushalte erhoben. Im Vorort A wurden bei einem Stichprobenumfang von n_1 = 400 x_1 = 39 Haushalte gezählt, die ein Jahreseinkommen von mehr als 60.000 DM erzielten. Im Vorort B ergab eine Stichprobe vom Umfang n_2 = 300 x_2 = 45 Haushalte mit einem Jahreseinkommen von mehr als 60.000 DM.

Ist der Anteil der Haushalte mit einem Jahreseinkommen von mehr als 60.000 DM in beiden Vororten verschieden groß (α = 0,05)?

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	$\Theta = [0,1] \times [0,1]$	bzw.	$\Theta = [-1,1]$
Nullhypothese:	$H_0: p_1 = p_2$		$H_0: p_1 - p_2 = 0$
Alternativhypothese:	$H_A: p_1 \neq p_2$		$H_A: p_1 - p_2 \neq 0$

Signifikanzniveau: $\alpha = 0,05$ $\alpha = 0,05$

2. Schritt: Prüfgröße, Testverteilung

Unter den Modellvoraussetzungen ist

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

standardnormalverteilt, wobei $\hat{p}_i = \frac{x_i}{n_i}$, $i = 1,2$.

Mit der Gültigkeit von $H_0: p_1 = p_2 = p$ ergibt sich:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p \cdot (1-p)}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \sim N(0,1).$$

Da p nicht bekannt ist, muß p geschätzt werden:

$$\hat{p} = \frac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

Mit dieser Schätzung erhält man als Prüfgröße

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot (1-\hat{p})}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}},$$

die annähernd standardnormalverteilt ist.

3. Schritt: Kritischer Bereich

Wir haben hier eine zweiseitige Fragestellung, so daß α symmetrisch aufgeteilt wird. Als kritischer Bereich ergibt sich hier:

$$(-\infty, z_{\alpha/2}] \cup [z_{1-\alpha/2}, +\infty)$$

Für α = 0,05 entnimmt man der Tabelle

$$z_o = z_{1-\alpha/2} = z_{0,975} = 1,96$$
$$z_u = z_{\alpha/2} = -z_{1-\alpha/2} = -1,96.$$

4. Schritt: Wert der Prüfgröße

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{39}{400} = 0,0975$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{45}{300} = 0,15$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{39 + 45}{400 + 300} = 0,12$$

$$z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\cdot(1-\hat{p})}}\cdot\sqrt{\frac{n_1\cdot n_2}{n_1+n_2}}$$

$$=\frac{0{,}0975-0{,}15}{\sqrt{0{,}12\cdot 0{,}88}}\cdot\sqrt{\frac{400\cdot 300}{400+300}}$$

$$=\frac{-0{,}0525}{0{,}325}\cdot 13{,}093$$

$$=-\,2{,}1150$$

5. Schritt: Entscheidung, Interpretation

Da $z = -2{,}115 < -1{,}96 = z_u$ ist, wird die Hypothese H_0: $p_1 = p_2$ verworfen. Die Hypothese „Der Anteil der Haushalte mit einem Jahreseinkommen von über 60.000 DM ist in beiden Vororten gleich“ wird also abgelehnt, d. h., H_A: $p_1 \neq p_2$ wird angenommen.

Beispiel:———————————————————————————

ET: Hypothesis Tests for Two Variables
Tests: t, F, fits, distributions (2 SAMPLE)

Aus zwei Stichproben seien wöchentliche Arbeitszeiten (HOURS1, HOURS2) (wie unter DATA LISTING) ermittelt worden. Die Stichproben enthalten unabhängige Beobachtungen. Wie lauten die ET-Testentscheidungen auf Gleichheit/Ungleichheit der Stichproben?

Generell (ET):

Prob $< \alpha$ vorgegeben $\Rightarrow$ H_0 ablehnen
oder
Prob $> \alpha$ vorgegeben $\Rightarrow$ H_0 beibehalten.

```
DATA LISTING (Current sample)
Observation      HOURS1      HOURS2
      1        38.000      32.000
      2        40.000      34.000
      3        42.000      28.000
      4        18.000      15.000
      5        20.000      15.000
      6        35.000      22.000
      7        19.000      23.000
      8        36.000      25.000
      9        44.000      32.000
     10        38.000      24.000
     11        40.000      28.000
     12        44.000      36.000
     13        48.000      35.000
     14        28.000      27.000
     15        43.000      36.000
```

```
   Tests based on 2 variables
 X1=HOURS1           X2=HOURS2

  0 OK, now compute test results
  1 Equality of means,     t test
  2 Equality of variances, F test
  3 Paired t test for equal means
  4 Z and t test of 0 correlation
  5 Kol.-Smir. test, same distrib
  6 ANOVA, F test for equal means
  7 Specify different variables

  Choose test(s). ESC = Exit.
```

```
 1 Equality of means,     t test
Means     are     35.5333 and      27.4667
Variances are     94.8381 and      47.5524
Pooled variance =     71.1952
If variances are     assumed to be equal
t[ 28] =     2.62, Prob = .0141
If variances are not assumed to be equal
t[ 25] =     2.62, Prob = .0148
```

```
 2 Equality of variances, F test
Means     are     35.5333 and      27.4667
Variances are     94.8381 and      47.5524

F ratio [ 14, 14] =   1.994, Prob =  .1044
```

```
3 Paired t test for equal means
Mean difference    =      8.0667
Standard deviation =      1.3783
t[ 15] =     5.85, Prob = .0000
```

```
 4 z and t test of 0 correlation
Correlation =    .8480
Approximate t[ 13]=  5.769, Prob =    .000
Fisher z (normal) =  4.327, Prob =    .000
```

Beispiel:

SPSS: Mittelwertvergleich zweier unabhängiger Stickproben

Gruppenstatistiken

	SAMPLENO	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
HOURS	1,00	15	35,5333	9,73849	2,51447
	2,00	15	27,4667	6,89582	1,78049

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
HOURS	Varianzen sind gleich	1,539	,225	2,618	28	,014	8,06667	3,08102	1,75548	14,37785
	Varianzen sind nicht gleich			2,618	25,219	,015	8,06667	3,08102	1,72397	14,40936

1. Test auf Gleichheit der Mittelwerte
   bei Annahme gleicher Grundgesamtheitsvarianzen:

   $t(\nu = 28) = 2.62, \Pr ob = .0141$

   $H_0: \overline{X}_1 = \overline{X}_2$, mit $\alpha \leq 1,41\%$ kann H_0 verworfen werden.

   $H_A: \overline{X}_1 \neq \overline{X}_2$

   bei Annahme nicht gleicher Grundgesamtheitsvarianzen:

   $t(\nu = 25)$ H_0 wird mit $\alpha \leq 1,48\%$ verworfen.

2. Test auf Gleichheit der Varianzen (F-Test)

   $H_0: Var(X_1) = Var(X_2)$

   $H_A: Var(X_1) \neq Var(X_2)$

   H_0 kann mit $\alpha \leq 10,44\,\%$ verworfen werden.

3. Gepaarter t-Test auf Gleichheit der Mittelwerte
   Paarweise (z. B. gleiche Person produziere beide Testreihen)

   $H_0: \overline{X}_1 = \overline{X}_2$ wird mit $\alpha \leq .0000$ verworfen.

4. Test auf Null-Korrelation

   $$H_0: r^2 = \frac{\sigma_{X_1 X_2}}{\sigma_{X_1} \cdot \sigma_{X_2}} = 0$$

   $H_A: r^2 \neq 0$

Für einen approximativen t-Test und Fishers standardnormalverteilten z-Wert zeigen größere errechnete als kritische Tabellenwerte an, daß H_0 einer Nullkorrelation abgelehnt wird.

Mit $r^2 = 0{,}8480$ und hohen t- bzw. z-Werten wird H_0 bei $\alpha \leq .000$ verworfen zugunsten einer Assoziation/Korrelation zwischen HOURS1 und HOURS2 (kleine (große) Werte HOURS1 entsprechen kleinen (großen) Werten von HOURS2).

8 Tests im klassischen linearen Regressionsmodell

Das klassische lineare Regressionsmodell spezifiziert einen Erklärungsansatz aus[10]

$$y_i = \underbrace{\beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_k \cdot x_{ik} + \ldots + \beta_K \cdot x_{iK}}_{\text{deterministischer Teil}} + \underbrace{\varepsilon_i}_{\text{stochastischer Teil}} \quad ,(i=1,\ldots,n)$$

mit den zu schätzenden Parametern β_k der Grundgesamtheit (k = 0, …, K).

Der Zufallsterm (Störterm, Errorterm, Residual) ε genügt einer bestimmten Verteilung, wobei häufig die Normalverteilung verwendet wird.

Das klassische lineare Regressionsmodell ist also ein stochastischer Ansatz. y_i wird zum Teil stochastisch erklärt. Die Koeffizienten β_k der Regressionsgleichung werden mit Hilfe der OLS-Methode geschätzt:

$$b_{OLS} = (X'X)^{-1} X'y$$

Man erhält als geschätzte Regressionsgleichung:

$$\hat{y}_i = b_0 + b_1 \cdot x_{i1} + b_2 \cdot x_{i2} + \ldots + b_k \cdot x_{ik} + \ldots + b_K \cdot x_{iK}.$$

Die geschätzten Koeffizienten b_k (k = 0, …, K) sind dabei stochastisch.

Die Schätzergebnisse sollten statistisch abgesichert, d.h., sie sollten getestet werden. Dies geschieht unter anderem mit Hilfe von zwei Tests:

- Gesamterklärungsgüte: Test auf R^2 (F-Test);
- Signifikanztest der geschätzten Parameter b_k (t-Test).

8.1 Test der Gesamterklärungsgüte R^2 (F-Test)

Das Bestimmtheitsmaß B = R^2 ist ein Maß für die Anpassungsgüte ('goodness of fit'):

$$B = R^2 = \frac{\text{erklärte Varianz}}{\text{Gesamtvarianz}} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{s_{\hat{y}}^2}{s_y^2}.$$

B = R^2 ist normiert: $0 \leq B \leq 1$.

Je größer B ist, desto besser ist die Anpassung.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = IR^K$

[10] Vgl. hierzu auch den Abschnitt VI.3 und Merz, J., Regressionsanalyse – Einführung in die Ökonometrie, Skriptum, Lüneburg 1993 und Merz, J., Statistik I – Deskription, Skriptum, Lüneburg 1994.

Nullhypothese: $H_0: \beta_1 = \beta_2 = ... = \beta_k = ... = \beta_K = 0$

Alternativhypothese: $H_A: \beta_k \neq 0 \ (k = 1, ..., K)$

Signifikanzniveau: $\alpha = 0,05$

Es wird also getestet, ob alle Parameter außer der Konstanten β_0 Null sind. Statt $\alpha = 0,05$ kann natürlich auch ein anderes Signifikanzniveau gewählt werden.

2. Schritt: Prüfgröße, Testverteilung

Prüfgröße:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{(n - K - 1)}{K}.$$

Diese Prüfgröße ist F-verteilt mit $\nu_1 = K$ und $\nu_2 = n - K - 1$ Freiheitsgraden:

$$F \sim F_{1-\alpha;\, K;\, n-K-1}.$$

3. Schritt: Kritischer Bereich

Kritischer Bereich:

$$[F_c; +\infty).$$

Den kritischen Wert F_c entnimmt man der Tabelle der F-Verteilung. $(F_c = F_{1-\alpha; K; n-K-1})$.

4. Schritt: Wert der Prüfgröße

$$F_{beob.} = \frac{R^2}{1 - R^2} \cdot \frac{n - K - 1}{K}.$$

Je größer R^2, desto eher liegt der Wert der Prüfgröße im kritschen Bereich (einseitiger Test). Diesen Wert liefert auch der Output eines statistischen Programmpakets (z.B. ET).

5. Schritt: Entscheidung, Interpretation

H_0 wird verworfen, falls

$$F_{beob.} > F_c.$$

H_0 kann nicht abgelehnt werden, falls

$$F_{beob.} \leq F_c.$$

Wenn H_0 verworfen wird, leistet der Erklärungsansatz einen signifikanten Beitrag zur Erklärung von y_i. Kann H_0 dagegen nicht abgelehnt werden, ist der Erklärungsansatz neu zu spezifizieren.

8.2 Signifikanztest für die einzelnen MKQ/OLS-Koeffizienten b_k (t-Test)

Ausgehend von der OLS-Schätzung

$$b_{OLS} = (X'X)^{-1} X'y$$

überprüft man nun die Signifikanz der *einzelnen* Koeffizienten β_k.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: $\Theta = \mathrm{IR}$

Nullhypothese: $H_0 : \beta_k = 0$

Alternativhypothese: $H_A : \beta_k \neq 0$

Signifikanzniveau: $\alpha = 0{,}05$

Es wird also ein einzelner Koeffizient β_k getestet, ob er signifikant von Null verschieden ist. Statt α = 0,05 kann natürlich auch hier ein anderes Signifikanzniveau gewählt werden.

2. Schritt: Prüfgröße, Testverteilung

Prüfgröße:

$$t = \frac{b_k - (\beta_k = 0)}{s_{b_k}},$$

wobei

$$s_{b_k} = \sqrt{\hat{Var}(b_k)} = \frac{e'e}{n-K-1}(X'X)^{-1}_{kk}$$

und

$$e = y - \hat{y}.$$

Die Prüfgröße t ist t-verteilt mit ν = n – K –1 Freiheitsgraden.

Im Computerprogramm ET wird die Prüfgröße t als t-ratio bezeichnet und angegeben.

3. Schritt: Kritischer Bereich

Kritischer Bereich:

$$]-\infty; -t_c] \cup [t_c; +\infty[$$

Die kritischen Werte entnimmt man der Tabelle der t-Verteilung ($t_c = t_{1-\alpha/2, n-K-1}$). Man verwendet hier α/2, weil es sich um einen zweiseitigen Test handelt.

4. Schritt: Wert der Prüfgröße

$$t_{beob.} = \frac{b_k - 0}{s_{b_k}}$$

Dieser Wert muß entweder ausgerechnet oder dem Output eines statistischen Programmpakets entnommen werden (z.B. ET).

5. Schritt: Entscheidung, Interpretation

H_0 wird verworfen, falls

$$|t_{beob.}| > t_c.$$

H_0 kann nicht abgelehnt werden, falls

$$|t_{beob.}| \leq t_c.$$

Wenn H_0 verworfen wird, ist β_k signifikant von Null verschieden, d.h., die Variable x_k leistet einen signifikanten Beitrag zur Erklärung von y.

Kann H_0 dagegen nicht abgelehnt werden, dann leistet die erklärende Variable keinen signfikanten Beitrag zur Erklärung von y.

Beispiel:

Das individuelle Arbeitsangebot von Frauen und Männern soll anhand von Mikrodaten erklärt werden. Die mikroökonomische Theorie liefert einen Erklärungsansatz

$$\text{hours} = f\left(\text{wage, sozioökonomische Charakteristika}\right) + \varepsilon .$$

Mit ET werden anhand von n = 15 Beobachtungen mit OLS die Parameter geschätzt (vgl. das folgende ET-Ergebnisprotokoll), wobei

hours = Wochenstunden
wage = Stundenlohn
sex = 0 = Männer, 1 = Frauen
age = Alter
hhsize = Haushaltsgröße

ET-Befehle:

```
? --------------------------------
? ET hours: ols by matrix algebra
? --------------------------------
?
? read data
?      y : hours
?      x0: one
?      x1: wage
?      x2: sex (0=male,1=female)
?      x3: age
?      x4: hhsize
?
read; file=hours.dat;nvar=5;nobs=15 ;names=1$
list; hours, wage, sex, age, hhsize$
?
? create X'X (=XSX) and X'y (XSy)
? ------------------------------
namelist; X=one,wage,sex,age,hhsize$
matrix; XSX=xdot(one,wage,sex,age,hhsize)$
matrix; XSy=xdot(X,hours)$
?
? create invers of (X'X)
? ------------------------------
matrix; XSXINV=ginv(XSX)$
?
? compute ols
? ------------------------------
matrix; bols=XSXINV | XSy$
?
? regression by et
? ------------------------------
regres; dep=hours; ind=one,wage,sex,age,hhsize$
```

ET-Ergebnis:

```
DATA LISTING (Current sample)
Observation      HOURS        WAGE         SEX          AGE          HHSIZE
      1         38.000       20.000       .00000       27.000       2.0000
      2         40.000       24.000       .00000       32.000       3.0000
      3         42.000       28.000       .00000       35.000       4.0000
      4         18.000       15.000       .00000       22.000       1.0000
      5         20.000       15.000       1.0000       21.000       1.0000
      6         35.000       22.000       1.0000       27.000       3.0000
      7         19.000       23.000       1.0000       32.000       4.0000
      8         36.000       25.000       .00000       34.000       3.0000
      9         44.000       32.000       1.0000       45.000       3.0000
     10         38.000       16.000       1.0000       48.000       3.0000
     11         40.000       28.000       1.0000       36.000       4.0000
     12         44.000       36.000       1.0000       38.000       5.0000
     13         48.000       35.000       1.0000       35.000       1.0000
     14         28.000       22.000       .00000       29.000       3.0000
     15         43.000       42.000       1.0000       34.000       2.0000

     1. Matrix -> XSX=XDOT(ONE,WAGE,SEX,AGE,HHSIZE)


      <<<< XSX      >>>>   COLUMN
               1             2             3             4             5
 ROW   1   15.0000       383.000       9.00000       495.000       42.0000
 ROW   2   383.000       10681.0       249.000       12988.0       1108.00
 ROW   3   9.00000       249.000       9.00000       316.000       26.0000
 ROW   4   495.000       12988.0       316.000       17103.0       1447.00
 ROW   5   42.0000       1108.00       26.0000       1447.00       138.000


     1. Matrix -> XSY=XDOT(X,HOURS)

      <<<< XSY      >>>>   COLUMN
               1
 ROW   1   533.000
 ROW   2   14395.0
 ROW   3   331.000
 ROW   4   18239.0
 ROW   5   1535.00

     1. Matrix -> XSXINV=GINV(XSX)

     <<<< XSXINV   >>>>   COLUMN
              1             2             3             4             5
 ROW  1    1.67913      -.159590E-01  .123515      -.389481E-01  .221385E-02
 ROW  2  -.159590E-01   .142732E-02 -.503406E-02  -.448413E-03 -.952554E-03
 ROW  3    .123515      -.503406E-02  .340583      -.757404E-02  .180766E-01
 ROW  4  -.389481E-01  -.448413E-03 -.757404E-02   .210819E-02 -.522435E-02
 ROW  5    .221385E-02  -.952554E-03  .180766E-01  -.522435E-02  .655948E-01

     1. Matrix -> BOLS=XSXINV|XSY

      <<<< BOLS     >>>>   COLUMN
               1
 ROW   1  -.844741
 ROW   2   .733063
 ROW   3  -4.29396
 ROW   4   .710601
 ROW   5  -1.14748
```

```
============================================================================
Ordinary    Least Squares
Dependent Variable        HOURS      Number of Observations             15
Mean of Dep. Variable      35.5333   Std. Dev. of Dep. Var.       9.738485
Std. Error of Regr.         6.2195   Sum of Squared Residuals   386.825
R - squared                 .70866   Adjusted R - squared           .59212
F( 4,  10)                  6.0810   Prob. Value for F              .00954
============================================================================
Variable  Coefficient Std. Error  t-ratio Prob|t|>x Mean of X  Std.Dev.of X
----------------------------------------------------------------------------
Constant  -.844767      8.059       -.105   .91859
WAGE        .733063      .2350       3.120   .01088   25.53333       8.02555
SEX       -4.29397      3.630      -1.183   .26417     .60000        .50709
AGE         .710602      .2856       2.488   .03208   33.00000       7.40656
HHSIZE    -1.14748      1.593       -.720   .48780    2.80000       1.20712
```

F-Test auf Gesamterklärungsgüte:

$F_{beob.} = F(4,10) = 6{,}0810$

Prob. Value for $\mathrm{F} = 0{,}00954$

d.h., ab einem Signifikanzniveau von 0,95 % kann $H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$ zugunsten von $H_A : \beta_k \neq 0 \left(k = 1, \ldots, K = 4\right)$ abgelehnt werden.

Für $\alpha = 5\%$ gibt es einen signifikanten Gesamterklärungsansatz.

t-Test auf Signifikanz der einzelnen Parameter:

Die $t_{beob.}$ sind als t-ratio zu jedem $b_k \left(k = 0, \ldots, 4\right)$ sowie die zugehörige Irrtumswahrscheinlichkeit $\alpha = \mathrm{Prob}\left|\,\mathrm{t}\,\right| > \mathrm{x}$ angegeben.

Für $\alpha = 5\%$ sind somit wage und age signifikant von Null verschieden.

$t_c\left(\alpha = 0{,}05\right) = t_c\left(10{,}0{,}05\right) = 2{,}2281 \qquad \nu = n - k - 1$

Beispiel:

SPSS-Ausdruck

Modellzusammenfassung				
Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,842[a]	,709	,592	6,21952
a. Einflußvariablen : (Konstante), HHSIZE, SEX, WAGE, AGE				

ANOVA[b]						
Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	940,909	4	235,227	6,081	,010[a]
	Nicht standardisierte Residuen	386,825	10	38,682		
	Gesamt	1327,733	14			
a. Einflußvariablen : (Konstante), HHSIZE, SEX, WAGE, AGE						
b. Abhängige Variable: HOURS						

Koeffizienten[a]						
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		RegressionskoeffizientB	Standardfehler	Beta		
1	(Konstante)	-,845	8,059		-,105	,919
	WAGE	,733	,235	,604	3,120	,011
	SEX	-4,294	3,630	-,224	-1,183	,264
	AGE	,711	,286	,540	2,488	,032
	HHSIZE	-1,147	1,593	-,142	-,720	,488
a. Abhängige Variable: HOURS						

8.3 p-value/prob-value und Testentscheidung

Eine gewählte Irrtumswahrscheinlichkeit α. bestimmt den zugehörigen kritischen Wert einer Testverteilung.

Zu dem kritischen Wert einer Testverteilung generell gehört also eine bestimmte Irrtumswahrscheinlichkeit α, das Signifikanzniveau des Tests.

Zu dem Prüfwert aus dem aktuellen Problem und der zugehörigen konkreten Stichprobe gehört die sogenannte exakte Irrtumswahrscheinlichkeit, die auch als p-value oder probability-value bezeichnet wird.

Der p-value ist also der für diese Stichprobe exakte Fehler 1. Art. Damit kann eine Testentscheidung über zwei Wege gefunden werden:

H0 wird nicht abgelehnt, wenn

Wert der Prüfgröße im Konfidenzbereich
oder
p-value > Signifikanzniveau

H_0 wird abgelehnt, wenn

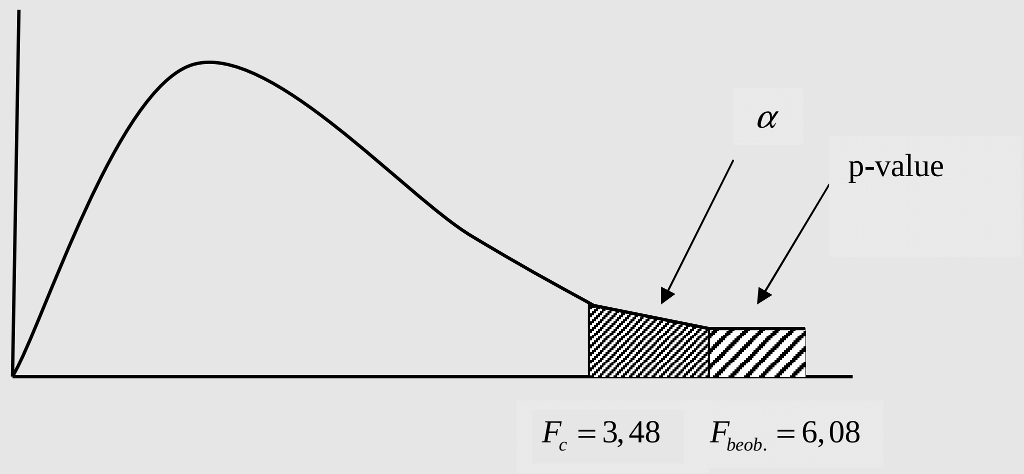
: Wert der Prüfgröße im kritischen Bereich
oder
p-value < Signifikanzniveau

Beispiel:

Regression zum Arbeitsangebot

F-Test (einseitiger Test)

So gehört im obigen Arbeitsangebotsbeispiel für den Test der Gesamterklärungsgüte mittels eines F-Tests zu dem Prüfwert des Beispiels Fbeob.= F(4,10) = 6,081 der p-value von 0,00954.



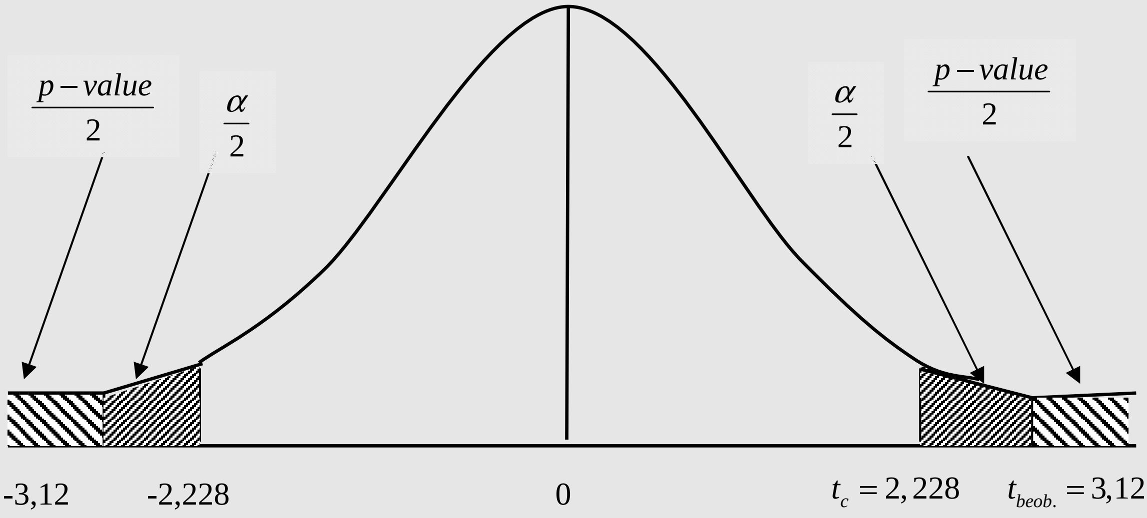
Testentscheidung: Fbeob. = F(4,10) = 6,081 > Fc = 3,478 (α=0,05)

p-value = 0,00954 < α=0,05

Also: H_0 = „kein signifikanter Gesamterklärungsansatz“ wird zugunsten einer signifikanten Gesamterklärungskraft abgelehnt

t-Test (zweiseitiger Test):

Im obigen Arbeitsangebotsbeispiel ist für den geschätzten WAGE-Koeffizienten der Prüfwert t_{beob} = 3,120 der entsprechende p-value 0,01088.



Testentscheidung: tbeob. = 3,120 > tc = 2,2281 (ν= n-K-1=10, α=0,05)

p-value = 0,01088 < α=0,05 bzw.

p-value/2 = 0,00544 < α/2=0,025

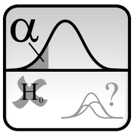
Also: H_0 = „kein signifikanter Einfluss des Lohnsatzes" wird zugunsten eines signifikanten Einflusses abgelehnt

---

Keyconcepts

Prinzip und Aufbau eines Tests
Kritischer Bereich
Prufgröße
Fehler 1. Art (α-Fehler)
Fehler 2. Art (β-Fehler)
Signifikanzniveau
Irrtumswahrscheinlichkeit
Güte eines Tests
Einstichprobentest: Anteilswert, arithmetisches Mittel und Varianz
Zweistichprobentest für die Differenz zweier arithmetischer Mittelwerte, Quotienten zweier Varianzen und Differenz zweier Annteilswerte
Test im klassischen linearen Regressionsmodell: F-Test, t-Test
P-Value/Prob-Value

IX Verteilungstests



Annahmen über die Verteilung eines Merkmals in der Grundgesamtheit überprüfen

Mit Verteilungstests wird überprüft, ob die in einer Stichprobe beobachtete Verteilung mit einer theoretischen Verteilung aus der Grundgesamtheit übereinstimmt. Man untersucht also die **Güte der Anpassung** einer empirischen Verteilung an eine theoretische Verteilung.

1 Chi-Quadrat-Verteilungstest

1.1 Einfache Hypothesen

Wir wollen ein Beispiel für eine diskrete Verteilung einer Grundgesamtheit betrachten. Bei 90 Würfen eines Würfels seien folgende absolute Häufigkeiten beobachtet worden:

Augenzahl	1	2	3	4	5	6
n_i	19	13	14	12	17	15

Es soll überprüft werden, ob ein idealer (symmetrischer) Würfel den Würfen zugrundeliegt $(\alpha=0{,}05)$.

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: Menge aller Verteilungen auf $\{1,2,\ldots,6\}$

Nullhypothese: H_0: Augenzahlen sind gleichverteilt auf der Menge $\{1,2,\ldots,6\}$

Alternativhypothese**:** H_A: Augenzahlen sind nicht gleichverteilt auf der Menge $\{1,2,\ldots,6\}$

Signifikanzniveau: $\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Unter der Gültigkeit von H_0 würde jede Augenzahl gleich häufig auftreten. Theoretische Häufigkeiten: $n \cdot p_i = 90 \cdot \frac{1}{6} = 15$ für $i = 1, 2, \ldots, 6$

Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}.$$

Diese Prüfgröße ist χ^2-verteilt mit $\nu = k - 1$ Freiheitsgraden, wobei k die Anzahl der Merkmalsausprägungen darstellt.

Anmerkung: Die Anzahl der Freiheitsgrade ist $\nu = k - 1$, weil mit der Kenntnis von $p_1, \ldots, p_{k-1}$ auch p_k festgelegt ist.

Man sollte bei der Anwendung folgende Faustregel beachten:

$$n \cdot p_i \geq 5 \quad \text{für } i = 1, \ldots, k$$

3. Schritt: Kritischer Bereich

Kritischer Bereich:

$$\left[\chi_c^2, +\infty\right).$$

Bei einem Signifikanzniveau von $\alpha = 0{,}05$ und $\nu = 6 - 1 = 5$ erhält man aus der Tabelle zur χ^2-Verteilung den Wert

$$\chi_c^2 = 11{,}07.$$

Einseitiger Test, da Prüfgröße bei größerem (+,–) Abstand 'einseitig' immer größer wird.

4. Schritt: Wert der Prüfgröße

Man berechnet den normierten quadrierten „Abstand“ zwischen den Verteilungen:

$$\begin{aligned}\chi^2 &= \frac{(19-15)^2}{15} + \frac{(13-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(12-15)^2}{15} + \frac{(17-15)^2}{15} + \frac{(15-15)^2}{15} \\ &= \frac{16+4+1+9+4+0}{15} \\ &= 2{,}267\end{aligned}$$

5. Schritt: Entscheidung, Interpretation

Da $\chi^2 = 2{,}267 \leq 11{,}07 = \chi_c^2$, kann H_0 nicht abgelehnt werden, d.h., man kann davon ausgehen, daß ein idealer Würfel vorliegt.

1.2 Zusammengesetzte Hypothesen

Als Beispiel soll die Frage untersucht werden, ob die Lebensdauer eines bestimmten Bauteils einer Maschine normalverteilt ist ($\alpha = 0{,}05$).

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge:	Menge aller möglichen Verteilungen ist die Menge aller Verteilungen auf IR
Nullhypothese:	H_0: Die unbekannte Verteilung ist eine Normalverteilung
Alternativhypothese:	H_A: Die unbekannte Verteilung ist eine Verteilung auf IR, aber keine Normalverteilung
Signifikanzniveau:	$\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Wiederum sollte die Prüfgröße, wie schon im Fall der einfachen Hypothese, die „Entfernung“ zwischen theoretischer und empirischer Verteilung messen. Zunächst muß die Frage beantwortet werden, welche Verteilung aus H_0 als theoretische Verteilung herangezogen wird.

Um bei unserem Beispiel zu bleiben: Welche Normalverteilung ist hier die theoretische Verteilung? Die Normalverteilung hängt von den Parametern μ und σ^2 ab. Über eine Parameterschätzung kann man μ und σ^2 schätzen: $\hat{\mu}, \hat{\sigma}^2$. Die theoretische Verteilung ist dann $N\left(\hat{\mu}, \hat{\sigma}^2\right)$.

Allgemein wählt man über eine Schätzung eine Verteilung aus H_0 aus. Wenn sich die Verteilungen aus H_0 nur durch einen oder mehrere Parameter unterscheiden, kann man diese Parameter am besten mit Hilfe der Maximum-Likelihood-Methode schätzen.

In unserem Beispiel soll sich bei 80 Beobachtungen ein arithmetisches Mittel $\overline{x} = 3,41$ und eine Varianz $s^2 = 0,49$ ergeben haben. Als theoretische Verteilung verwenden wir daher $N(3,41;0,49)$.

Weil hier ein stetiges Merkmal vorliegt, führen wir eine Unterteilung in k Klassen durch:

Klasse i	Lebensdauer in Jahren $x_i^u < x \le x_i^o$	Anzahl der Bauteile n_i
1	$0 < x \le 1,95$	4
2	$1,95 < x \le 2,45$	2
3	$2,45 < x \le 2,95$	8
4	$2,95 < x \le 3,45$	30
5	$3,45 < x \le 3,95$	20
6	$3,95 < x \le 4,45$	10
7	$x > 4,45$	6
	$\sum$	80

Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i},$$

die mit $\nu = k - 1 - m$ Freiheitsgraden χ^2-verteilt ist, wobei m die Anzahl der geschätzten Parameter der Verteilung darstellt. Im Beispiel ist m = 2.

Zur Berechnung von χ^2 sind die theoretischen relativen Häufigkeiten p_i in der Klasse i zu ermitteln. Dazu bestimmt man die Werte der Normalverteilung für die oberen Klassengrenzen x_i^o:

$$F_N(x_i^o \mid \hat{\mu}, \hat{\sigma}) = F_N(z_i^o),$$

wobei

$$z_i^o = \frac{x_i^o - \hat{\mu}}{\hat{\sigma}} \quad , \; i = 1, 2, \ldots, k$$

x_i^o	$z_i^o = \frac{x_i^o - 3,41}{0,7}$	$F_N = (z_i^o)$
1,95	-2,086	0,0185
2,45	-1,371	0,0852
2,95	-0,657	0,2556
3,45	0,057	0,5227
3,95	0,771	0,7797
4,45	1,486	0,9314
>4,45		1,0000

Nun ermitteln wir p_i:

$$p_i = F_N(z_i^o) - F_N(z_{i-1}^o)$$

Klasse i	Lebensdauer $x_i^u < x \le x_i^o$	Anzahl der Bauteile n_i	rel. Häufigkeiten p_i	$n \cdot p_i = 80 \cdot p_i$ (theoretische Anzahl)
	$0 < x \le 1{,}95$	4	0,0185	1,48
2	$1{,}95 < x \le 2{,}45$	2	0,0667	5,34
3	$2{,}45 < x \le 2{,}95$	8	0,1704	13,63
4	$2{,}95 < x \le 3{,}45$	30	0,2671	21,37
5	$3{,}45 < x \le 3{,}95$	20	0,2570	20,56
6	$3{,}95 < x \le 4{,}45$	10	0,1517	12,14
7	$x > 4{,}45$	6	0,0686	5,49
	$\sum$	80	1,0000	80,01

Als Faustregel für die Anwendung gilt:

$$n \cdot p_i \ge 5 \quad \text{für } i = 1, \ldots, k.$$

Da in der 1. Klasse $n \cdot p_i < 5$ ist, fassen wir die beiden ersten Klassen zu einer Klasse zusammen:

Klasse i	Lebensdauer $x_i^u < x \le x_i^o$	Anzahl der Bauteile n_i	rel. Häufigkeiten p_i	$n \cdot p_i = 80 \cdot p_i$ (theoretische Anzahl)
1	$0 < x \le 2{,}45$	6	0,0852	6,82
2	$2{,}45 < x \le 2{,}95$	8	0,1704	13,63
3	$2{,}95 < x \le 3{,}45$	30	0,2671	21,37
4	$3{,}45 < x \le 3{,}95$	20	0,2570	20,56
5	$3{,}95 < x \le 4{,}45$	10	0,1517	12,14
6	$x > 4{,}45$	6	0,0686	5,49

3. Schritt: Kritischer Bereich

Der kritische Bereich hat die Form:

$$[\chi_c^2; +\infty).$$

Wir haben nach der Zusammenfassung k = 6 Klassen. Die Zahl der geschätzten Parameter beträgt m = 2. Damit ergibt sich für die Zahl der Freiheitsgrade:

$$\nu = k - m - 1 = 6 - 2 - 1 = 3.$$

Da $\alpha = 0{,}05$ gewählt wurde, erhält man als kritischen Wert:

$$\chi_c^2 = 7{,}815$$

4. Schritt: Wert der Prüfgröße

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-n\cdot p_i)^2}{n\cdot p_i}$$
$$=\frac{(6-6{,}82)^2}{6{,}82}+\frac{(8-13{,}63)^2}{13{,}63}+\frac{(30-21{,}37)^2}{21{,}37}+\frac{(20-20{,}56)^2}{20{,}56}+\frac{(10-12{,}14)^2}{12{,}14}+\frac{(6-5{,}49)^2}{5{,}49}$$
$$=6{,}349$$

5. Schritt: Entscheidung, Interpretation

Da $\chi^2=6{,}349<\chi_c^2=7{,}815$, kann H_0 nicht abgelehnt werden. Die Lebensdauer des Bauteils ist also normalverteilt.

2 Kolmogorov-Smirnov-Verteilungstest

Auch der Kolmogorov-Smirnov-Anpassungstest vergleicht die empirische und theoretische Verteilungsfunktion. Er testet einfache Hypothesen. Im Gegensatz zum Chi-Quadrat-Anpassungstest kann er schon bei kleinen Stichproben angewendet werden.

Die Anwendung wird am folgenden Beispiel illustriert:

Es soll mit $\alpha = 0{,}05$ geprüft werden, ob die Körpergröße von zwölfjährigen Kindern normalverteilt ist. Eine Stichprobe vom Umfang n = 10 liefert die folgenden Werte (Körpergröße in m):

i	Körpergröße in m
1	1,66
2	1,38
3	1,56
4	1,53
5	1,47
6	1,59
7	1,34
8	1,57
9	1,41
10	1,49

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau

Parametermenge: Alle Verteilungen auf IR sind möglich.

Nullhypothese: H_0: Die Körpergröße 12-jähriger Kinder ist normalverteilt mit μ = 1,60 m und σ = 0,10 m.

Alternativhypothese**:** H_A: Die Körpergröße 12-jähriger Kinder ist **nicht** normalverteilt mit μ = 1,60 m und σ = 0,10 m.

Signifikanzniveau: $\alpha=0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Für die Konstruktion der Prüfgröße verwendet man die **Verteilungsfunktion der Grundgesamtheit** ($F^e(z)$, expected) und die **Summenhäufigkeitsfunktion der Stichprobe** ($F^o(z)$, observed). Als Prüfgröße wird jetzt die größte Abweichung der beobachteten von der erwarteten Verteilungsfunktion verwendet:

$$d = \max_z \left| F^e(z) - F^o(z) \right|$$
$$= \max_{z_i} \begin{cases} \left| F^e(z_i) - F^o(z_{i-1}) \right| \\ \left| F^e(z_i) - F^o(z_i) \right| \end{cases}$$

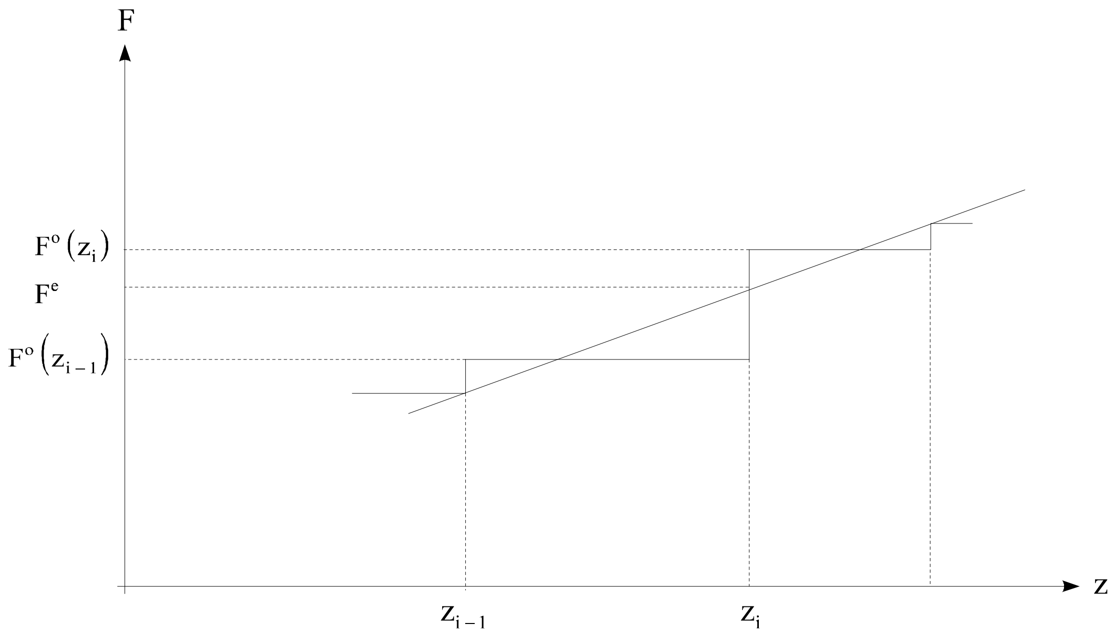


Abb.: Beobachtete und erwartete Verteilungsfunktion (Hujer 1991, S.212)

Die Verteilung der Prüfgröße d ist nach dem Satz von Kolmogorov für alle stetigen Verteilungen dieselbe. Sie ist nur vom Stichprobenumfang n abhängig.

3. Schritt: Kritischer Bereich

Der kritische Bereich hat die Form

$$[d_c; +\infty).$$

Als kritischer Wert ergibt sich für n = 10 und $\alpha = 0{,}05$ aus der Tabelle der Kolmogorov-Smirnov-Verteilung $d_c = 0{,}409$.

4. Schritt: Wert der Prüfgröße

x_i	$z_i = \frac{x_i - 1{,}6}{0{,}1}$	$F^o(z_i)$	$F^e(z_i)$	$\left\|F^e(z_i) - F^o(z_{i-1})\right\|$	$\left\|F^e(z_i) - F^o(z_i)\right\|$
1,34	–2,6	0,1	0,005	0,005	0,095
1,38	–2,2	0,2	0,014	0,086	0,186
1,41	–1,9	0,3	0,029	0,171	0,271
1,47	–1,3	0,4	0,097	0,203	0,303
1,49	–1,1	0,5	0,136	0,264	0,364
1,53	–0,7	0,6	0,242	0,258	0,358
1,56	–0,4	0,7	0,345	0,255	0,355
1,57	–0,3	0,8	0,382	0,318	0,418
1,59	–0,1	0,9	0,460	0,340	**d = 0,440**
1,66	0,6	1,0	0,726	0,174	0,274

z.B. $F^e(-2{,}6) = 1 - F^e(2{,}6) = 1 - 0{,}9953 = 0{,}0047 \approx 0{,}005$

5. Schritt: Entscheidung, Interpretation

Da d = 0,440 > 0,409 = d_c, wird H_0 verworfen. Die Körpergröße zwölfjähriger Kinder ist nicht normalverteilt mit μ = 1,60 m und σ = 0,10 m.

3 Chi-Quadrat-Unabhängigkeitstest

Der Chi-Quadrat-Unabhängigkeitstest untersucht, ob zwei nominalskalierte (qualitative) Merkmale voneinander stochastisch unabhängig sind.

Wir wollen auch hier anhand eines Beispiels vorgehen. Über einen spezifischen Eignungstest hat man entschieden, ob die Bewerber mit verschiedenen Abschlüssen (Diplom-Sozialökonom, Diplom-Kaufmann, Diplom-Volkswirt) für eine Aufgabe geeignet sind oder nicht. Die Beobachtungen wurden in der folgenden Kontingenztabelle festgehalten:

Abschluß (B) Eignung (A)	Diplom-Sozial-ökonom (B_1)	Diplom-Kauf-mann (B_2)	Diplom-Volkswirt (B_3)	$\sum$
geeignet (A_1)	n_{11} = 14	n_{12} = 10	n_{13} = 16	$n_{1.}$ = 40
ungeeignet(A_2)	n_{21} = 16	n_{22} = 25	n_{23} = 19	$n_{2.}$ = 60
$\sum$	$n_{.1}$ = 30	$n_{.2}$ = 35	$n_{.3}$ = 35	n = 100

1. Schritt: Nullhypothese, Alternativhypothese, Signifikanzniveau

Nullhypothese: H_0: Eignung (A) und Abschluß (B) sind voneinander unabhängig.

Alternativhypothese: H_A: Eignung (A) und Abschluß (B) sind voneinander abhängig.

Signifikanzniveau: $\alpha = 0{,}05$

2. Schritt: Prüfgröße, Testverteilung

Die marginalen relativen Häufigkeiten (Randhäufigkeiten) der Merkmale betragen

$$h(A_i) = \frac{n_{i\cdot}}{n}, \text{ wobei } i=1,...,r$$

und

$$h(B_j) = \frac{n_{\cdot j}}{n}, \text{ wobei } j=1,...,s\,.$$

Bei Unabhängigkeit von A und B gilt:

$$h(A_i \cap B_j) = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n}$$

bzw.

$$n \cdot h(A_i \cap B_j) = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}.$$

Wir wollen die absoluten Häufigkeiten, die sich bei Unabhängigkeit von A und B ergeben, mit $\tilde{n}_{ij}$ bezeichnen:

$$\tilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n} \qquad \text{(absolute Häufigkeiten bei Unabhängigkeit)}$$

Der Unterschied zwischen n_{ij} und $\tilde{n}_{ij}$ wird der Prüfgröße zugrunde gelegt.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$$

Diese Prüfgröße ist χ^2-verteilt mit $\nu = (r-1)(s-1)$ Freiheitsgraden.

3. Schritt: Kritischer Bereich

Der kritische Bereich hat die Form:

$$[\chi_c^2, +\infty).$$

Mit $\alpha = 0{,}05$ und $\nu = (r-1)(s-1) = 1 \cdot 2 = 2$ Freiheitsgraden ergibt sich der kritische Wert zu

$$\chi_c^2 = 5{,}991.$$

4. Schritt: Wert der Prüfgröße

Um die Prüfgröße zu berechnen, muß man die absoluten Häufigkeiten bei Unabhängigkeit $\tilde{n}_{ij}$ ermitteln:

Abschluß (B) / Eignung (A)	Diplom-Sozial-ökonom (B_1)	Diplom-Kauf-mann (B_2)	Diplom-Volkswirt (B_3)	$\sum$
geeignet (A_1)	$\tilde{n}_{11}=12$	$\tilde{n}_{12}=14$	$\tilde{n}_{13}=14$	$n_{1.}=40$
ungeeignet(A_2)	$\tilde{n}_{21}=18$	$\tilde{n}_{22}=21$	$\tilde{n}_{23}=21$	$n_{2.}=60$
$\sum$	$\tilde{n}_{.1}=30$	$\tilde{n}_{.2}=35$	$\tilde{n}_{.3}=35$	$n=100$

Damit erhält man für χ^2:

$$\chi^2=\frac{(14-12)^2}{12}+\frac{(10-14)^2}{14}+\frac{(16-14)^2}{14}+\frac{(16-18)^2}{18}+\frac{(25-21)^2}{21}+\frac{(19-21)^2}{21}$$
$$=2{,}937.$$

5. Schritt: Entscheidung, Interpretation

Da $\chi^2=2{,}937<\chi^2_c=5{,}991$, wird H_0 nicht abgelehnt. Es kann hier nicht nachgewiesen werden, daß die Merkmale Eignung (A) und Abschluß (B) voneinander abhängen.

Beispiel:

ET: Tests: t, F, fits, distribution (NORMALITY)

Mit ET können Tests auf die wichtige Annahme einer Normalverteilung vorgenommen werden (Greene 1993, S. 98 ff.)
Mit den Wochenarbeitszeiten des letzten ET-Beispiels ergibt sich

```
            DATA LISTING (Current sample)
            Observation      HOURS1      HOURS2
                  1        38.000      32.000
                  2        40.000      34.000
                  3        42.000      28.000
                  4        18.000      15.000
                  5        20.000      15.000
                  6        35.000      22.000
                  7        19.000      23.000
                  8        36.000      25.000
                  9        44.000      32.000
                 10        38.000      24.000
                 11        40.000      28.000
                 12        44.000      36.000
                 13        48.000      35.000
                 14        28.000      27.000
                 15        43.000      36.000

Normality Tests
Variable  Chi-squared  P-Value  Kolmogorov-Smirnov  P-Value
HOURS1         1.9576   .37575               .2115      .51337
HOURS2         1.0114   .60308               .1445      .91264
```

Der Chi-Quadrat Test verwendet 'skewness (Schiefe)' und 'kurtosis (Wölbung)'. (Bei einer Normalverteilung: skewness = 0, kurtosis = 3)

$$\chi^2(\nu = k-1) = \chi^2(14, \alpha = 5\%)$$

$$\chi_c^2 = 23{,}685$$

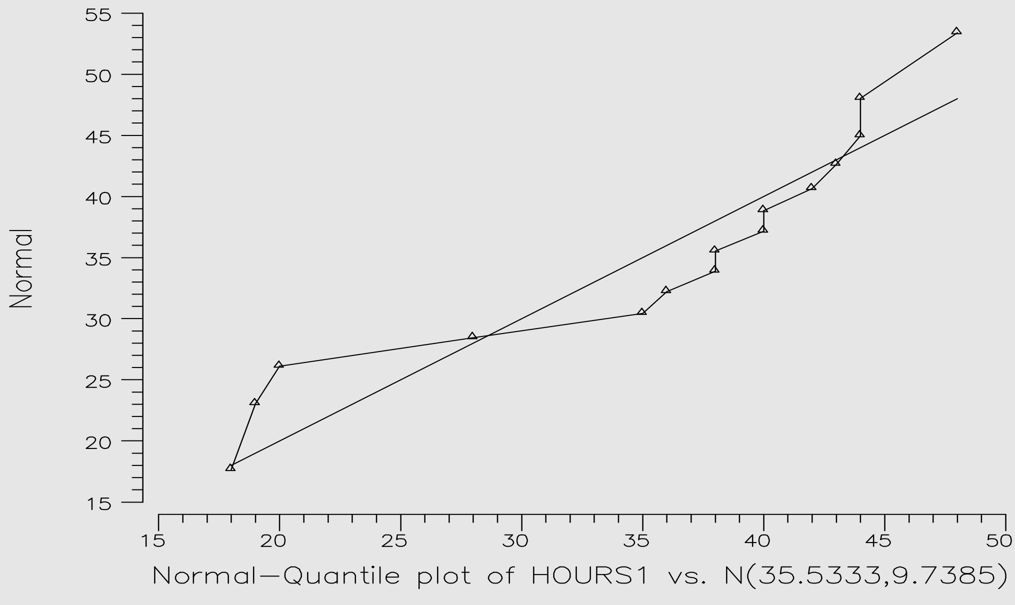
Da $1{,}9576 \leq \chi_c^2 = 23{,}685 \Rightarrow \mathrm{H}_0$ (HOURS ist normalverteilt) nicht ablehnen $\Rightarrow$ HOURS ist normalverteilt.

$$\left.\begin{array}{ll} \text{P - Value} > \alpha \text{ vorgegeben :} & 37{,}575\% > 5\% \Rightarrow \\ \text{Chi - squared} < \chi_c^2 = 23{,}685(5\%): & 1{,}9576 < 23{,}685 \Rightarrow \end{array}\right\} H_0 \text{ beibehalten}$$

Mit relativ kleinen χ^2-Werten (und relativ großen $\alpha = \text{P} - \text{Value - Werten}$) ist die zugrundeliegende Verteilung für HOURS 1 aber auch für HOURS 2 normalverteilt.

Der Kolmogorov-Smirnov Test basiert auf der Differenz zwischen empirischer kumulativer Verteilung und theoretischer Normalverteilung mit gleichem Mittelwert und Varianz. Auch hiermit ist die Normalverteilungshypothese beizubehalten.

Der 'Normal-Quantile Plot' vergleicht die empirische (hier für HOURS1) mit der theoretischen Normalverteilung. Wäre die empirische Verteilung normalverteilt, dann lägen alle Punkte auf der Diagonalen.

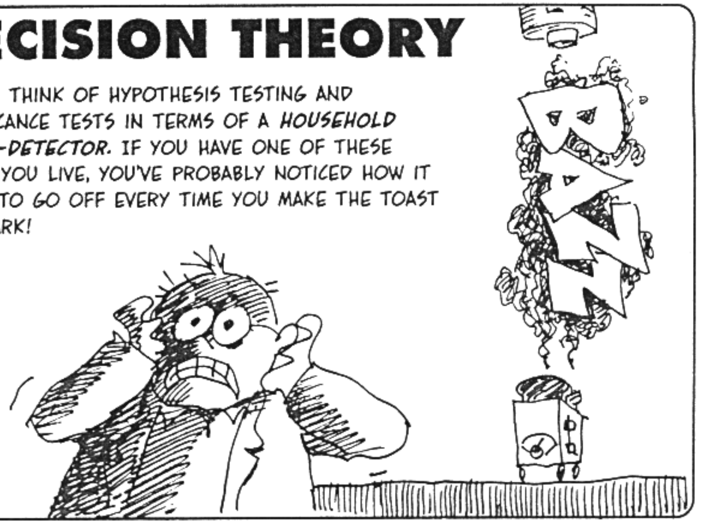


Keyconcepts

Chi-Quadrat-Verteilungstest
Kolmogorov-Smirnov-Veteilungstest
Chi-Quadrat-Unabhängigkeitstest

DECISION THEORY

WE CAN THINK OF HYPOTHESIS TESTING AND SIGNIFICANCE TESTS IN TERMS OF A ***HOUSEHOLD SMOKE-DETECTOR.*** IF YOU HAVE ONE OF THESE WHERE YOU LIVE, YOU'VE PROBABLY NOTICED HOW IT TENDS TO GO OFF EVERY TIME YOU MAKE THE TOAST TOO DARK!



THIS IS WHAT IS CALLED A ***TYPE I ERROR:*** AN ALARM WITHOUT A FIRE. CONVERSELY, A ***TYPE II ERROR*** IS A FIRE WITHOUT AN ALARM. EVERY COOK KNOWS HOW TO AVOID A TYPE I ERROR: JUST ***REMOVE THE BATTERIES.*** UNFORTUNATELY, THIS INCREASES THE INCIDENCE OF TYPE II ERRORS!

SIMILARLY, REDUCING THE CHANCES OF TYPE II ERROR, FOR EXAMPLE BY MAKING THE ALARM HYPERSENSITIVE, CAN INCREASE THE NUMBER OF FALSE ALARMS.

WE CAN SUMMARIZE THIS IN A TWO-BY-TWO ***DECISION TABLE.***

	NO FIRE	FIRE
NO ALARM	NO ERROR	TYPE II
ALARM	TYPE I	NO ERROR

NOW THINK OF THE NULL HYPOTHESIS AS THE CONDITION OF ***NO FIRE,*** WHILE THE ALTERNATE HYPOTHESIS IS THAT A FIRE IS BURNING. THE ALARM CORRESPONDS TO REJECTION OF THE NULL HYPOTHESIS:

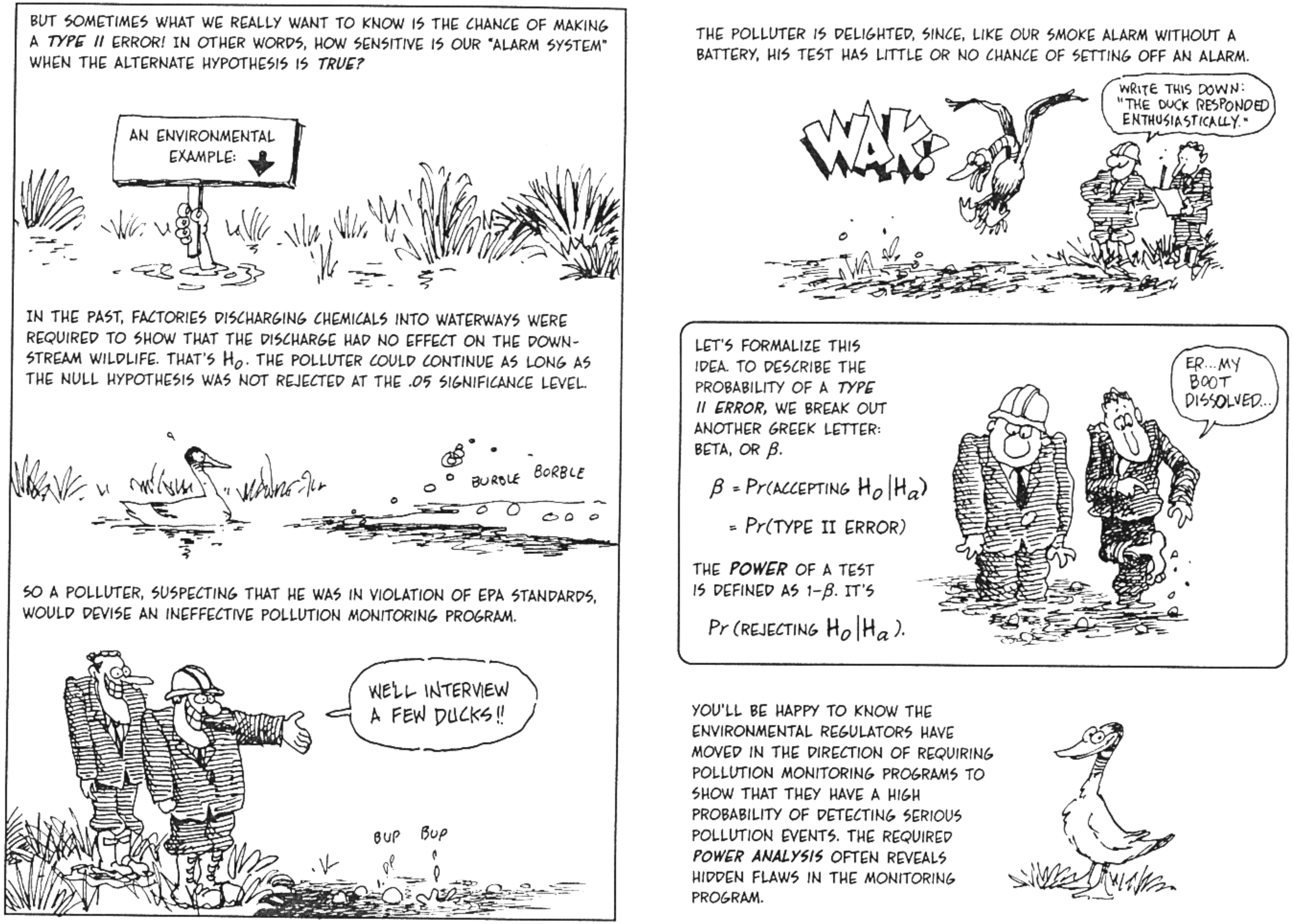
	TRUE STATE	
	H_0	H_a
ACCEPT H_0	NO ERROR	TYPE II
REJECT H_0	TYPE I	NO ERROR

ALL THE SIGNIFICANCE TESTS WE DID EARLIER IN THIS CHAPTER EMPHASIZED THE PROBABILITY OF COMMITTING A TYPE I ERROR—I.E., THE PROBABILITY OF OUR OBSERVATIONS OCCURRING IF H_0 WAS TRUE. WE DEMANDED THAT

$$Pr(\text{REJECTING } H_0 \mid H_0) = Pr(\text{TYPE I ERROR} \mid H_0) < \alpha$$

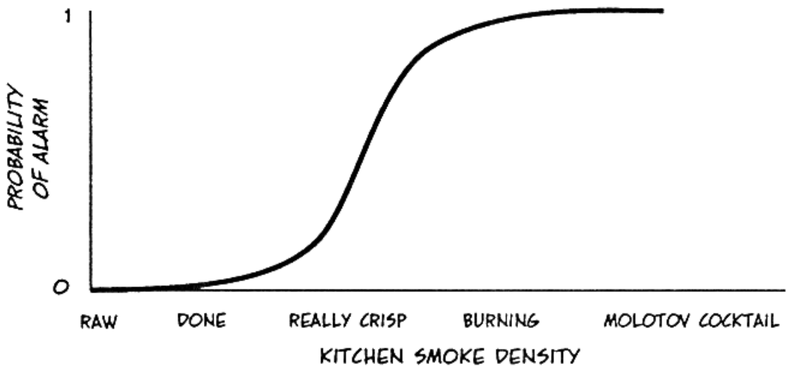
$1-\alpha$ MEASURES OUR CONFIDENCE THAT ANY ALARM BELLS WE HEAR ARE GENUINE. HIGH CONFIDENCE MEANS RARELY SETTING OFF FALSE ALARMS.

Gonick, Smith 1993

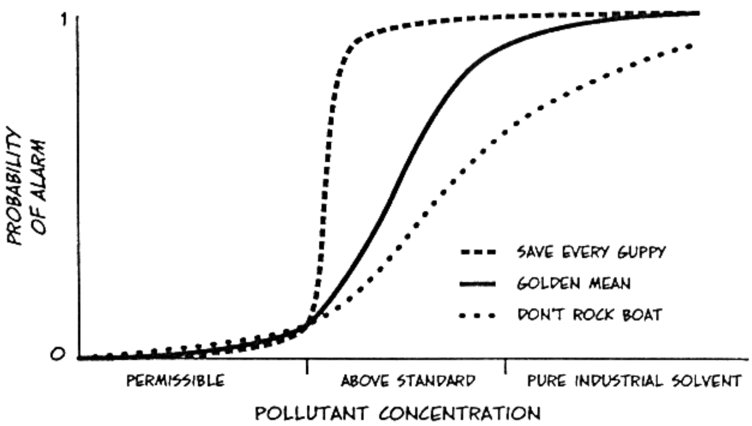


Gonick, Smith 1993

ONE WAY TO VISUALIZE THE EFFECT OF A TEST'S POWER IS BY GRAPHING THE PROBABILITY OF REJECTING H_0 AGAINST THE ACTUAL STATE OF THE SYSTEM. IN THE CASE OF A SMOKE ALARM, THE PROBABILITY CLIMBS TOWARD 1 AS THE SMOKE GETS THICKER.



FOR THE E.P.A. WATER QUALITY EXAMPLE, THE HORIZONTAL AXIS IS THE TRUE CONCENTRATION OF POLLUTANT IN THE WATER.



HERE ARE THE POWER CURVES FOR THREE MONITORING PROGRAMS. THE ***SAVE EVERY LAST GUPPY*** (COSTS \$5 MILLION), THE ***GOLDEN MEAN*** (COSTS \$500,000), AND ***DON'T ROCK THE BOAT*** (ALSO COSTS \$500,000, BUT THEY PUT ON A *GOOD SHOW!*). THE HIGHER THE TEST'S POWER, THE STEEPER THE CURVE.

Gonick, Smith 1993

X Computerprogramme zur Wahrscheinlichkeitsrechnung und induktiven Statistik

Kurze Vorstellung von Computerprogrammen, die in der Praxis für statistische Berechnungen eingesetzt werden

1 Anwendungsmöglichkeiten im Rahmen allgemeiner Programmpakete

Berechnungen der Kombinatorik und der Prüfgrößen können durch allgemeine Berechnungsmodi fast in jedem Programm durchgeführt werden.

Einige Programme liefern auch Tabellenwerte von wichtigen Testverteilungen.

2 SPSS, SAS und BMDP

Wichtige statistische Programmpakete sind SPSS, SAS oder BMDP.

3 ET, LIMDEP, GAUSS, GLIM und Stata

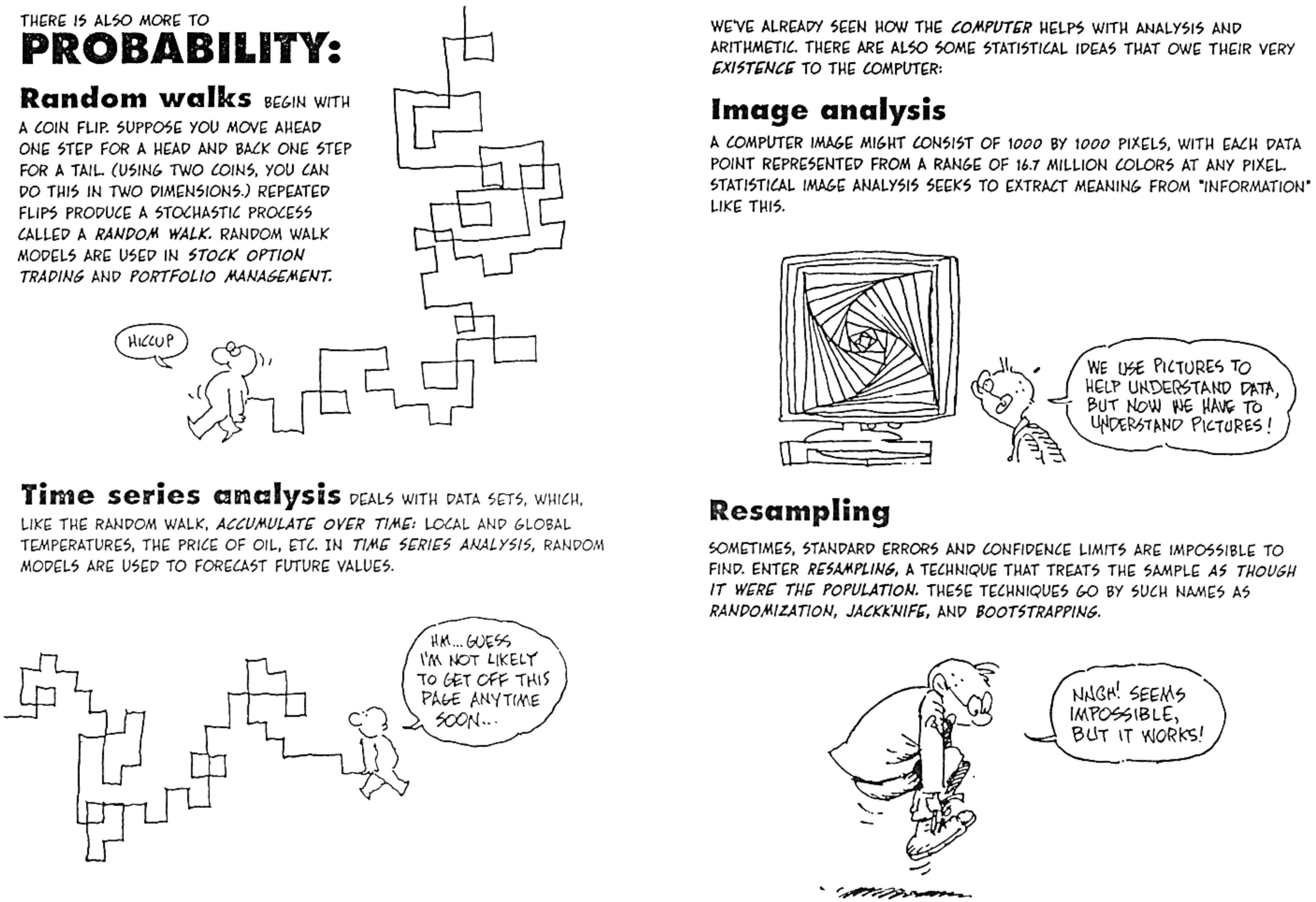
Für Statistik I – Deskription und hier Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik wird bspw. ET (Econometrics Toolkit) als ein einfaches aber schon mächtiges, menuegesteuertes PC-Programmpaket verwendet. Mit der ebenso möglichen Befehlssprache können viele Dinge für alle Beobachtungen alleine über Variablennamen berechnet werden (Tests, Matrizen, Inverse von Matrizen etc.).
Die gleiche Befehlssprache ist Grundlage von LIMDEP, einem Programmpaket speziell für beschränkt abhängige Variablen-Probleme (Limited Dependent Variables, LIMDEP). ET/LIMDEP ist deshalb auch Grundlage meiner weiteren Lehrveranstaltungen im Rahmen einer empirischen Wirtschaftsforschung (z. B. für Diskrete Entscheidungsmodelle – Mikroökonometrie).

GAUSS und GLIM sind Beispiele mächtiger, auf Matrizen aufbauender statistischer Programmpakete mit vielen einzelnen Bausteinen (z. B. MAXLIK bei GAUSS zur iterativen Maximum Likelihood Berechnung).

Stata ist ein mächtiges Programmpaket mit vielen ökonometrisch ausgerichteten Modulen, das bestimmte Probleme – z.B. Regressionsrechnung mit Paneldesign und vielen statistischen Tests – kompakt und komfortabel mit einer Metasprache löst.

Darüber hinaus gibt es noch eine Vielzahl von andere brauchbare statistische Programmpakete.

Enjoy it!



THERE IS ALSO MORE TO

PROBABILITY:

Random walks

BEGIN WITH A COIN FLIP. SUPPOSE YOU MOVE AHEAD ONE STEP FOR A HEAD AND BACK ONE STEP FOR A TAIL. (USING TWO COINS, YOU CAN DO THIS IN TWO DIMENSIONS.) REPEATED FLIPS PRODUCE A STOCHASTIC PROCESS CALLED A *RANDOM WALK.* RANDOM WALK MODELS ARE USED IN *STOCK OPTION TRADING* AND *PORTFOLIO MANAGEMENT.*

Time series analysis

DEALS WITH DATA SETS, WHICH, LIKE THE RANDOM WALK, *ACCUMULATE OVER TIME:* LOCAL AND GLOBAL TEMPERATURES, THE PRICE OF OIL, ETC. IN *TIME SERIES ANALYSIS,* RANDOM MODELS ARE USED TO FORECAST FUTURE VALUES.

WE'VE ALREADY SEEN HOW THE *COMPUTER* HELPS WITH ANALYSIS AND ARITHMETIC. THERE ARE ALSO SOME STATISTICAL IDEAS THAT OWE THEIR VERY *EXISTENCE* TO THE COMPUTER:

Image analysis

A COMPUTER IMAGE *MIGHT CONSIST* OF *1000* BY *1000* PIXELS, WITH EACH DATA POINT REPRESENTED FROM A RANGE OF *16.7 MILLION COLORS* AT ANY PIXEL. STATISTICAL IMAGE ANALYSIS SEEKS TO EXTRACT MEANING FROM "INFORMATION" LIKE THIS.

Resampling

SOMETIMES, STANDARD ERRORS AND CONFIDENCE LIMITS ARE IMPOSSIBLE TO FIND. ENTER *RESAMPLING,* A TECHNIQUE THAT TREATS THE SAMPLE *AS THOUGH IT WERE THE POPULATION.* THESE TECHNIQUES GO BY SUCH NAMES AS *RANDOMIZATION, JACKKNIFE,* AND *BOOTSTRAPPING.*

Gonick, Smith 1993

Anhang I: Formelsammlung

Zusammenfassung der Formeln

I Grundzüge der Wahrscheinlichkeitsrechnung

Laplacesche Wahrscheinlichkeitsdefinition

$$P(A) = \frac{\text{Anzahl der Elementarereignisse in A}}{\text{Anzahl der Elementarereignisse in G}}$$

G = Menge aller Elementarereignisse (Ereignisraum)
A = Ereignis (Teilmenge von G)

Kolmogorov-Axiome

1. Axiom: $P(A) \geq 0$, $P(A) \in R_0^+$
2. Axiom: $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$, falls A_1, A_2,... einander paarweise ausschließende Ereignisse aus demselben Ereignisraum sind.
3. Axiom: P(G) = 1

Bedingte Wahrscheinlichkeit

$$P(B|\ A) = \frac{P(A \cap B)}{P(A)}$$

Stochastische Unabhängigkeit von Ereignissen

$$P(B|\ A) = P(B|\ \overline{A}) = P(B)$$

Sätze der Wahrscheinlichkeitsrechnung

Additionssatz

- für disjunkte Ereignisse:

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

- für nicht-disjunkte Ereignisse:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Multiplikationssatz

- für stochastisch unabhängige Ereignisse:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

- für stochastisch abhängige Ereignisse:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

Satz von der totalen Wahrscheinlichkeit

Die Ereignisse A_i sind disjunkt und bilden eine endliche Zerlegung der Grundgesamtheit. Für ein beliebiges Ereignis ε innerhalb G gilt dann:

$$P(\varepsilon) = \sum_{i=1}^{n} P(\varepsilon \cap A_i) = \sum_{i=1}^{n} P(A_i) \cdot P(\varepsilon | A_i)$$

Satz von Bayes

$$P(A_j | \varepsilon) = \frac{P(A_j) \cdot P(\varepsilon | A_j)}{\sum_{i=1}^{n} P(A_i) \cdot P(\varepsilon | A_i)}$$

II Zufallsvariablen, Wahrscheinlichkeitsverteilungen

Diskrete Zufallsvariable

Wahrscheinlichkeitsfunktion

$$P(X = x_i) = f(x_i)$$

Eigenschaften:

1. $0 \le f(x_i) \le 1$
2. $f(x_i) \ge 0$
3. $\sum_i f(x_i) = 1$

Verteilungsfunktion

$$F(X) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

Eigenschaften:

$0 \le F(x) \le 1$

Für $x_1 \le x_2$ gilt $F(x_1) \le F(x_2)$

Erwartungswert

$$E(X) = \sum_i x_i \cdot f(x_i) = \mu$$

Varianz

$$Var(X) = \sum_i x_i^2 \cdot f(x_i) - \mu^2 = \sigma_x^2$$

Stetige Zufallsvariable

Wahrscheinlichkeitsdichte

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Eigenschaften:

1. $f(x) \ge 0$

2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Verteilungsfunktion

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

Eigenschaften:

1. $0 \le F(x) \le 1$
2. $Für\ x_1 \le x_2\ gilt\ F(x_1) \le F(x_2)$
3. $\lim_{x \to -\infty} F(x) = 0$
4. $\lim_{x \to +\infty} F(x) = 1$

Erwartungswert

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\,dx = \mu$$

Varianz

$$Var(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2 = \sigma_x^2$$

III Diskrete Verteilungen

Gleichverteilung

Parameter n = Stichprobenumfang

Wahrscheinlichkeitsfunktion

$$P(X = x_i) = f(x_i) = \frac{1}{n} \qquad i = 1, \ldots, n$$

Verteilungsfunktion

$$F(x_i)=\begin{cases}0 & \textit{für } x<x_1\\ \dfrac{i}{n} & \textit{für } x_i \le x < x_{i+1} \quad (i=1,\ldots,n-1)\\ 1 & \textit{für } x_n \le x\end{cases}$$

Binomialverteilung

Parameter n = Stichprobenumfang
p = Erfolgswahrscheinlichkeit des Elementarereignisses

Wahrscheinlichkeitsfunktion

$$P(X=x)=f_B(x|\,n,p)=f_B(n-x|\,n,1-p)$$

$$f_B(x|\,n,p)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}$$

Verteilungsfunktion

$$P(X\le x)=F_B(x)=F_B(x|\,n,p)=1-F_B(n-x-1|\,n,1-p)$$

$$F_B(x|n,p)=\sum_{k=0}^{x}\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

Erwartungswert

$$E(X)=n\cdot p$$

Varianz

$$Var(X)=n\cdot p\cdot(1-p)$$

Hypergeometrische Verteilung

Parameter n = Stichprobenumfang
N = Umfang der Grundgesamtheit
M = Zahl der Elemente in der Grundgesamtheit mit einer bestimmten Eigenschaft

Wahrscheinlichkeitsfunktion

$$P(X = x) = f_H(x \mid N, n, M)$$

$$f_H(x \mid N, n, M) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Verteilungsfunktion

$$F_H(x \mid N, n, M) = \sum_{k=0}^{x} \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

Erwartungswert

$$E(X) = n \cdot \frac{M}{N}$$

Varianz

$$Var(X) = n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$$

Poissonverteilung

Parameter μ: Erwartungswert = Varianz

Wahrscheinlichkeitsfunktion

$$P(X = x) = f_P(x \mid \mu)$$

$$f_P(x \mid \mu) = \frac{\mu^x \cdot e^{-\mu}}{x!}$$

Verteilungsfunktion

$$P(X \le x) = F_P(x \mid \mu) = \sum_{k=0}^{x} \frac{\mu^k \cdot e^{-\mu}}{k!}$$

Rekursionsformel

$$f\left((x+1)\,|\,\mu\right) = \frac{\mu}{x+1} \cdot f\left(x\,|\,\mu\right)$$

Erwartungswert

$$E(X) = \mu$$

Varianz

$$Var(X) = \mu$$

Geometrische Verteilung

Parameter p

Wahrscheinlichkeitsfunktion

$$f_g\left(x\,|\,p\right) = p\left(1-p\right)^{x-1} \qquad x = 1, 2, \ldots$$

Verteilungsfunktion

$$F_g\left(x\,|\,p\right) = \begin{cases} 0 & \text{für } x < 1 \\ 1-\left(1-p\right)^m & \text{für } m \le x < m+1; \quad m = 1, 2, \ldots \end{cases}$$

Erwartungswert

$$E(X) = \frac{1}{p}$$

Varianz

$$Var(X) = \frac{1-p}{p^2}$$

Multinomialverteilung

Parameter n = Stichprobenumfang
p_i = Erfolgswahrscheinlichkeiten der Elementarereignisse

Wahrscheinlichkeitsfunktion

$$P\left(X_1 = x_1, ..., X_k = x_k\right) = f_M\left(x_1, x_2, ..., x_k \middle| n, p_1, ..., p_k\right)$$

$$f_M\left(x_1, x_2, ..., x_k \mid n, p_1, ..., p_k\right) = \frac{n!}{x_1!\, x_2! ... x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot ... \cdot p_k^{x_k}$$

mit $\sum_{i=1}^{k} x_i = n$ und $\sum_{i=1}^{k} p_i = 1.$

Erwartungswert

$$E\left(X_i\right) = n \cdot p_i$$

Varianz

$$Var\left(X_i\right) = n \cdot p_i \cdot \left(1 - p_i\right)$$

Allgemeine hypergeometrische Verteilung

Parameter n = Stichprobenumfang
N_i = Zahl der Elemente in der Grundgesamtheit mit einer bestimmten, für jedes i verschiedenen Eigenschaft

Wahrscheinlichkeitsfunktion

$$P\left(X_i = x_1, ..., X_k = x_k\right) = f_{AH}\left(x_1, x_2, ..., x_k \middle| n, N_1, N_2, ..., N_k\right)$$

$$f_{AH}\left(x_1, x_2, ..., x_k \mid n, N_1, ..., N_k\right) = \frac{\binom{N_1}{x_1} \cdot \binom{N_2}{x_2} \cdot ... \cdot \binom{N_k}{x_k}}{\binom{N}{n}}$$

mit $\sum_{i=1}^{k} x_i = n$ und $\sum_{i=1}^{k} N_i = N$.

Erwartungswert

$$E\left(X_i\right) = n \cdot \frac{N_i}{N}$$

Varianz

$$Var(X_i) = n \cdot \frac{N_i}{N} \cdot \left(1 - \frac{N_i}{N}\right) \cdot \frac{N-n}{N-1}$$

IV Stetige Verteilungen

Gleichverteilung

Parameter a, b → Grenzen des Intervalls

Dichtefunktion

$$f_g(x|\,a,b) = \begin{cases} \dfrac{1}{b-a} & \text{für } a \le x \le b \\ 0 & \text{sonst} \end{cases}$$

Verteilungsfunktion

$$F_g(x|\,a,b) = \begin{cases} 0 & \text{für x<a} \\ \dfrac{x-a}{b-a} & \text{für } \text{a} \le \text{x} \le \text{b} \\ 1 & \text{für x>b} \end{cases}$$

Erwartungswert

$$E(X) = \frac{a+b}{2}$$

Varianz

$$Var(X) = \frac{(b-a)^2}{12}$$

Exponentialverteilung

Dichtefunktion

$$f_E(x|\,\lambda) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{für } x \ge 0 \text{ mit } \lambda > 0 \\ 0 & \text{sonst} \end{cases}$$

Verteilungsfunktion

$$F_E(x|\lambda) = \begin{cases} 0 & \text{\textit{für} } x < 0 \\ 1 - e^{-\lambda \cdot x} & \text{\textit{für} } x \geq 0 \end{cases}$$

Erwartungswert

$$E(X) = \frac{1}{\lambda}$$

Varianz

$$Var(X) = \frac{1}{\lambda^2}$$

Gammaverteilung

Parameter r, λ

Dichtefunktion

$$f_G = \frac{\lambda^r}{\Gamma(r)} \cdot e^{-\lambda x} \cdot x^{r-1} \quad , \quad x \geq 0$$

$$\textit{wobei } \Gamma(r) = \int_0^\infty t^r\, e^{-t}\, dt \ , \ \Gamma(r) = (r-1)! \ , \ \textit{falls } r \ \textit{ganzzahlig}$$

Erwartungswert

$$E(X) = \frac{r}{\lambda}$$

Varianz

$$Var(X) = \frac{r}{\lambda^2}$$

Normalverteilung

Parameter μ = Erwartungswert
σ^2 = Varianz = Streuung um den Erwartungswert

Dichtefunktion

$$f_N\left(x \middle| \mu, \sigma^2\right) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$$

für $-\infty < x < +\infty$

$-\infty < \mu < +\infty$

$0 < \sigma^2 < +\infty$

Verteilungsfunktion

$$F_N\left(x \middle| \mu, \sigma^2\right) = \int_{-\infty}^{x} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Erwartungswert

$$E(X) = \mu$$

Varianz

$$Var(X) = \sigma^2$$

Standardnormalverteilung

$$N\left(\mu, \sigma^2\right) = N(0,1)$$

= Normalverteilung mit den Parametern μ=0 und $\sigma^2 = 1$

Standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma}$$

Dichtefunktion

$$\phi(z) = f_N(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\cdot z^2}$$

Verteilungsfunktion

$$\Phi(z) = F_N(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2}\cdot t^2} dt$$

Chi-Quadratverteilung

Parameter ν

Dichtefunktion

$$f_{\chi^2}\left(z \middle| \nu\right) = C\left(\nu\right) \cdot z^{\frac{\nu}{2}-1} \cdot e^{-\frac{z}{2}} \qquad z > 0$$

$$\text{mit } C\left(\nu\right) = \left[2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)\right]^{-1}$$

$$\Gamma\left(x\right) = \int_0^\infty e^{-t} \cdot t^{x-1} dt$$

Erwartungswert

$$E\left(\chi^2\right) = \nu$$

Varianz

$$Var\left(\chi^2\right) = 2\nu$$

Studentverteilung (t-Verteilung)

Parameter ν

Dichtefunktion

$$\mathrm{f_t}(\mathrm{z} \mid \nu) = C(\nu) \cdot \frac{1}{\left(1 + \frac{z^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad -\infty \leq z \leq \infty$$

$$mit\ C(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu \cdot \pi} \cdot \Gamma\left(\frac{\nu}{2}\right)}$$

Erwartungswert

$$E\left(T\right) = 0$$

Varianz

$$Var(T) = \frac{\nu}{\nu - 2} \qquad \text{für } \nu > 2$$

F-Verteilung

Parameter ν_1, ν_2

Dichtefunktion

$$f_F(z|\nu_1, \nu_2) = C(\nu_1, \nu_2) \cdot \frac{z^{\frac{\nu_1}{2}-1}}{(\nu_2 + \nu_1 z)^{\frac{\nu_1+\nu_2}{2}}}, \qquad z > 0$$

$$\text{mit } C(\nu_1, \nu_2) = \frac{\nu_1^{(\nu_1/2)} \cdot \nu_2^{(\nu_2/2)}}{\dfrac{\Gamma\left(\dfrac{\nu_1}{2}\right) \cdot \Gamma\left(\dfrac{\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2} + \dfrac{\nu_2}{2}\right)}}$$

Erwartungswert

$$E(F) = \frac{\nu_2}{\nu_2 - 2} \qquad \text{für } \nu_2 \geq 2$$

Varianz

$$Var(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \qquad \text{für } \nu_2 > 4$$

Übergänge zwischen den Verteilungen

Normalverteilung

$$n \cdot p \cdot (1 - p) \geq 9$$

$$Z = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}$$

$$\mu \geq 9$$

$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

$$n \geq 100$$
$$p \leq 0{,}05$$

Binomialverteilung **Poissonverteilung**

$$\frac{n}{N} \leq 0{,}05$$

$$p = \frac{M}{N}$$

$$n \cdot p \cdot (1 - p) \geq 9$$

$$\frac{n}{N} \leq 0{,}05$$

$$Z = \frac{X - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p) \cdot \frac{N - n}{N - 1}}}$$

Hypergeometrische Verteilung

V Induktive Statistik Stichprobenfunktionen und Testverteilungen

Arithmetisches Mittel der Stichprobe

Stichprobenfunktion

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i)$$

Erwartungswert

$$E\left(\overline{X}\right) = \mu$$

Varianz

$$Var\left(\overline{X}\right) = \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$$

Testverteilung

$$Z = \frac{\overline{X} - \mu}{\sigma} \cdot \sqrt{n}$$

ist standardnormalverteilt, falls X_i normalverteilt ist;

ist approximativ standardnormalverteilt, falls die X_i unabhängig identisch verteilt sind und n groß ist (Faustregel: $n \geq 100$).

Varianz der Stichprobe

Stichprobenfunktion

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

Testverteilung

$$\chi^2 = \frac{n \cdot S^2}{\sigma^2}, wobei\ S^2 = \frac{1}{n} \sum_i \left(X_i - \overline{X}\right)^2$$

χ^2 ist χ^2-verteilt mit $\nu = n - 1$ Freiheitsgraden.

Standardisierte Zufallsvariable bei unbekannter Varianz der Grundgesamtheit

Stichprobenfunktion

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n-1}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\chi^2}{(n-1)}}}$$

Testverteilung

T ist t-verteilt mit $\nu = n-1$ Freiheitsgraden

Quotienten zweier Varianzen

Stichprobenfunktion

für $\sigma_1^2 = \sigma_2^2$ gilt $F = \frac{n_1 \cdot S_1^2 \cdot (n_2-1)}{n_2 \cdot S_2^2 \cdot (n_1-1)} = \frac{\chi_1^2 \cdot (n_2-1)}{\chi_2^2 \cdot (n_1-1)}$.

Testverteilung

F ist F-verteilt mit $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ Freiheitsgraden.

VI Punktschätzung

Eigenschaften von Schätzfunktionen

- **Erwartungstreue**: $E(\hat{\theta}) = \theta$

  Verzerrung (bias): $E(\hat{\theta}) \neq \theta$

  Asymptotische Erwartungstreue, wenn gilt: $\lim_{n\to\infty} E(\hat{\theta}^{(n)}) = \theta$

- **Konsistenz**: $\lim_{n\to\infty} prob(|\hat{\theta}^n - \theta| \geq \delta) = 0$
  - Eine Schätzfunktion $\hat{\theta}$ ist konsistent, falls sie
    1. asymptotisch erwartungstreu ist und
    2. ihre Varianz $Var(\hat{\theta})$ für $n \to \infty$ gegen Null strebt.

- **Effizienz**:
  - relative Effizienz:

    $Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$, wobei $\hat{\theta}_1, \hat{\theta}_2$ erwartungstreue Schätzfunktionen für θ sind.

  - absolute Effizienz:

    Die $Var\left(\hat{\theta}\right)$ ist minimal im Vergleich zu jeder anderen erwartungstreuen Schätzfunktion.

Schätzmethoden

Methode der Momente

Unbekannte Parameter der Grundgesamtheit werden den entsprechenden Parametern der Stichprobe gleichgesetzt.

Methode der kleinsten Quadrate

Unbekannte Parameter der Grundgesamtheit $\beta_k \; (k=0,...,K)$ aus

$$y_i = \underbrace{\underbrace{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_K x_{Ki}}_{\mu_i}}_{\text{systematischer Einfluß } f(x_i)} + \underbrace{\underbrace{\varepsilon_i}_{\varepsilon_i}}_{\text{zufälliger Einfluß}}$$

werden über $b^{OLS} = \left(X'X\right)^{-1} X'y$ geschätzt.

Maximum-Likelihood-Methode

Likelihoodfunktion:

$$L(\theta) = f\left(x_1 \mid \theta\right) \cdot f\left(x_2 \mid \theta\right) \cdot \ldots \cdot f\left(x_n \mid \theta\right)$$

$$= \prod_{i=1}^{n} f\left(x_i \mid \theta\right) \qquad \textit{bzw.}$$

$$\ln\; L(\theta) = \sum_{i=1}^{n} \ln\; f\left(x_i \mid \theta\right)$$

mit $f\left(x_i \mid \theta\right) =$ Dichtefunktion der Zufallsvariable X_i

Den Schätzwert $\hat{\theta}$ erhält man über die Maximierung von $L(\theta)$ bzw. $\ln\; L(\theta)$:

$$\frac{\partial \mathrm{L}}{\partial \theta} = 0 \text{ bzw. } \frac{\partial \ln \mathrm{L}}{\partial \ln \theta} = 0.$$

VII Intervallschätzung

Konfidenzintervall für μ

- **bei bekannter Varianz σ^2 der Grundgesamtheit (GG normalverteilt)**

$$P\left(\bar{X} - z_{1-\alpha/2} \cdot \sigma_{\bar{X}} \le \mu \le \bar{X} + z_{1-\alpha/2} \cdot \sigma_{\bar{X}}\right) = 1-\alpha$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$ (Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \ge 0,05$)

- **bei unbekannter Varianz σ^2 der Grundgesamtheit (GG normalverteilt)**

$$P\left(\bar{X} - t \cdot \hat{\sigma}_{\bar{X}} \le \mu \le \bar{X} + t \cdot \hat{\sigma}_{\bar{X}}\right) = 1-\alpha, \qquad \textit{wobei } t = t_{1-\alpha/2,n-1}$$

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n-1}}$$

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N}}$$ (Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \ge 0,05$)

Wenn die Anzahl der Freiheitsgrade $\nu > 30$ ist, kann die Studentverteilung durch die Normalverteilung approximiert werden.

Konfidenzintervall für die Varianz σ^2 (GG normalverteilt)

$$P\left(\frac{n \cdot S^2}{\chi^2_{1-\alpha/2,n-1}} \le \sigma^2 \le \frac{n \cdot S^2}{\chi^2_{\alpha/2,n-1}}\right) = 1-\alpha$$

Konfidenzintervall für den Anteilswert p

$$P\left(\hat{p} - z_{1-\alpha/2} \cdot \hat{\sigma}_{\hat{p}} \le p \le \hat{p} + z_{1-\alpha/2} \cdot \hat{\sigma}_{\hat{p}}\right) = 1-\alpha$$

mit $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$ Modell mit Zurücklegen

mit $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n} \cdot \frac{N-n}{N-1}}$ Modell ohne Zurücklegen und $\frac{n}{N} \ge 0,05$.

Bestimmung des notwendigen Stichprobenumfangs

- **bei Schätzung für μ**

  $\Delta\mu = z \cdot \sigma_{\overline{X}}$ = absoluter Fehler

  $$n = \frac{z^2 \cdot \sigma^2}{(\Delta\mu)^2}$$

- **bei Schätzung für Anteilswert p**

  $\Delta p = z \cdot \sigma_{\hat{p}} = z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$ = absoluter Fehler — Modell mit Zurücklegen

  $$n = \frac{z^2 \cdot p \cdot (1-p)}{(\Delta p)^2}$$ Modell mit Zurücklegen

  $\Delta p = z \cdot \sqrt{\frac{p \cdot (p-1)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ = *absoluter Fehler* — Modell ohne Zurücklegen und $\frac{n}{N} \geq 0{,}05$

  $$n = \frac{z^2 \cdot N \cdot p \cdot (1-p)}{(\Delta p)^2 \cdot (N-1) + z^2 \cdot p \cdot (1-p)}$$ Modell ohne Zurücklegen und $\frac{n}{N} \geq 0{,}05$

  $\hat{p}$ ist Schätzwert für p.

  Wenn keine Informationen über $\hat{p}$ vorliegen: $\hat{p} = 0{,}5$.

Konfidenzintervall für die Differenz zweier arithmetischer Mittel

Voraussetzung:
- Unabhängigkeit der Stichproben
- genügend große Stichprobenumfänge (Faustregel: $n_1 > 30, n_2 > 30$)

- **σ_1^2 und σ_2^2 bekannt**

  $$P\left[\left(\overline{X}_1 - \overline{X}_2\right) - z \cdot \sigma_D \leq \mu_1 - \mu_2 \leq \left(\overline{X}_1 - \overline{X}_2\right) + z \cdot \sigma_D\right] = 1 - \alpha$$

  wobei $z = z_{1-\alpha/2}$

  $$\sigma_D{}^2 = \frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}$$

- **σ_1^2 und σ_2^2 unbekannt, aber $\sigma_1^2 = \sigma_2^2$**

$$P\left[\left(\bar{X}_1 - \bar{X}_2\right) - t \cdot \hat{\sigma}_D \le \mu_1 - \mu_2 \le \left(\bar{X}_1 - \bar{X}_2\right) + t \cdot \hat{\sigma}_D\right] = 1 - \alpha$$

wobei $t = t_{1-\alpha/2, n_1+n_2-2}$

$$\hat{\sigma}^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2} \cdot S^2$$

$$S^2 = \frac{1}{n_1 + n_2} \cdot \left(n_1 \cdot S_1^{\,2} + n_2 \cdot S_2^{\,2}\right)$$

$$\hat{\sigma}_D^{\,2} = \frac{n_1 \cdot S_1^{\,2} + n_2 \cdot S_2^{\,2}}{n_1 + n_2 - 2} \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)$$

Konfidenzintervall für die Differenz zweier Anteilswerte

Bedingung:

$$n_1 \cdot \hat{p}_1 \cdot (1 - \hat{p}_1) \ge 9,\ n_2 \cdot \hat{p}_2 \cdot (1 - \hat{p}_2) \ge 9$$

$$P[(\hat{p}_1 - \hat{p}_2) - z \cdot \hat{\sigma}_D \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z \cdot \hat{\sigma}_D] = 1 - \alpha$$

wobei

$$\hat{\sigma}_D^2 = \frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}$$

$$z = z_{1-\alpha/2}$$

VIII Parametertests

Prüfgrößen und Testverteilungen

Grundlagen

1. Schritt: Parametermenge, Nullhypothese, Alternativhypothese, Signifikanzniveau
2. Schritt: Prüfgröße, Testverteilung
3. Schritt: Kritischer Bereich
4. Schritt: Wert der Prüfgröße
5. Schritt: Entscheidung, Interpretation

Einstichprobentests für den Anteilswert

Zweiseitige Fragestellung

Nullhypothese: $H_0 : p = p_0$

Alternativhypothese: $H_A : p \neq p_0$

Prüfgröße: $$Z = \frac{\hat{p} - p_0}{\sqrt{Var(\hat{p}|p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} ,$$

wobei $\hat{p} = \frac{X}{n}$ und $X \sim B(n, p)$

Testverteilung: $Z \sim N(0, 1)$ für $n \cdot p \cdot (1-p) > 9$

Kritischer Bereich: $(-\infty; z_u] \cup [z_o; +\infty)$

$z_u = z_{\alpha/2}$, $z_o = z_{1-\alpha/2}$

Einseitige Fragestellung

Nullhypothese: $H_0 : p \leq p_0$

Alternativhypothese: $H_A : p > p_0$

Prüfgröße: $$Z = \frac{\hat{p} - p_0}{\sqrt{Var(\hat{p}|p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} ,$$

wobei $\hat{p} = \frac{X}{n}$ und $X \sim B(n, p)$

Testverteilung: $Z \sim N(0, 1)$ für $n \cdot p \cdot (1-p) > 9$

Kritischer Bereich: $[z_c; +\infty$

$z_c = z_{1-\alpha}$

Einstichprobentests für das arithmetische Mittel bei normalverteilter Grundgesamtheit

Einstichprobentest für μ bei bekannter Varianz σ^2 der GG

Zweiseitige Fragestellung

Nullhypothese: $H_0 : \mu = \mu_0$

Alternativhypothese: $H_A : \mu \neq \mu_0$

Prüfgröße: $$Z = \frac{\bar{X} - \mu_0}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} ,$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$ (Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0,05$)

Testverteilung: $Z \sim N(0, 1)$

Kritischer Bereich: $(-\infty; z_u] \cup [z_o; +\infty)$

$z_u = z_{\alpha/2}$, $z_o = z_{1-\alpha/2}$

Einstichprobentest für μ bei unbekannter Varianz σ^2 der GG

Zweiseitige Fragestellung

Nullhypothese: $H_0 : \mu = \mu_0$

Alternativhypothese: $H_A : \mu \neq \mu_0$

Prüfgröße:

$$T = \frac{\overline{X} - \mu_0}{\sqrt{V\hat{a}r(\overline{X})}} = \frac{\overline{X} - \mu_0}{\hat{\sigma}_{\overline{X}}},$$

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n-1}}$$

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N}}$$ (Korrektur bei einer Stichprobe ohne Zurücklegen und $\frac{n}{N} \geq 0,05$)

Testverteilung: T ~ t - verteilt mit n – 1 Freiheitsgraden

Kritischer Bereich: $(-\infty; t_u] \cup [t_o; +\infty)$

$t_u = t_{\alpha/2;n-1}$, $t_o = t_{1-\alpha/2;n-1}$

Einstichprobentests für die Varianz bei normalverteilter Grundgesamtheit

Zweiseitige Fragestellung

Nullhypothese: $H_0 : \sigma = \sigma_0$

Alternativhypothese: $H_A : \sigma \neq \sigma_0$

Prüfgröße:

$$\chi_0^2 = \frac{n \cdot S^2}{\sigma_0^2}$$

Testverteilung: $\chi_0^2 \sim \chi^2$- verteilt mit n – 1 Freiheitsgraden

Kritischer Bereich: $\left[0;\chi_u^{\;2}\right]\cup\left[\chi_o^{\;2};+\infty\right)$

$\chi_u^{\;2}=\chi^2_{\alpha/2}\,(n-1)$, $\chi_o^{\;2}=\chi^2_{1-\alpha/2}\left(n-1\right)$

Einseitige Fragestellung

Nullhypothese: $H_0:\sigma\leq\sigma_0$

Alternativhypothese: $H_A:\sigma>\sigma_0$

Prüfgröße: $\chi_0^{\;2}=\dfrac{n\cdot S^2}{\sigma_0^{\;2}}$

Testverteilung: $\chi_0^{\;2}\sim\chi^2$- verteilt mit n-1 Freiheitsgraden

Kritischer Bereich: $\left[\chi_c^{\;2};+\infty\right)$

$\chi_c^{\;2}=\chi^2_{1-\alpha}\left(n-1\right)$

Zweistichprobentests für die Differenz zweier arithmetischer Mittel

Voraussetzungen:

- Beide Stichproben sind unabhängig voneinander.
- Beide Stichproben stammen aus normalverteilten Grundgesamtheiten.

- **Die Varianzen σ_1^2 und σ_2^2 sind bekannt**

Fall 1: $\sigma_1^{\;2}=\sigma_2^{\;2}$

Nullhypothese: $H_0:\mu_1-\mu_2=0$

Alternativhypothese: $H_A:\mu_1-\mu_2\neq0$

Prüfgröße: $Z=\dfrac{\left(\bar{X}_1-\bar{X}_2\right)}{\sigma}\cdot\sqrt{\dfrac{n_1\cdot n_2}{n_1+n_2}}$

Testverteilung: Z ~ N (0 , 1)

Kritischer Bereich: $\left(-\infty;z_u\right]\cup\left[z_o;+\infty\right)$

$z_u=z_{\alpha/2}$, $z_o=z_{1-\alpha/2}$

Fall 2: $\sigma_1^{\;2}\neq\sigma_2^{\;2}$

Nullhypothese: $H_0:\mu_1-\mu_2=0$

Alternativhypothese: $H_A:\mu_1-\mu_2\neq0$

Prüfgröße: $Z=\dfrac{\left(\bar{X}_1-\bar{X}_2\right)}{\sqrt{n_2\cdot\sigma_1^{\;2}+n_1\cdot\sigma_2^{\;2}}}\cdot\sqrt{n_1\cdot n_2}$

Testverteilung: Z ~ N (0 , 1)

Kritischer Bereich: $\left(-\infty; z_u\right] \cup \left[z_o; +\infty\right)$

$z_u = z_{\alpha/2}$, $z_o = z_{1-\alpha/2}$

- **Die Varianzen σ_1^2 und σ_2^2 sind unbekannt**

Fall 1: $\sigma_1^2 = \sigma_2^2$

Nullhypothese: $H_0 : \mu_1 - \mu_2 = 0$

Alternativhypothese: $H_A : \mu_1 - \mu_2 \neq 0$

Prüfgröße:
$$T = \frac{\bar{X}_1 - \bar{X}_2}{\left(n_1 + n_2\right) \cdot \frac{S}{\sqrt{n_1 + n_2 - 2}}} \cdot \sqrt{n_1 \cdot n_2}$$

wobei $S^2 = \frac{1}{n_1 + n_2} \cdot \left(n_1 \cdot S_1^2 + n_2 \cdot S_2^2\right)$ bzw.

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{n_2 \cdot \hat{\sigma}_1^2 + n_1 \cdot \hat{\sigma}_2^2}} \cdot \sqrt{n_1 \cdot n_2} \quad \text{mit } \hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} S_1^2 \quad \hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} S_2^2$$

Testverteilung: T ~ t-verteilt mit $\nu = n_1 + n_2 - 2$ Freiheitsgraden

Kritischer Bereich: $\left(-\infty; t_u\right] \cup \left[t_o; +\infty\right)$

$t_u = t_{\alpha/2,\nu}$, $t_o = t_{1-\alpha/2,\nu}$

Fall 2: $\sigma_1^2 \neq \sigma_2^2$

Nullhypothese: $H_0 : \mu_1 - \mu_2 = 0$

Alternativhypothese: $H_A : \mu_1 - \mu_2 \neq 0$

Prüfgröße:
$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{n_2 \cdot \hat{\sigma}_1^2 + n_1 \cdot \hat{\sigma}_2^2}} \cdot \sqrt{n_1 \cdot n_2}$$

wobei $\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} \cdot S_1^2$

$\hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} \cdot S_2^2$

Testverteilung: T ist annähernd t-verteilt mit ν Freiheitsgraden:

$$\nu^* = \frac{1}{\frac{w^2}{n_1 - 1} + \frac{(1-w)^2}{n_2 - 1}} \quad \textit{mit}$$

$$0 < w = \frac{\hat{\sigma}_1^2 \cdot n_2}{\hat{\sigma}_1^2 \cdot n_2 + \hat{\sigma}_2^2 \cdot n_1} < 1$$

$$\nu = \left\lfloor \nu^* \right\rfloor$$

Kritischer Bereich: $\left(-\infty; t_u\right] \cup \left[t_o; +\infty\right)$

$t_u = t_{\alpha/2,\nu}$, $t_o = t_{1-\alpha/2,\nu}$

Zweistichprobentests für den Quotienten zweier Varianzen

Voraussetzungen:

- Beide Stichproben sind unabhängig voneinander.
- Beide Stichproben stammen aus normalverteilten Grundgesamtheiten.

Zweiseitige Fragestellung

Nullhypothese: $H_0: \sigma_1^2 = \sigma_2^2$

Alternativhypothese: $H_A: \sigma_1^2 \neq \sigma_2^2$

Prüfgröße: $F = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \dfrac{n_1 \cdot S_1^2 / (n_1 - 1)}{n_2 \cdot S_2^2 / (n_2 - 1)}$

Testverteilung: F ist F-verteilt mit $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ Freiheitsgraden

Kritischer Bereich: $[0; F_u] \cup [F_o; +\infty$

$F_o = F_{1-\alpha/2, \nu_1, \nu_2}$, $F_u = F_{\alpha/2, \nu_1, \nu_2} = \dfrac{1}{F_{1-\alpha/2, \nu_2, \nu_1}}$

Einseitige Fragestellung

Nullhypothese: $H_0: \dfrac{\sigma_1^2}{\sigma_2^2} \leq 1$

Alternativhypothese: $H_A: \dfrac{\sigma_1^2}{\sigma_2^2} > 1$

Prüfgröße: $F = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \dfrac{n_1 \cdot S_1^2 / (n_1 - 1)}{n_2 \cdot S_2^2 / (n_2 - 1)}$

Testverteilung: F ist F-verteilt mit $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ Freiheitsgraden

Kritischer Bereich: $[F_c; +\infty$

$F_c = F_{1-\alpha, \nu_1, \nu_2}$

Zweistichprobentests für die Differenz zweier Anteilswerte

Voraussetzungen:

- Beide Stichproben sind voneinander unabhängig.
- n_1 und n_2 sind so groß, daß die Anteilswerte als normalverteilt angesehen werden können.

Nullhypothese: $H_0: p_1 - p_2 = 0$

Alternativhypothese: $H_A: p_1 - p_2 \neq 0$

Prüfgröße: $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$

wobei $\hat{p} = \frac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$

Testverteilung: Z ist annähernd N (0,1)-verteilt

Kritischer Bereich: $(-\infty; z_u] \cup [z_o; +\infty)$

$z_u = z_{\alpha/2}$, $z_o = z_{1-\alpha/2}$

Tests im klassischen linearen Regressionsmodell

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_k \cdot x_{ik} + \ldots + \beta_K \cdot x_{iK} + \varepsilon_i \qquad (i = 1, \ldots, n)$$

Test der Gesamterklärungsgüte (F-Test)

Nullhypothese: $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = \ldots = \beta_K = 0$

Alternativhypothese: $H_A : \beta_k \neq 0 \quad (k = 1, \ldots, K)$

Prüfgröße: $F = \frac{R^2}{1 - R^2} \cdot \frac{n - K - 1}{K}$

Testverteilung: F ist F-verteilt mit $\nu_1 = K$ und $\nu_2 = n - K - 1$ Freiheitsgraden

Kritischer Bereich: $[F_c; +\infty$

$F_c = F_{1-\alpha, K, n-K-1}$

Signifikanztest für die einzelnen MKQ/OLS-Koeffizienten b_k (t-Test)

Nullhypothese: $H_0 : \beta_k = 0$

Alternativhypothese: $H_A : \beta_k \neq 0$

Prüfgröße: $t = \frac{b_k - (\beta_k = 0)}{s_{b_k}}$

wobei $s_{b_k} = \sqrt{V\hat{a}r(b_k)} = \frac{e'e}{n - K - 1} \cdot (X'X)^{-1}_{kk}$ *und*

$e = y - \hat{y}$

Testverteilung: t ist t-verteilt mit $\nu = n - K - 1$ Freiheitsgraden

Kritischer Bereich: $]-\infty; -t_u] \cup [t_o; +\infty[$, $t_u = -t_c = t_{\alpha/2, n-K-1}$, $t_o = t_c = t_{1-\alpha/2, n-K-1}$

IX Verteilungstests

Chi-Quadrat-Verteilungstest

Einfache Hypothese

Nullhypothese: H_0: Das Merkmal folgt einer bestimmten, genau festgelegten Verteilung.

Alternativhypothese: H_A: Das Merkmal folgt dieser bestimmten, genau festgelegten Verteilung nicht.

Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

Testverteilung: χ^2 ist χ^2-verteilt mit $\nu = k - 1$ Freiheitsgraden.
(k ist die Anzahl der Merkmalsausprägungen)
Faustregel für die Anwendung: $n \cdot p_i \geq 5$ für $i = 1, \ldots, k$

Kritischer Bereich:

$$\left[\chi_c^2; +\infty\right)$$

$$\chi_c^2 = \chi_{1-\alpha}^2 (k-1)$$

Zusammengesetzte Hypothese

Nullhypothese: H_0: Das Merkmal folgt einer bestimmten Verteilung.

Alternativhypothese: H_A: Das Merkmal folgt dieser bestimmten Verteilung nicht.

Anmerkung: Parameter der Verteilung müssen geschätzt werden.

Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

Testverteilung: χ^2 ist χ^2-verteilt mit $\nu = k - 1 - m$ Freiheitsgraden.
(k ist die Anzahl der Merkmalsausprägungen,
m ist die Anzahl der geschätzten Parameter der Verteilung)
Faustregel für die Anwendung: $n \cdot p_i \geq 5$ für $i = 1, \ldots, k$

Kritischer Bereich:

$$\left[\chi_c^2; +\infty\right)$$

$$\chi_c^2 = \chi_{1-\alpha}^2 (k-1-m)$$

Kolmogorov-Smirnov-Verteilungstest

Nullhypothese: H_0: Das Merkmal folgt einer bestimmten, genau festgelegten Verteilung.

Alternativhypothese: H_A: Das Merkmal folgt dieser bestimmten, genau festgelegten Verteilung nicht.

Prüfgröße: $d = \max_{z_i} \begin{cases} \left|F^e(z_i) - F^o(z_{i-1})\right| \\ \left|F^e(z_i) - F^o(z_i)\right| \end{cases}$

wobei $F^e(z)$ die Verteilungsfunktion der Grundgesamtheit (expected) und $F^o(z)$ die Summenhäufigkeitsfunktion der Stichprobe (observed) ist.

Testverteilung: d ist Kolmogorov-Smirnov-verteilt: d ~ $d_{\alpha,n}$
(n ist der Stichprobenumfang)

Kritischer Bereich: $[d_c; +\infty$
$d_c = d_{\alpha,n}$

Tabelle der Kolmogorov-Smirnov-Verteilung für $\alpha = 0{,}1$ und $\alpha = 0{,}05$
(Quelle: J. Schwarze, Grundlagen der Statistik II, 1993, S. 254):

n	$\alpha = 0{,}1$	$\alpha = 0{,}05$	n	$\alpha = 0{,}1$	$\alpha = 0{,}05$	n	$\alpha = 0{,}1$	$\alpha = 0{,}05$	n	$\alpha = 0{,}1$	$\alpha = 0{,}05$
3	0,636	0,708	13	0,325	0,361	23	0,247	0,275	33	0,208	0,231
4	0,565	0,624	14	0,314	0,349	24	0,242	0,269	34	0,205	0,227
5	0,509	0,563	15	0,304	0,338	25	0,238	0,264	35	0,202	0,224
6	0,468	0,519	16	0,295	0,327	26	0,233	0,259	36	0,199	0,221
7	0,436	0,483	17	0,286	0,318	27	0,229	0,254	37	0,196	0,218
8	0,410	0,454	18	0,278	0,309	28	0,225	0,250	38	0,194	0,215
9	0,387	0,430	19	0,271	0,301	29	0,221	0,246	39	0,191	0,213
10	0,369	0,409	20	0,265	0,294	30	0,218	0,242	40	0,189	0,210
11	0,352	0,391	21	0,259	0,287	31	0,214	0,238	50	0,170	0,188
12	0,338	0,375	22	0,253	0,281	32	0,211	0,234	100	0,121	0,134

Chi-Quadrat-Unabhängigkeitstest

Nullhypothese: H_0: Die Merkmale A und B sind voneinander unabhängig.

Alternativhypothese: H_A: Die Merkmale A und B sind voneinander abhängig.

Prüfgröße: $$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \tilde{n}_{ij}\right)^2}{\tilde{n}_{ij}}$$

wobei n_{ij} die beobachteten absoluten Häufigkeiten und $\tilde{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ die absoluten Häufigkeiten bei Unabhängigkeit sind.

Testverteilung: χ^2 ist χ^2-verteilt mit $\nu = (r-1) \cdot (s-1)$ Freiheitsgraden

Kritischer Bereich: $\left[\chi_c^2; +\infty\right); \chi_c^2 = \chi_{1-\alpha}\left((r-1) \cdot (s-1)\right)$

Binomialverteilung

(Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
1	0	0,9500	0,9500	0,9000	0,9000	0,8500	0,8500	0,8000	0,8000	0,7500	0,7500	0,7000	0,7000	0,6500	0,6500	0,6000	0,6000	0,5500	0,5500	0,5000	0,5000
1	1	0,0500	1,0000	0,1000	1,0000	0,1500	1,0000	0,2000	1,0000	0,2500	1,0000	0,3000	1,0000	0,3500	1,0000	0,4000	1,0000	0,4500	1,0000	0,5000	1,0000
2	0	0,9025	0,9025	0,8100	0,8100	0,7225	0,7225	0,6400	0,6400	0,5625	0,5625	0,4900	0,4900	0,4225	0,4225	0,3600	0,3600	0,3025	0,3025	0,2500	0,2500
2	1	0,0950	0,9975	0,1800	0,9900	0,2550	0,9775	0,3200	0,9600	0,3750	0,9375	0,4200	0,9100	0,4550	0,8775	0,4800	0,8400	0,4950	0,7975	0,5000	0,7500
2	2	0,0025	1,0000	0,0100	1,0000	0,0225	1,0000	0,0400	1,0000	0,0625	1,0000	0,0900	1,0000	0,1225	1,0000	0,1600	1,0000	0,2025	1,0000	0,2500	1,0000
3	0	0,8574	0,8574	0,7290	0,7290	0,6141	0,6141	0,5120	0,5120	0,4219	0,4219	0,3430	0,3430	0,2746	0,2746	0,2160	0,2160	0,1664	0,1664	0,1250	0,1250
3	1	0,1354	0,9928	0,2430	0,9720	0,3251	0,9393	0,3840	0,8960	0,4219	0,8438	0,4410	0,7840	0,4436	0,7183	0,4320	0,6480	0,4084	0,5748	0,3750	0,5000
3	2	0,0071	0,9999	0,0270	0,9990	0,0574	0,9966	0,0960	0,9920	0,1406	0,9844	0,1890	0,9730	0,2389	0,9571	0,2880	0,9360	0,3341	0,9089	0,3750	0,8750
3	3	0,0001	1,0000	0,0010	1,0000	0,0034	1,0000	0,0080	1,0000	0,0156	1,0000	0,0270	1,0000	0,0429	1,0000	0,0640	1,0000	0,0911	1,0000	0,1250	1,0000
4	0	0,8145	0,8145	0,6561	0,6561	0,5220	0,5220	0,4096	0,4096	0,3164	0,3164	0,2401	0,2401	0,1785	0,1785	0,1296	0,1296	0,0915	0,0915	0,0625	0,0625
4	1	0,1715	0,9860	0,2916	0,9477	0,3685	0,8905	0,4096	0,8192	0,4219	0,7383	0,4116	0,6517	0,3845	0,5630	0,3456	0,4752	0,2995	0,3910	0,2500	0,3125
4	2	0,0135	0,9995	0,0486	0,9963	0,0975	0,9880	0,1536	0,9728	0,2109	0,9492	0,2646	0,9163	0,3105	0,8735	0,3456	0,8208	0,3675	0,7585	0,3750	0,6875
4	3	0,0005	1,0000	0,0036	0,9999	0,0115	0,9995	0,0256	0,9984	0,0469	0,9961	0,0756	0,9919	0,1115	0,9850	0,1536	0,9744	0,2005	0,9590	0,2500	0,9375
4	4	0,0000	1,0000	0,0001	1,0000	0,0005	1,0000	0,0016	1,0000	0,0039	1,0000	0,0081	1,0000	0,0150	1,0000	0,0256	1,0000	0,0410	1,0000	0,0625	1,0000
5	0	0,7738	0,7738	0,5905	0,5905	0,4437	0,4437	0,3277	0,3277	0,2373	0,2373	0,1681	0,1681	0,1160	0,1160	0,0778	0,0778	0,0503	0,0503	0,0313	0,0313
5	1	0,2036	0,9774	0,3281	0,9185	0,3915	0,8352	0,4096	0,7373	0,3955	0,6328	0,3602	0,5282	0,3124	0,4284	0,2592	0,3370	0,2059	0,2562	0,1563	0,1875
5	2	0,0214	0,9988	0,0729	0,9914	0,1382	0,9734	0,2048	0,9421	0,2637	0,8965	0,3087	0,8369	0,3364	0,7648	0,3456	0,6826	0,3369	0,5931	0,3125	0,5000
5	3	0,0011	1,0000	0,0081	0,9995	0,0244	0,9978	0,0512	0,9933	0,0879	0,9844	0,1323	0,9692	0,1811	0,9460	0,2304	0,9130	0,2757	0,8688	0,3125	0,8125
5	4	0,0000	1,0000	0,0005	1,0000	0,0022	0,9999	0,0064	0,9997	0,0146	0,9990	0,0284	0,9976	0,0488	0,9947	0,0768	0,9898	0,1128	0,9815	0,1563	0,9688
5	5	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0010	1,0000	0,0024	1,0000	0,0053	1,0000	0,0102	1,0000	0,0185	1,0000	0,0313	1,0000
6	0	0,7351	0,7351	0,5314	0,5314	0,3771	0,3771	0,2621	0,2621	0,1780	0,1780	0,1176	0,1176	0,0754	0,0754	0,0467	0,0467	0,0277	0,0277	0,0156	0,0156
6	1	0,2321	0,9672	0,3543	0,8857	0,3993	0,7765	0,3932	0,6554	0,3560	0,5339	0,3025	0,4202	0,2437	0,3191	0,1866	0,2333	0,1359	0,1636	0,0938	0,1094
6	2	0,0305	0,9978	0,0984	0,9842	0,1762	0,9527	0,2458	0,9011	0,2966	0,8306	0,3241	0,7443	0,3280	0,6471	0,3110	0,5443	0,2780	0,4415	0,2344	0,3438
6	3	0,0021	0,9999	0,0146	0,9987	0,0415	0,9941	0,0819	0,9830	0,1318	0,9624	0,1852	0,9295	0,2355	0,8826	0,2765	0,8208	0,3032	0,7447	0,3125	0,6563
6	4	0,0001	1,0000	0,0012	0,9999	0,0055	0,9996	0,0154	0,9984	0,0330	0,9954	0,0595	0,9891	0,0951	0,9777	0,1382	0,9590	0,1861	0,9308	0,2344	0,8906
6	5	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0015	0,9999	0,0044	0,9998	0,0102	0,9993	0,0205	0,9982	0,0369	0,9959	0,0609	0,9917	0,0938	0,9844
6	6	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000	0,0007	1,0000	0,0018	1,0000	0,0041	1,0000	0,0083	1,0000	0,0156	1,0000
7	0	0,6983	0,6983	0,4783	0,4783	0,3206	0,3206	0,2097	0,2097	0,1335	0,1335	0,0824	0,0824	0,0490	0,0490	0,0280	0,0280	0,0152	0,0152	0,0078	0,0078
7	1	0,2573	0,9556	0,3720	0,8503	0,3960	0,7166	0,3670	0,5767	0,3115	0,4449	0,2471	0,3294	0,1848	0,2338	0,1306	0,1586	0,0872	0,1024	0,0547	0,0625
7	2	0,0406	0,9962	0,1240	0,9743	0,2097	0,9262	0,2753	0,8520	0,3115	0,7564	0,3177	0,6471	0,2985	0,5323	0,2613	0,4199	0,2140	0,3164	0,1641	0,2266
7	3	0,0036	0,9998	0,0230	0,9973	0,0617	0,9879	0,1147	0,9667	0,1730	0,9294	0,2269	0,8740	0,2679	0,8002	0,2903	0,7102	0,2918	0,6083	0,2734	0,5000
7	4	0,0002	1,0000	0,0026	0,9998	0,0109	0,9988	0,0287	0,9953	0,0577	0,9871	0,0972	0,9712	0,1442	0,9444	0,1935	0,9037	0,2388	0,8471	0,2734	0,7734
7	5	0,0000	1,0000	0,0002	1,0000	0,0012	0,9999	0,0043	0,9996	0,0115	0,9987	0,0250	0,9962	0,0466	0,9910	0,0774	0,9812	0,1172	0,9643	0,1641	0,9375
7	6	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0013	0,9999	0,0036	0,9998	0,0084	0,9994	0,0172	0,9984	0,0320	0,9963	0,0547	0,9922
7	7	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000	0,0006	1,0000	0,0016	1,0000	0,0037	1,0000	0,0078	1,0000

Binomialverteilung (Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
8	0	0,6634	0,6634	0,4305	0,4305	0,2725	0,2725	0,1678	0,1678	0,1001	0,1001	0,0576	0,0576	0,0319	0,0319	0,0168	0,0168	0,0084	0,0084	0,0039	0,0039
8	1	0,2793	0,9428	0,3826	0,8131	0,3847	0,6572	0,3355	0,5033	0,2670	0,3671	0,1977	0,2553	0,1373	0,1691	0,0896	0,1064	0,0548	0,0632	0,0313	0,0352
8	2	0,0515	0,9942	0,1488	0,9619	0,2376	0,8948	0,2936	0,7969	0,3115	0,6785	0,2965	0,5518	0,2587	0,4278	0,2090	0,3154	0,1569	0,2201	0,1094	0,1445
8	3	0,0054	0,9996	0,0331	0,9950	0,0839	0,9786	0,1468	0,9437	0,2076	0,8862	0,2541	0,8059	0,2786	0,7064	0,2787	0,5941	0,2568	0,4770	0,2188	0,3633
8	4	0,0004	1,0000	0,0046	0,9996	0,0185	0,9971	0,0459	0,9896	0,0865	0,9727	0,1361	0,9420	0,1875	0,8939	0,2322	0,8263	0,2627	0,7396	0,2734	0,6367
8	5	0,0000	1,0000	0,0004	1,0000	0,0026	0,9998	0,0092	0,9988	0,0231	0,9958	0,0467	0,9887	0,0808	0,9747	0,1239	0,9502	0,1719	0,9115	0,2188	0,8555
8	6	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0011	0,9999	0,0038	0,9996	0,0100	0,9987	0,0217	0,9964	0,0413	0,9915	0,0703	0,9819	0,1094	0,9648
8	7	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0012	0,9999	0,0033	0,9998	0,0079	0,9993	0,0164	0,9983	0,0313	0,9961
8	8	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000	0,0007	1,0000	0,0017	1,0000	0,0039	1,0000
9	0	0,6302	0,6302	0,3874	0,3874	0,2316	0,2316	0,1342	0,1342	0,0751	0,0751	0,0404	0,0404	0,0207	0,0207	0,0101	0,0101	0,0046	0,0046	0,0020	0,0020
9	1	0,2985	0,9288	0,3874	0,7748	0,3679	0,5995	0,3020	0,4362	0,2253	0,3003	0,1556	0,1960	0,1004	0,1211	0,0605	0,0705	0,0339	0,0385	0,0176	0,0195
9	2	0,0629	0,9916	0,1722	0,9470	0,2597	0,8591	0,3020	0,7382	0,3003	0,6007	0,2668	0,4628	0,2162	0,3373	0,1612	0,2318	0,1110	0,1495	0,0703	0,0898
9	3	0,0077	0,9994	0,0446	0,9917	0,1069	0,9661	0,1762	0,9144	0,2336	0,8343	0,2668	0,7297	0,2716	0,6089	0,2508	0,4826	0,2119	0,3614	0,1641	0,2539
9	4	0,0006	1,0000	0,0074	0,9991	0,0283	0,9944	0,0661	0,9804	0,1168	0,9511	0,1715	0,9012	0,2194	0,8283	0,2508	0,7334	0,2600	0,6214	0,2461	0,5000
9	5	0,0000	1,0000	0,0008	0,9999	0,0050	0,9994	0,0165	0,9969	0,0389	0,9900	0,0735	0,9747	0,1181	0,9464	0,1672	0,9006	0,2128	0,8342	0,2461	0,7461
9	6	0,0000	1,0000	0,0001	1,0000	0,0006	1,0000	0,0028	0,9997	0,0087	0,9987	0,0210	0,9957	0,0424	0,9888	0,0743	0,9750	0,1160	0,9502	0,1641	0,9102
9	7	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0012	0,9999	0,0039	0,9996	0,0098	0,9986	0,0212	0,9962	0,0407	0,9909	0,0703	0,9805
9	8	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0013	0,9999	0,0035	0,9997	0,0083	0,9992	0,0176	0,9980
9	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0008	1,0000	0,0020	1,0000
10	0	0,5987	0,5987	0,3487	0,3487	0,1969	0,1969	0,1074	0,1074	0,0563	0,0563	0,0282	0,0282	0,0135	0,0135	0,0060	0,0060	0,0025	0,0025	0,0010	0,0010
10	1	0,3151	0,9139	0,3874	0,7361	0,3474	0,5443	0,2684	0,3758	0,1877	0,2440	0,1211	0,1493	0,0725	0,0860	0,0403	0,0464	0,0207	0,0233	0,0098	0,0107
10	2	0,0746	0,9885	0,1937	0,9298	0,2759	0,8202	0,3020	0,6778	0,2816	0,5256	0,2335	0,3828	0,1757	0,2616	0,1209	0,1673	0,0763	0,0996	0,0439	0,0547
10	3	0,0105	0,9990	0,0574	0,9872	0,1298	0,9500	0,2013	0,8791	0,2503	0,7759	0,2668	0,6496	0,2522	0,5138	0,2150	0,3823	0,1665	0,2660	0,1172	0,1719
10	4	0,0010	0,9999	0,0112	0,9984	0,0401	0,9901	0,0881	0,9672	0,1460	0,9219	0,2001	0,8497	0,2377	0,7515	0,2508	0,6331	0,2384	0,5044	0,2051	0,3770
10	5	0,0001	1,0000	0,0015	0,9999	0,0085	0,9986	0,0264	0,9936	0,0584	0,9803	0,1029	0,9527	0,1536	0,9051	0,2007	0,8338	0,2340	0,7384	0,2461	0,6230
10	6	0,0000	1,0000	0,0001	1,0000	0,0012	0,9999	0,0055	0,9991	0,0162	0,9965	0,0368	0,9894	0,0689	0,9740	0,1115	0,9452	0,1596	0,8980	0,2051	0,8281
10	7	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0008	0,9999	0,0031	0,9996	0,0090	0,9984	0,0212	0,9952	0,0425	0,9877	0,0746	0,9726	0,1172	0,9453
10	8	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0014	0,9999	0,0043	0,9995	0,0106	0,9983	0,0229	0,9955	0,0439	0,9893
10	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	1,0000	0,0016	0,9999	0,0042	0,9997	0,0098	0,9990
10	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0010	1,0000
11	0	0,5688	0,5688	0,3138	0,3138	0,1673	0,1673	0,0859	0,0859	0,0422	0,0422	0,0198	0,0198	0,0088	0,0088	0,0036	0,0036	0,0014	0,0014	0,0005	0,0005
11	1	0,3293	0,8981	0,3835	0,6974	0,3248	0,4922	0,2362	0,3221	0,1549	0,1971	0,0932	0,1130	0,0518	0,0606	0,0266	0,0302	0,0125	0,0139	0,0054	0,0059
11	2	0,0867	0,9848	0,2131	0,9104	0,2866	0,7788	0,2953	0,6174	0,2581	0,4552	0,1998	0,3127	0,1395	0,2001	0,0887	0,1189	0,0513	0,0652	0,0269	0,0327
11	3	0,0137	0,9984	0,0710	0,9815	0,1517	0,9306	0,2215	0,8389	0,2581	0,7133	0,2568	0,5696	0,2254	0,4256	0,1774	0,2963	0,1259	0,1911	0,0806	0,1133
11	4	0,0014	0,9999	0,0158	0,9972	0,0536	0,9841	0,1107	0,9496	0,1721	0,8854	0,2201	0,7897	0,2428	0,6683	0,2365	0,5328	0,2060	0,3971	0,1611	0,2744
11	5	0,0001	1,0000	0,0025	0,9997	0,0132	0,9973	0,0388	0,9883	0,0803	0,9657	0,1321	0,9218	0,1830	0,8513	0,2207	0,7535	0,2360	0,6331	0,2256	0,5000
11	6	0,0000	1,0000	0,0003	1,0000	0,0023	0,9997	0,0097	0,9980	0,0268	0,9924	0,0566	0,9784	0,0985	0,9499	0,1471	0,9006	0,1931	0,8262	0,2256	0,7256
11	7	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0017	0,9998	0,0064	0,9988	0,0173	0,9957	0,0379	0,9878	0,0701	0,9707	0,1128	0,9390	0,1611	0,8867
11	8	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0011	0,9999	0,0037	0,9994	0,0102	0,9980	0,0234	0,9941	0,0462	0,9852	0,0806	0,9673
11	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	1,0000	0,0018	0,9998	0,0052	0,9993	0,0126	0,9978	0,0269	0,9941
11	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0007	1,0000	0,0021	0,9998	0,0054	0,9995
11	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0005	1,0000

Binomialverteilung

(Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
12	0	0,5404	0,5404	0,2824	0,2824	0,1422	0,1422	0,0687	0,0687	0,0317	0,0317	0,0138	0,0138	0,0057	0,0057	0,0022	0,0022	0,0008	0,0008	0,0002	0,0002
12	1	0,3413	0,8816	0,3766	0,6590	0,3012	0,4435	0,2062	0,2749	0,1267	0,1584	0,0712	0,0850	0,0368	0,0424	0,0174	0,0196	0,0075	0,0083	0,0029	0,0032
12	2	0,0988	0,9804	0,2301	0,8891	0,2924	0,7358	0,2835	0,5583	0,2323	0,3907	0,1678	0,2528	0,1088	0,1513	0,0639	0,0834	0,0339	0,0421	0,0161	0,0193
12	3	0,0173	0,9978	0,0852	0,9744	0,1720	0,9078	0,2362	0,7946	0,2581	0,6488	0,2397	0,4925	0,1954	0,3467	0,1419	0,2253	0,0923	0,1345	0,0537	0,0730
12	4	0,0021	0,9998	0,0213	0,9957	0,0683	0,9761	0,1329	0,9274	0,1936	0,8424	0,2311	0,7237	0,2367	0,5833	0,2128	0,4382	0,1700	0,3044	0,1208	0,1938
12	5	0,0002	1,0000	0,0038	0,9995	0,0193	0,9954	0,0532	0,9806	0,1032	0,9456	0,1585	0,8822	0,2039	0,7873	0,2270	0,6652	0,2225	0,5269	0,1934	0,3872
12	6	0,0000	1,0000	0,0005	0,9999	0,0040	0,9993	0,0155	0,9961	0,0401	0,9857	0,0792	0,9614	0,1281	0,9154	0,1766	0,8418	0,2124	0,7393	0,2256	0,6128
12	7	0,0000	1,0000	0,0000	1,0000	0,0006	0,9999	0,0033	0,9994	0,0115	0,9972	0,0291	0,9905	0,0591	0,9745	0,1009	0,9427	0,1489	0,8883	0,1934	0,8062
12	8	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0024	0,9996	0,0078	0,9983	0,0199	0,9944	0,0420	0,9847	0,0762	0,9644	0,1208	0,9270
12	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	1,0000	0,0015	0,9998	0,0048	0,9992	0,0125	0,9972	0,0277	0,9921	0,0537	0,9807
12	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0008	0,9999	0,0025	0,9997	0,0068	0,9989	0,0161	0,9968
12	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0010	0,9999	0,0029	0,9998
12	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000
13	0	0,5133	0,5133	0,2542	0,2542	0,1209	0,1209	0,0550	0,0550	0,0238	0,0238	0,0097	0,0097	0,0037	0,0037	0,0013	0,0013	0,0004	0,0004	0,0001	0,0001
13	1	0,3512	0,8646	0,3672	0,6213	0,2774	0,3983	0,1787	0,2336	0,1029	0,1267	0,0540	0,0637	0,0259	0,0296	0,0113	0,0126	0,0045	0,0049	0,0016	0,0017
13	2	0,1109	0,9755	0,2448	0,8661	0,2937	0,6920	0,2680	0,5017	0,2059	0,3326	0,1388	0,2025	0,0836	0,1132	0,0453	0,0579	0,0220	0,0269	0,0095	0,0112
13	3	0,0214	0,9969	0,0997	0,9658	0,1900	0,8820	0,2457	0,7473	0,2517	0,5843	0,2181	0,4206	0,1651	0,2783	0,1107	0,1686	0,0660	0,0929	0,0349	0,0461
13	4	0,0028	0,9997	0,0277	0,9935	0,0838	0,9658	0,1535	0,9009	0,2097	0,7940	0,2337	0,6543	0,2222	0,5005	0,1845	0,3530	0,1350	0,2279	0,0873	0,1334
13	5	0,0003	1,0000	0,0055	0,9991	0,0266	0,9925	0,0691	0,9700	0,1258	0,9198	0,1803	0,8346	0,2154	0,7159	0,2214	0,5744	0,1989	0,4268	0,1571	0,2905
13	6	0,0000	1,0000	0,0008	0,9999	0,0063	0,9987	0,0230	0,9930	0,0559	0,9757	0,1030	0,9376	0,1546	0,8705	0,1968	0,7712	0,2169	0,6437	0,2095	0,5000
13	7	0,0000	1,0000	0,0001	1,0000	0,0011	0,9998	0,0058	0,9988	0,0186	0,9944	0,0442	0,9818	0,0833	0,9538	0,1312	0,9023	0,1775	0,8212	0,2095	0,7095
13	8	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0011	0,9998	0,0047	0,9990	0,0142	0,9960	0,0336	0,9874	0,0656	0,9679	0,1089	0,9302	0,1571	0,8666
13	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0009	0,9999	0,0034	0,9993	0,0101	0,9975	0,0243	0,9922	0,0495	0,9797	0,0873	0,9539
13	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0006	0,9999	0,0022	0,9997	0,0065	0,9987	0,0162	0,9959	0,0349	0,9888
13	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0012	0,9999	0,0036	0,9995	0,0095	0,9983
13	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	1,0000	0,0016	0,9999
13	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000
14	0	0,4877	0,4877	0,2288	0,2288	0,1028	0,1028	0,0440	0,0440	0,0178	0,0178	0,0068	0,0068	0,0024	0,0024	0,0008	0,0008	0,0002	0,0002	0,0001	0,0001
14	1	0,3593	0,8470	0,3559	0,5846	0,2539	0,3567	0,1539	0,1979	0,0832	0,1010	0,0407	0,0475	0,0181	0,0205	0,0073	0,0081	0,0027	0,0029	0,0009	0,0009
14	2	0,1229	0,9699	0,2570	0,8416	0,2912	0,6479	0,2501	0,4481	0,1802	0,2811	0,1134	0,1608	0,0634	0,0839	0,0317	0,0398	0,0141	0,0170	0,0056	0,0065
14	3	0,0259	0,9958	0,1142	0,9559	0,2056	0,8535	0,2501	0,6982	0,2402	0,5213	0,1943	0,3552	0,1366	0,2205	0,0845	0,1243	0,0462	0,0632	0,0222	0,0287
14	4	0,0037	0,9996	0,0349	0,9908	0,0998	0,9533	0,1720	0,8702	0,2202	0,7415	0,2290	0,5842	0,2022	0,4227	0,1549	0,2793	0,1040	0,1672	0,0611	0,0898
14	5	0,0004	1,0000	0,0078	0,9985	0,0352	0,9885	0,0860	0,9561	0,1468	0,8883	0,1963	0,7805	0,2178	0,6405	0,2066	0,4859	0,1701	0,3373	0,1222	0,2120
14	6	0,0000	1,0000	0,0013	0,9998	0,0093	0,9978	0,0322	0,9884	0,0734	0,9617	0,1262	0,9067	0,1759	0,8164	0,2066	0,6925	0,2088	0,5461	0,1833	0,3953
14	7	0,0000	1,0000	0,0002	1,0000	0,0019	0,9997	0,0092	0,9976	0,0280	0,9897	0,0618	0,9685	0,1082	0,9247	0,1574	0,8499	0,1952	0,7414	0,2095	0,6047
14	8	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0020	0,9996	0,0082	0,9978	0,0232	0,9917	0,0510	0,9757	0,0918	0,9417	0,1398	0,8811	0,1833	0,7880
14	9	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0018	0,9997	0,0066	0,9983	0,0183	0,9940	0,0408	0,9825	0,0762	0,9574	0,1222	0,9102
14	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0014	0,9998	0,0049	0,9989	0,0136	0,9961	0,0312	0,9886	0,0611	0,9713
14	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0010	0,9999	0,0033	0,9994	0,0093	0,9978	0,0222	0,9935
14	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0019	0,9997	0,0056	0,9991
14	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000	0,0009	0,9999
14	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000

Binomialverteilung

(Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
15	0	0,4633	0,4633	0,2059	0,2059	0,0874	0,0874	0,0352	0,0352	0,0134	0,0134	0,0047	0,0047	0,0016	0,0016	0,0005	0,0005	0,0001	0,0001	0,0000	0,0000
15	1	0,3658	0,8290	0,3432	0,5490	0,2312	0,3186	0,1319	0,1671	0,0668	0,0802	0,0305	0,0353	0,0126	0,0142	0,0047	0,0052	0,0016	0,0017	0,0005	0,0005
15	2	0,1348	0,9638	0,2669	0,8159	0,2856	0,6042	0,2309	0,3980	0,1559	0,2361	0,0916	0,1268	0,0476	0,0617	0,0219	0,0271	0,0090	0,0107	0,0032	0,0037
15	3	0,0307	0,9945	0,1285	0,9444	0,2184	0,8227	0,2501	0,6482	0,2252	0,4613	0,1700	0,2969	0,1110	0,1727	0,0634	0,0905	0,0318	0,0424	0,0139	0,0176
15	4	0,0049	0,9994	0,0428	0,9873	0,1156	0,9383	0,1876	0,8358	0,2252	0,6865	0,2186	0,5155	0,1792	0,3519	0,1268	0,2173	0,0780	0,1204	0,0417	0,0592
15	5	0,0006	0,9999	0,0105	0,9978	0,0449	0,9832	0,1032	0,9389	0,1651	0,8516	0,2061	0,7216	0,2123	0,5643	0,1859	0,4032	0,1404	0,2608	0,0916	0,1509
15	6	0,0000	1,0000	0,0019	0,9997	0,0132	0,9964	0,0430	0,9819	0,0917	0,9434	0,1472	0,8689	0,1906	0,7548	0,2066	0,6098	0,1914	0,4522	0,1527	0,3036
15	7	0,0000	1,0000	0,0003	1,0000	0,0030	0,9994	0,0138	0,9958	0,0393	0,9827	0,0811	0,9500	0,1319	0,8868	0,1771	0,7869	0,2013	0,6535	0,1964	0,5000
15	8	0,0000	1,0000	0,0000	1,0000	0,0005	0,9999	0,0035	0,9992	0,0131	0,9958	0,0348	0,9848	0,0710	0,9578	0,1181	0,9050	0,1647	0,8182	0,1964	0,6964
15	9	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0007	0,9999	0,0034	0,9992	0,0116	0,9963	0,0298	0,9876	0,0612	0,9662	0,1048	0,9231	0,1527	0,8491
15	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0007	0,9999	0,0030	0,9993	0,0096	0,9972	0,0245	0,9907	0,0515	0,9745	0,0916	0,9408
15	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0006	0,9999	0,0024	0,9995	0,0074	0,9981	0,0191	0,9937	0,0417	0,9824
15	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	0,9999	0,0016	0,9997	0,0052	0,9989	0,0139	0,9963
15	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0010	0,9999	0,0032	0,9995
15	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	1,0000
15	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
16	0	0,4401	0,4401	0,1853	0,1853	0,0743	0,0743	0,0281	0,0281	0,0100	0,0100	0,0033	0,0033	0,0010	0,0010	0,0003	0,0003	0,0001	0,0001	0,0000	0,0000
16	1	0,3706	0,8108	0,3294	0,5147	0,2097	0,2839	0,1126	0,1407	0,0535	0,0635	0,0228	0,0261	0,0087	0,0098	0,0030	0,0033	0,0009	0,0010	0,0002	0,0003
16	2	0,1463	0,9571	0,2745	0,7892	0,2775	0,5614	0,2111	0,3518	0,1336	0,1971	0,0732	0,0994	0,0353	0,0451	0,0150	0,0183	0,0056	0,0066	0,0018	0,0021
16	3	0,0359	0,9930	0,1423	0,9316	0,2285	0,7899	0,2463	0,5981	0,2079	0,4050	0,1465	0,2459	0,0888	0,1339	0,0468	0,0651	0,0215	0,0281	0,0085	0,0106
16	4	0,0061	0,9991	0,0514	0,9830	0,1311	0,9209	0,2001	0,7982	0,2252	0,6302	0,2040	0,4499	0,1553	0,2892	0,1014	0,1666	0,0572	0,0853	0,0278	0,0384
16	5	0,0008	0,9999	0,0137	0,9967	0,0555	0,9765	0,1201	0,9183	0,1802	0,8103	0,2099	0,6598	0,2008	0,4900	0,1623	0,3288	0,1123	0,1976	0,0667	0,1051
16	6	0,0001	1,0000	0,0028	0,9995	0,0180	0,9944	0,0550	0,9733	0,1101	0,9204	0,1649	0,8247	0,1982	0,6881	0,1983	0,5272	0,1684	0,3660	0,1222	0,2272
16	7	0,0000	1,0000	0,0004	0,9999	0,0045	0,9989	0,0197	0,9930	0,0524	0,9729	0,1010	0,9256	0,1524	0,8406	0,1889	0,7161	0,1969	0,5629	0,1746	0,4018
16	8	0,0000	1,0000	0,0001	1,0000	0,0009	0,9998	0,0055	0,9985	0,0197	0,9925	0,0487	0,9743	0,0923	0,9329	0,1417	0,8577	0,1812	0,7441	0,1964	0,5982
16	9	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0012	0,9998	0,0058	0,9984	0,0185	0,9929	0,0442	0,9771	0,0840	0,9417	0,1318	0,8759	0,1746	0,7728
16	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0014	0,9997	0,0056	0,9984	0,0167	0,9938	0,0392	0,9809	0,0755	0,9514	0,1222	0,8949
16	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0013	0,9997	0,0049	0,9987	0,0142	0,9951	0,0337	0,9851	0,0667	0,9616
16	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0011	0,9998	0,0040	0,9991	0,0115	0,9965	0,0278	0,9894
16	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0008	0,9999	0,0029	0,9994	0,0085	0,9979
16	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0018	0,9997
16	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	1,0000
16	16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
17	0	0,4181	0,4181	0,1668	0,1668	0,0631	0,0631	0,0225	0,0225	0,0075	0,0075	0,0023	0,0023	0,0007	0,0007	0,0002	0,0002	0,0000	0,0000	0,0000	0,0000
17	1	0,3741	0,7922	0,3150	0,4818	0,1893	0,2525	0,0957	0,1182	0,0426	0,0501	0,0169	0,0193	0,0060	0,0067	0,0019	0,0021	0,0005	0,0006	0,0001	0,0001
17	2	0,1575	0,9497	0,2800	0,7618	0,2673	0,5198	0,1914	0,3096	0,1136	0,1637	0,0581	0,0774	0,0260	0,0327	0,0102	0,0123	0,0035	0,0041	0,0010	0,0012
17	3	0,0415	0,9912	0,1556	0,9174	0,2359	0,7556	0,2393	0,5489	0,1893	0,3530	0,1245	0,2019	0,0701	0,1028	0,0341	0,0464	0,0144	0,0184	0,0052	0,0064
17	4	0,0076	0,9988	0,0605	0,9779	0,1457	0,9013	0,2093	0,7582	0,2209	0,5739	0,1868	0,3887	0,1320	0,2348	0,0796	0,1260	0,0411	0,0596	0,0182	0,0245
17	5	0,0010	0,9999	0,0175	0,9953	0,0668	0,9681	0,1361	0,8943	0,1914	0,7653	0,2081	0,5968	0,1849	0,4197	0,1379	0,2639	0,0875	0,1471	0,0472	0,0717
17	6	0,0001	1,0000	0,0039	0,9992	0,0236	0,9917	0,0680	0,9623	0,1276	0,8929	0,1784	0,7752	0,1991	0,6188	0,1839	0,4478	0,1432	0,2902	0,0944	0,1662
17	7	0,0000	1,0000	0,0007	0,9999	0,0065	0,9983	0,0267	0,9891	0,0668	0,9598	0,1201	0,8954	0,1685	0,7872	0,1927	0,6405	0,1841	0,4743	0,1484	0,3145
17	8	0,0000	1,0000	0,0001	1,0000	0,0014	0,9997	0,0084	0,9974	0,0279	0,9876	0,0644	0,9597	0,1134	0,9006	0,1606	0,8011	0,1883	0,6626	0,1855	0,5000
17	9	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0021	0,9995	0,0093	0,9969	0,0276	0,9873	0,0611	0,9617	0,1070	0,9081	0,1540	0,8166	0,1855	0,6855
17	10	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0004	0,9999	0,0025	0,9994	0,0095	0,9968	0,0263	0,9880	0,0571	0,9652	0,1008	0,9174	0,1484	0,8338
17	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0026	0,9993	0,0090	0,9970	0,0242	0,9894	0,0525	0,9699	0,0944	0,9283

Binomialverteilung

(Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
17	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0006	0,9999	0,0024	0,9994	0,0081	0,9975	0,0215	0,9914	0,0472	0,9755
17	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0021	0,9995	0,0068	0,9981	0,0182	0,9936
17	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0004	0,9999	0,0016	0,9997	0,0052	0,9988
17	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000	0,0010	0,9999
17	16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000
17	17	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
18	0	0,3972	0,3972	0,1501	0,1501	0,0536	0,0536	0,0180	0,0180	0,0056	0,0056	0,0016	0,0016	0,0004	0,0004	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000
18	1	0,3763	0,7735	0,3002	0,4503	0,1704	0,2241	0,0811	0,0991	0,0338	0,0395	0,0126	0,0142	0,0042	0,0046	0,0012	0,0013	0,0003	0,0003	0,0001	0,0001
18	2	0,1683	0,9419	0,2835	0,7338	0,2556	0,4797	0,1723	0,2713	0,0958	0,1353	0,0458	0,0600	0,0190	0,0236	0,0069	0,0082	0,0022	0,0025	0,0006	0,0007
18	3	0,0473	0,9891	0,1680	0,9018	0,2406	0,7202	0,2297	0,5010	0,1704	0,3057	0,1046	0,1646	0,0547	0,0783	0,0246	0,0328	0,0095	0,0120	0,0031	0,0038
18	4	0,0093	0,9985	0,0700	0,9718	0,1592	0,8794	0,2153	0,7164	0,2130	0,5187	0,1681	0,3327	0,1104	0,1886	0,0614	0,0942	0,0291	0,0411	0,0117	0,0154
18	5	0,0014	0,9998	0,0218	0,9936	0,0787	0,9581	0,1507	0,8671	0,1988	0,7175	0,2017	0,5344	0,1664	0,3550	0,1146	0,2088	0,0666	0,1077	0,0327	0,0481
18	6	0,0002	1,0000	0,0052	0,9988	0,0301	0,9882	0,0816	0,9487	0,1436	0,8610	0,1873	0,7217	0,1941	0,5491	0,1655	0,3743	0,1181	0,2258	0,0708	0,1189
18	7	0,0000	1,0000	0,0010	0,9998	0,0091	0,9973	0,0350	0,9837	0,0820	0,9431	0,1376	0,8593	0,1792	0,7283	0,1892	0,5634	0,1657	0,3915	0,1214	0,2403
18	8	0,0000	1,0000	0,0002	1,0000	0,0022	0,9995	0,0120	0,9957	0,0376	0,9807	0,0811	0,9404	0,1327	0,8609	0,1734	0,7368	0,1864	0,5778	0,1669	0,4073
18	9	0,0000	1,0000	0,0000	1,0000	0,0004	0,9999	0,0033	0,9991	0,0139	0,9946	0,0386	0,9790	0,0794	0,9403	0,1284	0,8653	0,1694	0,7473	0,1855	0,5927
18	10	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0008	0,9998	0,0042	0,9988	0,0149	0,9939	0,0385	0,9788	0,0771	0,9424	0,1248	0,8720	0,1669	0,7597
18	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0010	0,9998	0,0046	0,9986	0,0151	0,9938	0,0374	0,9797	0,0742	0,9463	0,1214	0,8811
18	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0012	0,9997	0,0047	0,9986	0,0145	0,9942	0,0354	0,9817	0,0708	0,9519
18	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0012	0,9997	0,0045	0,9987	0,0134	0,9951	0,0327	0,9846
18	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0011	0,9998	0,0039	0,9990	0,0117	0,9962
18	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0009	0,9999	0,0031	0,9993
18	16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0006	0,9999
18	17	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000
18	18	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
19	0	0,3774	0,3774	0,1351	0,1351	0,0456	0,0456	0,0144	0,0144	0,0042	0,0042	0,0011	0,0011	0,0003	0,0003	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000
19	1	0,3774	0,7547	0,2852	0,4203	0,1529	0,1985	0,0685	0,0829	0,0268	0,0310	0,0093	0,0104	0,0029	0,0031	0,0008	0,0008	0,0002	0,0002	0,0000	0,0000
19	2	0,1787	0,9335	0,2852	0,7054	0,2428	0,4413	0,1540	0,2369	0,0803	0,1113	0,0358	0,0462	0,0138	0,0170	0,0046	0,0055	0,0013	0,0015	0,0003	0,0004
19	3	0,0533	0,9868	0,1796	0,8850	0,2428	0,6841	0,2182	0,4551	0,1517	0,2631	0,0869	0,1332	0,0422	0,0591	0,0175	0,0230	0,0062	0,0077	0,0018	0,0022
19	4	0,0112	0,9980	0,0798	0,9648	0,1714	0,8556	0,2182	0,6733	0,2023	0,4654	0,1491	0,2822	0,0909	0,1500	0,0467	0,0696	0,0203	0,0280	0,0074	0,0096
19	5	0,0018	0,9998	0,0266	0,9914	0,0907	0,9463	0,1636	0,8369	0,2023	0,6678	0,1916	0,4739	0,1468	0,2968	0,0933	0,1629	0,0497	0,0777	0,0222	0,0318
19	6	0,0002	1,0000	0,0069	0,9983	0,0374	0,9837	0,0955	0,9324	0,1574	0,8251	0,1916	0,6655	0,1844	0,4812	0,1451	0,3081	0,0949	0,1727	0,0518	0,0835
19	7	0,0000	1,0000	0,0014	0,9997	0,0122	0,9959	0,0443	0,9767	0,0974	0,9225	0,1525	0,8180	0,1844	0,6656	0,1797	0,4878	0,1443	0,3169	0,0961	0,1796
19	8	0,0000	1,0000	0,0002	1,0000	0,0032	0,9992	0,0166	0,9933	0,0487	0,9713	0,0981	0,9161	0,1489	0,8145	0,1797	0,6675	0,1771	0,4940	0,1442	0,3238
19	9	0,0000	1,0000	0,0000	1,0000	0,0007	0,9999	0,0051	0,9984	0,0198	0,9911	0,0514	0,9674	0,0980	0,9125	0,1464	0,8139	0,1771	0,6710	0,1762	0,5000
19	10	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0013	0,9997	0,0066	0,9977	0,0220	0,9895	0,0528	0,9653	0,0976	0,9115	0,1449	0,8159	0,1762	0,6762
19	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0018	0,9995	0,0077	0,9972	0,0233	0,9886	0,0532	0,9648	0,0970	0,9129	0,1442	0,8204
19	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0004	0,9999	0,0022	0,9994	0,0083	0,9969	0,0237	0,9884	0,0529	0,9658	0,0961	0,9165
19	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0024	0,9993	0,0085	0,9969	0,0233	0,9891	0,0518	0,9682
19	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0006	0,9999	0,0024	0,9994	0,0082	0,9972	0,0222	0,9904
19	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0022	0,9995	0,0074	0,9978
19	16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0005	0,9999	0,0018	0,9996
19	17	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	1,0000
19	18	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
19	19	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000

Binomialverteilung

(Quelle: eigene Berechnungen)

p->		0,05		0,1		0,15		0,2		0,25		0,3		0,35		0,4		0,45		0,5	
n	x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
20	0	0,3585	0,3585	0,1216	0,1216	0,0388	0,0388	0,0115	0,0115	0,0032	0,0032	0,0008	0,0008	0,0002	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
20	1	0,3774	0,7358	0,2702	0,3917	0,1368	0,1756	0,0576	0,0692	0,0211	0,0243	0,0068	0,0076	0,0020	0,0021	0,0005	0,0005	0,0001	0,0001	0,0000	0,0000
20	2	0,1887	0,9245	0,2852	0,6769	0,2293	0,4049	0,1369	0,2061	0,0669	0,0913	0,0278	0,0355	0,0100	0,0121	0,0031	0,0036	0,0008	0,0009	0,0002	0,0002
20	3	0,0596	0,9841	0,1901	0,8670	0,2428	0,6477	0,2054	0,4114	0,1339	0,2252	0,0716	0,1071	0,0323	0,0444	0,0123	0,0160	0,0040	0,0049	0,0011	0,0013
20	4	0,0133	0,9974	0,0898	0,9568	0,1821	0,8298	0,2182	0,6296	0,1897	0,4148	0,1304	0,2375	0,0738	0,1182	0,0350	0,0510	0,0139	0,0189	0,0046	0,0059
20	5	0,0022	0,9997	0,0319	0,9887	0,1028	0,9327	0,1746	0,8042	0,2023	0,6172	0,1789	0,4164	0,1272	0,2454	0,0746	0,1256	0,0365	0,0553	0,0148	0,0207
20	6	0,0003	1,0000	0,0089	0,9976	0,0454	0,9781	0,1091	0,9133	0,1686	0,7858	0,1916	0,6080	0,1712	0,4166	0,1244	0,2500	0,0746	0,1299	0,0370	0,0577
20	7	0,0000	1,0000	0,0020	0,9996	0,0160	0,9941	0,0545	0,9679	0,1124	0,8982	0,1643	0,7723	0,1844	0,6010	0,1659	0,4159	0,1221	0,2520	0,0739	0,1316
20	8	0,0000	1,0000	0,0004	0,9999	0,0046	0,9987	0,0222	0,9900	0,0609	0,9591	0,1144	0,8867	0,1614	0,7624	0,1797	0,5956	0,1623	0,4143	0,1201	0,2517
20	9	0,0000	1,0000	0,0001	1,0000	0,0011	0,9998	0,0074	0,9974	0,0271	0,9861	0,0654	0,9520	0,1158	0,8782	0,1597	0,7553	0,1771	0,5914	0,1602	0,4119
20	10	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0020	0,9994	0,0099	0,9961	0,0308	0,9829	0,0686	0,9468	0,1171	0,8725	0,1593	0,7507	0,1762	0,5881
20	11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0005	0,9999	0,0030	0,9991	0,0120	0,9949	0,0336	0,9804	0,0710	0,9435	0,1185	0,8692	0,1602	0,7483
20	12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0008	0,9998	0,0039	0,9987	0,0136	0,9940	0,0355	0,9790	0,0727	0,9420	0,1201	0,8684
20	13	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0010	0,9997	0,0045	0,9985	0,0146	0,9935	0,0366	0,9786	0,0739	0,9423
20	14	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0012	0,9997	0,0049	0,9984	0,0150	0,9936	0,0370	0,9793
20	15	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0013	0,9997	0,0049	0,9985	0,0148	0,9941
20	16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0003	1,0000	0,0013	0,9997	0,0046	0,9987
20	17	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000	0,0011	0,9998
20	18	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	1,0000
20	19	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
20	20	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000

m=μ (Erwartungswert) = n · p

Poisson-Verteilung (Quelle: eigene Berechnungen)

m	0,01		0,02		0,03		0,04		0,05	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,9900	0,9900	0,9802	0,9802	0,9704	0,9704	0,9608	0,9608	0,9512	0,9512
1	0,0099	1,0000	0,0196	0,9998	0,0291	0,9996	0,0384	0,9992	0,0476	0,998
2	0,0000	1,0000	0,0002	1,0000	0,0004	1,0000	0,0008	1,0000	0,0012	1,000
3	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,000

m	0,06		0,07		0,08		0,09		0,10	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,9418	0,9418	0,9324	0,9324	0,9231	0,9231	0,9139	0,9139	0,9048	0,904
1	0,0565	0,9983	0,0653	0,9977	0,0738	0,9970	0,0823	0,9962	0,0905	0,995
2	0,0017	1,0000	0,0023	0,9999	0,0030	0,9999	0,0037	0,9999	0,0045	0,999
3	0,0000	1,0000	0,0001	1,0000	0,0001	1,0000	0,0001	1,0000	0,0002	1,000

m	0,11		0,12		0,15		0,20		0,25	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,8958	0,8958	0,8869	0,8869	0,8607	0,8607	0,8187	0,8187	0,7788	0,778
1	0,0985	0,9944	0,1064	0,9934	0,1291	0,9898	0,1637	0,9825	0,1947	0,973
2	0,0054	0,9998	0,0064	0,9997	0,0097	0,9995	0,0164	0,9989	0,0243	0,997
3	0,0002	1,0000	0,0003	1,0000	0,0005	1,0000	0,0011	0,9999	0,0020	0,999
4	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0001	1,000

m	0,30		0,35		0,40		0,45		0,50	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,7408	0,7408	0,7047	0,7047	0,6703	0,6703	0,6376	0,6376	0,6065	0,606
1	0,2222	0,9631	0,2466	0,9513	0,2681	0,9384	0,2869	0,9246	0,3033	0,909
2	0,0333	0,9964	0,0432	0,9945	0,0536	0,9921	0,0646	0,9891	0,0758	0,985
3	0,0033	0,9997	0,0050	0,9995	0,0072	0,9992	0,0097	0,9988	0,0126	0,998
4	0,0003	1,0000	0,0004	1,0000	0,0007	0,9999	0,0011	0,9999	0,0016	0,999
5	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0001	1,0000	0,0002	1,000

m	0,55		0,60		0,65		0,70		0,75	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,5769	0,5769	0,5488	0,5488	0,5220	0,5220	0,4966	0,4966	0,4724	0,472
1	0,3173	0,8943	0,3293	0,8781	0,3393	0,8614	0,3476	0,8442	0,3543	0,826
2	0,0873	0,9815	0,0988	0,9769	0,1103	0,9717	0,1217	0,9659	0,1329	0,959
3	0,0160	0,9975	0,0198	0,9966	0,0239	0,9956	0,0284	0,9942	0,0332	0,992
4	0,0022	0,9997	0,0030	0,9996	0,0039	0,9994	0,0050	0,9992	0,0062	0,998
5	0,0002	1,0000	0,0004	1,0000	0,0005	0,9999	0,0007	0,9999	0,0009	0,999
6	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0001	1,0000	0,0001	1,000

m	0,80		0,85		0,90		0,95		1,00	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,4493	0,4493	0,4274	0,4274	0,4066	0,4066	0,3867	0,3867	0,3679	0,367
1	0,3595	0,8088	0,3633	0,7907	0,3659	0,7725	0,3674	0,7541	0,3679	0,735
2	0,1438	0,9526	0,1544	0,9451	0,1647	0,9371	0,1745	0,9287	0,1839	0,919
3	0,0383	0,9909	0,0437	0,9889	0,0494	0,9865	0,0553	0,9839	0,0613	0,981
4	0,0077	0,9986	0,0093	0,9982	0,0111	0,9977	0,0131	0,9971	0,0153	0,996
5	0,0012	0,9998	0,0016	0,9997	0,0020	0,9997	0,0025	0,9995	0,0031	0,999
6	0,0002	1,0000	0,0002	1,0000	0,0003	1,0000	0,0004	0,9999	0,0005	0,999
7	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0001	1,000

Poisson-Verteilung (Quelle: eigene Berechnungen)

m	1,10		1,20		1,30		1,40		1,50	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,3329	0,3329	0,3012	0,3012	0,2725	0,2725	0,2466	0,2466	0,2231	0,2231
1	0,3662	0,6990	0,3614	0,6626	0,3543	0,6268	0,3452	0,5918	0,3347	0,5578
2	0,2014	0,9004	0,2169	0,8795	0,2303	0,8571	0,2417	0,8335	0,2510	0,8088
3	0,0738	0,9743	0,0867	0,9662	0,0998	0,9569	0,1128	0,9463	0,1255	0,9344
4	0,0203	0,9946	0,0260	0,9923	0,0324	0,9893	0,0395	0,9857	0,0471	0,9814
5	0,0045	0,9990	0,0062	0,9985	0,0084	0,9978	0,0111	0,9968	0,0141	0,9955
6	0,0008	0,9999	0,0012	0,9997	0,0018	0,9996	0,0026	0,9994	0,0035	0,9991
7	0,0001	1,0000	0,0002	1,0000	0,0003	0,9999	0,0005	0,9999	0,0008	0,9998
8	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0001	1,0000	0,0001	1,0000

m	1,60		1,70		1,80		1,90		2,00	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,2019	0,2019	0,1827	0,1827	0,1653	0,1653	0,1496	0,1496	0,1353	0,1353
1	0,3230	0,5249	0,3106	0,4932	0,2975	0,4628	0,2842	0,4337	0,2707	0,4060
2	0,2584	0,7834	0,2640	0,7572	0,2678	0,7306	0,2700	0,7037	0,2707	0,6767
3	0,1378	0,9212	0,1496	0,9068	0,1607	0,8913	0,1710	0,8747	0,1804	0,8571
4	0,0551	0,9763	0,0636	0,9704	0,0723	0,9636	0,0812	0,9559	0,0902	0,9473
5	0,0176	0,9940	0,0216	0,9920	0,0260	0,9896	0,0309	0,9868	0,0361	0,9834
6	0,0047	0,9987	0,0061	0,9981	0,0078	0,9974	0,0098	0,9966	0,0120	0,9955
7	0,0011	0,9997	0,0015	0,9996	0,0020	0,9994	0,0027	0,9992	0,0034	0,9989
8	0,0002	1,0000	0,0003	0,9999	0,0005	0,9999	0,0006	0,9998	0,0009	0,9998
9	0,0000	1,0000	0,0001	1,0000	0,0001	1,0000	0,0001	1,0000	0,0002	1,0000

m	2,10		2,20		2,30		2,40		2,50	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,1225	0,1225	0,1108	0,1108	0,1003	0,1003	0,0907	0,0907	0,0821	0,0821
1	0,2572	0,3796	0,2438	0,3546	0,2306	0,3309	0,2177	0,3084	0,2052	0,2873
2	0,2700	0,6496	0,2681	0,6227	0,2652	0,5960	0,2613	0,5697	0,2565	0,5438
3	0,1890	0,8386	0,1966	0,8194	0,2033	0,7993	0,2090	0,7787	0,2138	0,7576
4	0,0992	0,9379	0,1082	0,9275	0,1169	0,9162	0,1254	0,9041	0,1336	0,8912
5	0,0417	0,9796	0,0476	0,9751	0,0538	0,9700	0,0602	0,9643	0,0668	0,9580
6	0,0146	0,9941	0,0174	0,9925	0,0206	0,9906	0,0241	0,9884	0,0278	0,9858
7	0,0044	0,9985	0,0055	0,9980	0,0068	0,9974	0,0083	0,9967	0,0099	0,9958
8	0,0011	0,9997	0,0015	0,9995	0,0019	0,9994	0,0025	0,9991	0,0031	0,9989
9	0,0003	0,9999	0,0004	0,9999	0,0005	0,9999	0,0007	0,9998	0,0009	0,9997
10	0,0001	1,0000	0,0001	1,0000	0,0001	1,0000	0,0002	1,0000	0,0002	0,9999
11	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000

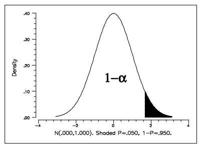
m	2,60		2,70		2,80		2,90		3,00	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,0743	0,0743	0,0672	0,0672	0,0608	0,0608	0,0550	0,0550	0,0498	0,0498
1	0,1931	0,2674	0,1815	0,2487	0,1703	0,2311	0,1596	0,2146	0,1494	0,1991
2	0,2510	0,5184	0,2450	0,4936	0,2384	0,4695	0,2314	0,4460	0,2240	0,4232
3	0,2176	0,7360	0,2205	0,7141	0,2225	0,6919	0,2237	0,6696	0,2240	0,6472
4	0,1414	0,8774	0,1488	0,8629	0,1557	0,8477	0,1622	0,8318	0,1680	0,8153
5	0,0735	0,9510	0,0804	0,9433	0,0872	0,9349	0,0940	0,9258	0,1008	0,9161
6	0,0319	0,9828	0,0362	0,9794	0,0407	0,9756	0,0455	0,9713	0,0504	0,9665
7	0,0118	0,9947	0,0139	0,9934	0,0163	0,9919	0,0188	0,9901	0,0216	0,9881
8	0,0038	0,9985	0,0047	0,9981	0,0057	0,9976	0,0068	0,9969	0,0081	0,9962
9	0,0011	0,9996	0,0014	0,9995	0,0018	0,9993	0,0022	0,9991	0,0027	0,9989
10	0,0003	0,9999	0,0004	0,9999	0,0005	0,9998	0,0006	0,9998	0,0008	0,9997
11	0,0001	1,0000	0,0001	1,0000	0,0001	1,0000	0,0002	0,9999	0,0002	0,9999
12	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000

Poisson-Verteilung (Quelle: eigene Berechnungen)

m	3,50		4,00		4,50		5,00		5,50	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,0302	0,0302	0,0183	0,0183	0,0111	0,0111	0,0067	0,0067	0,0041	0,0041
1	0,1057	0,1359	0,0733	0,0916	0,0500	0,0611	0,0337	0,0404	0,0225	0,0266
2	0,1850	0,3208	0,1465	0,2381	0,1125	0,1736	0,0842	0,1247	0,0618	0,0884
3	0,2158	0,5366	0,1954	0,4335	0,1687	0,3423	0,1404	0,2650	0,1133	0,2017
4	0,1888	0,7254	0,1954	0,6288	0,1898	0,5321	0,1755	0,4405	0,1558	0,3575
5	0,1322	0,8576	0,1563	0,7851	0,1708	0,7029	0,1755	0,6160	0,1714	0,5289
6	0,0771	0,9347	0,1042	0,8893	0,1281	0,8311	0,1462	0,7622	0,1571	0,6860
7	0,0385	0,9733	0,0595	0,9489	0,0824	0,9134	0,1044	0,8666	0,1234	0,8095
8	0,0169	0,9901	0,0298	0,9786	0,0463	0,9597	0,0653	0,9319	0,0849	0,8944
9	0,0066	0,9967	0,0132	0,9919	0,0232	0,9829	0,0363	0,9682	0,0519	0,9462
10	0,0023	0,9990	0,0053	0,9972	0,0104	0,9933	0,0181	0,9863	0,0285	0,9747
11	0,0007	0,9997	0,0019	0,9991	0,0043	0,9976	0,0082	0,9945	0,0143	0,9890
12	0,0002	0,9999	0,0006	0,9997	0,0016	0,9992	0,0034	0,9980	0,0065	0,9955
13	0,0001	1,0000	0,0002	0,9999	0,0006	0,9997	0,0013	0,9993	0,0028	0,9983
14	0,0000	1,0000	0,0001	1,0000	0,0002	0,9999	0,0005	0,9998	0,0011	0,9994
15	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0002	0,9999	0,0004	0,9998
16	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	0,9999
17	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000
18	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000

m	6,00		7,00		8,00		9,00		10,00	
x	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)	f(x)	F(x)
0	0,0025	0,0025	0,0009	0,0009	0,0003	0,0003	0,0001	0,0001	0,0000	0,0000
1	0,0149	0,0174	0,0064	0,0073	0,0027	0,0030	0,0011	0,0012	0,0005	0,0005
2	0,0446	0,0620	0,0223	0,0296	0,0107	0,0138	0,0050	0,0062	0,0023	0,0028
3	0,0892	0,1512	0,0521	0,0818	0,0286	0,0424	0,0150	0,0212	0,0076	0,0103
4	0,1339	0,2851	0,0912	0,1730	0,0573	0,0996	0,0337	0,0550	0,0189	0,0293
5	0,1606	0,4457	0,1277	0,3007	0,0916	0,1912	0,0607	0,1157	0,0378	0,0671
6	0,1606	0,6063	0,1490	0,4497	0,1221	0,3134	0,0911	0,2068	0,0631	0,1301
7	0,1377	0,7440	0,1490	0,5987	0,1396	0,4530	0,1171	0,3239	0,0901	0,2202
8	0,1033	0,8472	0,1304	0,7291	0,1396	0,5925	0,1318	0,4557	0,1126	0,3328
9	0,0688	0,9161	0,1014	0,8305	0,1241	0,7166	0,1318	0,5874	0,1251	0,4579
10	0,0413	0,9574	0,0710	0,9015	0,0993	0,8159	0,1186	0,7060	0,1251	0,5830
11	0,0225	0,9799	0,0452	0,9467	0,0722	0,8881	0,0970	0,8030	0,1137	0,6968
12	0,0113	0,9912	0,0263	0,9730	0,0481	0,9362	0,0728	0,8758	0,0948	0,7916
13	0,0052	0,9964	0,0142	0,9872	0,0296	0,9658	0,0504	0,9261	0,0729	0,8645
14	0,0022	0,9986	0,0071	0,9943	0,0169	0,9827	0,0324	0,9585	0,0521	0,9165
15	0,0009	0,9995	0,0033	0,9976	0,0090	0,9918	0,0194	0,9780	0,0347	0,9513
16	0,0003	0,9998	0,0014	0,9990	0,0045	0,9963	0,0109	0,9889	0,0217	0,9730
17	0,0001	0,9999	0,0006	0,9996	0,0021	0,9984	0,0058	0,9947	0,0128	0,9857
18	0,0000	1,0000	0,0002	0,9999	0,0009	0,9993	0,0029	0,9976	0,0071	0,9928
19	0,0000	1,0000	0,0001	1,0000	0,0004	0,9997	0,0014	0,9989	0,0037	0,9965
20	0,0000	1,0000	0,0000	1,0000	0,0002	0,9999	0,0006	0,9996	0,0019	0,9984
21	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000	0,0003	0,9998	0,0009	0,9993
22	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	0,9999	0,0004	0,9997
23	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0002	0,9999
24	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000	1,0000	0,0001	1,0000

Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung $N(0,1)$ (Quelle: Hartung et al. 1982, S. 734)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

Ablesebeispiel: $\Phi(1{,}56) = 0{,}9406$; $\Phi(z) = 1 - \alpha$

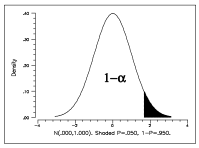
Erweiterung der Tafel: $\Phi(-z) = 1 - \Phi(z)$

Approximation nach Hastings für $z > 0$

$$\Phi(z) \cong 1 - \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{z^2}{2}} \cdot \left(a_1 \cdot t + a_2 \cdot t^2 + a_3 \cdot t^3 + a_4 \cdot t^4 + a_5 \cdot t^5\right) \text{ mit } t = \frac{1}{1 + b \cdot z},$$

$b = 0{,}2316419$, $a_1 = 0{,}31938153$, $a_2 = -0{,}356563782$, $a_3 = 1{,}781477937$,
$a_4 = -1{,}821255978$, $a_5 = 1{,}330274429$.

Quantile z der Standardnormalverteilung
$\mathbf{N(0,1)}$ (Quelle: Hartung et al. 1982, S. 735)



$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z
0,9999	3,7190	0,9975	2,8071	0,965	1,8119	0,83	0,9542
0,9998	3,5401	0,9970	2,7478	0,960	1,7507	0,82	0,9154
0,9997	3,4316	0,9965	2,6968	0,955	1,6954	0,81	0,8779
0,9996	3,3528	0,9960	2,6521	0,950	1,6449	0,80	0,8416
0,9995	3,2905	0,9955	2,6121	0,945	1,5982	0,79	0,8064
0,9994	3,2389	0,9950	2,5758	0,940	1,5548	0,78	0,7722
0,9993	3,1947	0,9945	2,5427	0,935	1,5141	0,76	0,7063
0,9992	3,1559	0,9940	2,5121	0,930	1,4758	0,74	0,6433
0,9991	3,1214	0,9935	2,4838	0,925	1,4395	0,72	0,5828
0,9990	3,0902	0,9930	2,4573	0,920	1,4051	0,70	0,5244
0,9989	3,0618	0,9925	2,4324	0,915	1,3722	0,68	0,4677
0,9988	3,0357	0,9920	2,4089	0,910	1,3408	0,66	0,4125
0,9987	3,0115	0,9915	2,3867	0,905	1,3106	0,64	0,3585
0,9986	2,9889	0,9910	2,3656	0,900	1,2816	0,62	0,3055
0,9985	2,9677	0,9905	2,3455	0,890	1,2265	0,60	0,2533
0,9984	2,9478	0,9900	2,3263	0,880	1,1750	0,58	0,2019
0,9983	2,9290	0,9850	2,1701	0,870	1,1264	0,56	0,1510
0,9982	2,9112	0,9800	2,0537	0,860	1,0803	0,54	0,1004
0,9981	2,8943	0,9750	1,9600	0,850	1,0364	0,52	0,0502
0,9980	2,8782	0,9700	1,8808	0,840	0,9945	0,50	0,0000

Ablesebeispiel: $z_{0,95} = 1,6449$

Erweiterung der Tafel: $z_{1-\alpha} = -z_{\alpha}$

Approximation nach Hastings für $0,5 < \Phi(z) < 1$

$$z \cong t - \frac{a_0 + a_1 \cdot t + a_2 \cdot t^2}{1 + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot t^3} \text{ mit } t = \sqrt{-2 \cdot \ln\left(1 - \Phi(z)\right)},$$

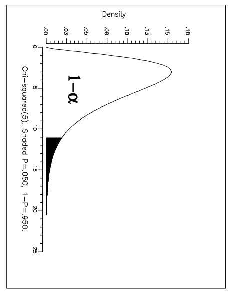
$a_0 = 2,515517$,	$a_1 = 0,802853$,	$a_2 = 0,010328$,
$b_1 = 1,432788$,	$b_2 = 0,189269$,	$b_3 = 0,001308$.

Häufig vorkommende z-Werte für die Bestimmung von Konfidenzintervallen

Z ist standardnormalverteilt		Konfidenzintervall zweiseitig	Konfidenzintervall einseitig
Häufig vorkommende z-Werte für:	$\alpha = 0,01$	$z = 2,58$	$z = 2,33$
	$\alpha = 0,05$	$z = 1,96$	$z = 1,65$
	$\alpha = 0,1$	$z = 1,65$	$z = 1,28$

χ^2-Verteilung für $P\left(O < \chi^2 \le \chi^2_\alpha\right) = 1 - \alpha$

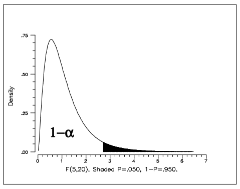
(Quelle: J. Schwarze: Grundlagen der Statistik II, 1993, S. 306 - 307):



ν	1 − α											
	0,001	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995	0,999
1	0,000	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879	10,828
2	0,002	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597	13,816
3	0,024	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838	16,266
4	0,091	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860	18,467
5	0,210	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750	20,515
6	0,381	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548	22,458
7	0,598	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278	24,322
8	0,857	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955	26,125
9	1,152	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589	27,877
10	1,479	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188	29,588
11	1,834	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757	31,264
12	2,214	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300	32,910
13	2,617	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819	34,528
14	3,041	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319	36,123
15	3,483	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801	37,697
16	3,942	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267	39,252
17	4,416	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718	40,790
18	4,905	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156	42,312
19	5,407	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582	43,820
20	5,921	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997	45,315
21	6,447	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401	46,797
22	6,983	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796	48,268
23	7,529	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181	49,728
24	8,085	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,558	51,179
25	8,649	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928	52,620

ν	$1-\alpha$											
	0,001	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995	0,999
26	9,222	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290	54,052
27	9,803	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645	55,476
28	10,391	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993	56,892
29	10,986	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336	58,301
30	11,588	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672	59,703
31	12,196	14,458	15,655	17,539	19,281	21,434	41,422	44,985	48,232	52,191	55,003	61,098
32	12,811	15,134	16,362	18,291	20,072	22,271	42,585	46,194	49,480	53,486	56,328	62,487
33	13,431	15,815	17,074	19,047	20,867	23,110	43,745	47,400	50,725	54,776	57,648	63,870
34	14,057	16,501	17,789	19,806	21,664	23,952	44,903	48,602	51,966	56,061	58,964	65,247
35	14,688	17,192	18,509	20,569	22,465	24,797	46,059	49,802	53,203	57,342	60,275	66,619
36	15,324	17,887	19,233	21,336	23,269	25,643	47,212	50,998	54,437	58,619	61,581	67,985
37	15,965	18,586	19,960	22,106	24,075	26,492	48,363	52,192	55,668	59,893	62,883	69,346
38	16,611	19,289	20,691	22,878	24,884	27,343	49,513	53,384	56,896	61,162	64,181	70,703
39	17,262	19,996	21,426	23,654	25,695	28,196	50,660	54,572	58,120	62,428	65,476	72,055
40	17,916	20,707	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691	66,766	73,402
41	18,575	21,421	22,906	25,215	27,326	29,907	52,949	56,942	60,561	64,950	68,053	74,745
42	19,239	22,138	23,650	25,999	28,144	30,765	54,090	58,124	61,777	66,206	69,336	76,084
43	19,906	22,859	24,398	26,785	28,965	31,625	55,230	59,303	62,990	67,459	70,616	77,419
44	20,576	23,584	25,148	27,575	29,787	32,487	56,369	60,481	64,201	68,710	71,893	78,750
45	21,251	24,311	25,901	28,366	30,612	33,350	57,505	61,656	65,410	69,957	73,166	80,077
46	21,929	25,041	26,657	29,160	31,439	34,215	58,641	62,830	66,617	71,201	74,437	81,400
47	22,610	25,775	27,416	29,956	32,268	35,081	59,774	64,001	67,821	72,443	75,704	82,720
48	23,295	26,511	28,177	30,755	33,098	35,949	60,907	65,171	69,023	73,683	76,969	84,037
49	23,983	27,249	28,941	31,555	33,930	36,818	62,038	66,339	70,222	74,919	78,231	85,351
50	24,674	27,991	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154	79,490	86,661

ν	$1-\alpha$											
	0,001	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995	0,999
51	25,368	28,735	30,475	33,162	35,600	38,560	64,295	68,669	72,616	77,386	80,747	87,968
52	26,065	29,481	31,246	33,968	36,437	39,433	65,422	69,832	73,810	78,616	82,001	89,272
53	26,765	30,230	32,018	34,776	37,276	40,308	66,548	70,993	75,002	79,843	83,253	90,573
54	27,468	30,981	32,793	35,586	38,116	41,183	67,673	72,153	76,192	81,069	84,502	91,872
55	28,173	31,735	33,570	36,398	38,958	42,060	68,796	73,311	77,380	82,292	85,749	93,168
56	28,881	32,490	34,350	37,212	39,801	42,937	69,919	74,468	78,567	83,513	86,994	94,461
57	29,592	33,248	35,131	38,027	40,646	43,816	71,040	75,624	79,752	84,733	88,236	95,751
58	30,305	34,008	35,913	38,844	41,492	44,696	72,160	76,778	80,936	85,950	89,477	97,039
59	31,020	34,770	36,698	39,662	42,339	45,577	73,279	77,931	82,117	87,166	90,715	98,324
60	31,738	35,534	37,485	40,482	43,188	46,459	74,397	79,082	83,298	88,379	91,952	99,607
61	32,459	36,300	38,273	41,303	44,038	47,342	75,514	80,232	84,476	89,591	93,186	100,888
62	33,181	37,068	39,063	42,126	44,889	48,226	76,630	81,381	85,654	90,802	94,419	102,166
63	33,906	37,838	39,855	42,950	45,741	49,111	77,745	82,529	86,830	92,010	95,649	103,442
64	34,633	38,610	40,649	43,776	46,595	49,996	78,860	83,675	88,004	93,217	96,878	104,716
65	35,362	39,383	41,444	44,603	47,450	50,883	79,973	84,821	89,177	94,422	98,105	105,988
66	36,093	40,158	42,240	45,431	48,305	51,770	81,085	85,965	90,349	95,626	99,330	107,258
67	36,826	40,935	43,038	46,261	49,162	52,659	82,197	87,108	91,519	96,828	100,554	108,526
68	37,561	41,713	43,838	47,092	50,020	53,548	83,308	88,250	92,689	98,028	101,776	109,791
69	38,298	42,494	44,639	47,924	50,879	54,438	84,418	89,391	93,856	99,228	102,996	111,055
70	39,036	43,275	45,442	48,758	51,739	55,329	85,527	90,531	95,023	100,425	104,215	112,317
75	42,757	47,206	49,475	52,942	56,054	59,795	91,061	96,217	100,839	106,393	110,286	118,599
80	46,520	51,172	53,540	57,153	60,391	64,278	96,578	101,879	106,629	112,329	116,321	124,839
85	50,320	55,170	57,634	61,389	64,749	68,777	102,079	107,522	112,393	118,236	122,325	131,041
90	54,155	59,196	61,754	65,647	69,126	73,291	107,565	113,145	118,136	124,116	128,299	137,208
95	58,022	63,250	65,898	69,925	73,520	77,818	113,038	118,752	123,858	129,973	134,247	143,344
100	61,918	67,328	70,065	74,222	77,929	82,358	118,498	124,342	129,561	135,807	140,169	149,449



F-Verteilung:

In Abhängigkeit von ν_1 und ν_2 sind die Werte x der F-verteilten Zufallsvariablen tabelliert, bei denen die Verteilungsfunktion $(1-\alpha)$ die Werte 0,90; 0,95; 0,975; 0,99 und 0,995 annimmt (Quelle: J. Schwarze, Grundlagen der Statistik II, 1993, S. 301 - 305)

$1-\alpha = 0{,}900$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	30	40	50	100	∞
1	39,86	49,50	53,59	55,83	57,24	58,20	59,14	59,69	60,12	60,46	60,98	61,36	61,64	61,86	62,04	62,58	62,85	63,01	63,33	63,33
2	8,526	9,000	9,162	9,244	9,293	9,326	9,350	9,367	9,381	9,392	9,409	9,421	9,430	9,437	9,442	9,459	9,467	9,472	9,482	9,491
3	5,538	5,462	5,390	5,343	5,309	5,285	5,266	5,251	5,240	5,230	5,215	5,204	5,196	5,189	5,184	5,167	5,159	5,154	5,143	5,134
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,895	3,877	3,864	3,853	3,844	3,817	3,803	3,795	3,777	3,761
5	4,060	3,780	3,619	3,520	3,453	3,404	3,368	3,339	3,316	3,297	3,268	3,247	3,230	3,217	3,206	3,174	3,157	3,147	3,126	3,105
6	3,776	3,463	3,289	3,181	3,107	3,055	3,014	2,983	2,958	2,937	2,905	2,881	2,863	2,848	2,836	2,800	2,781	2,770	2,746	2,722
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,668	2,643	2,623	2,607	2,595	2,555	2,535	2,523	2,497	2,471
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,502	2,475	2,454	2,438	2,425	2,383	2,361	2,348	2,321	2,293
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,379	2,351	2,329	2,312	2,298	2,255	2,232	2,218	2,189	2,159
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,284	2,255	2,233	2,215	2,201	2,155	2,132	2,117	2,087	2,055
11	3,225	2,859	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,209	2,179	2,156	2,138	2,123	2,076	2,052	2,036	2,005	1,972
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,147	2,117	2,094	2,075	2,060	2,011	1,986	1,970	1,938	1,904
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,097	2,066	2,042	2,023	2,007	1,958	1,931	1,915	1,882	1,846
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	2,054	2,022	1,998	1,978	1,962	1,912	1,885	1,869	1,834	1,797
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	2,017	1,985	1,961	1,941	1,924	1,873	1,845	1,828	1,793	1,755
16	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055	2,028	1,985	1,953	1,928	1,908	1,891	1,839	1,811	1,793	1,757	1,718
17	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028	2,001	1,958	1,925	1,900	1,879	1,862	1,809	1,781	1,763	1,726	1,686
18	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005	1,977	1,933	1,900	1,875	1,854	1,837	1,783	1,754	1,736	1,698	1,657
19	2,990	2,605	2,397	2,266	2,176	2,109	2,058	2,017	1,984	1,956	1,912	1,878	1,852	1,831	1,814	1,759	1,730	1,711	1,673	1,631
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,892	1,859	1,833	1,811	1,794	1,738	1,708	1,690	1,650	1,607
21	2,961	2,574	2,365	2,233	2,142	2,075	2,023	1,982	1,948	1,920	1,875	1,841	1,815	1,793	1,776	1,719	1,689	1,670	1,630	1,586
22	2,949	2,561	2,351	2,219	2,128	2,060	2,008	1,967	1,933	1,904	1,859	1,825	1,798	1,777	1,759	1,702	1,671	1,652	1,611	1,567
23	2,937	2,549	2,339	2,207	2,115	2,047	1,995	1,953	1,919	1,890	1,845	1,811	1,784	1,762	1,744	1,686	1,655	1,636	1,594	1,549
24	2,927	2,538	2,327	2,195	2,103	2,035	1,983	1,941	1,906	1,877	1,832	1,797	1,770	1,748	1,730	1,672	1,641	1,621	1,579	1,533
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895	1,866	1,820	1,785	1,758	1,736	1,718	1,659	1,627	1,607	1,565	1,518
26	2,909	2,519	2,307	2,174	2,082	2,014	1,961	1,919	1,884	1,855	1,809	1,774	1,747	1,724	1,706	1,647	1,615	1,594	1,551	1,504
27	2,901	2,511	2,299	2,165	2,073	2,005	1,951	1,909	1,874	1,845	1,799	1,764	1,736	1,714	1,695	1,636	1,603	1,583	1,539	1,491
28	2,894	2,503	2,291	2,157	2,064	1,996	1,943	1,900	1,865	1,836	1,790	1,754	1,726	1,704	1,685	1,625	1,592	1,572	1,528	1,478
29	2,887	2,495	2,283	2,149	2,057	1,988	1,934	1,892	1,857	1,827	1,781	1,745	1,717	1,695	1,676	1,616	1,583	1,562	1,517	1,467
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,773	1,737	1,709	1,686	1,667	1,606	1,573	1,552	1,507	1,456
40	2,835	2,440	2,226	2,091	1,997	1,927	1,872	1,829	1,793	1,763	1,715	1,678	1,649	1,625	1,605	1,541	1,506	1,483	1,434	1,377
50	2,809	2,412	2,197	2,061	1,966	1,895	1,840	1,796	1,760	1,729	1,680	1,643	1,613	1,588	1,568	1,502	1,465	1,441	1,388	1,327
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707	1,657	1,619	1,589	1,564	1,543	1,476	1,437	1,413	1,358	1,291
70	2,779	2,380	2,164	2,027	1,931	1,860	1,804	1,760	1,723	1,691	1,641	1,603	1,572	1,547	1,526	1,457	1,418	1,392	1,335	1,265
80	2,769	2,370	2,153	2,016	1,921	1,849	1,793	1,748	1,711	1,680	1,629	1,590	1,559	1,534	1,513	1,443	1,403	1,377	1,318	1,245
90	2,762	2,362	2,146	2,008	1,912	1,841	1,785	1,739	1,702	1,670	1,620	1,581	1,550	1,524	1,503	1,432	1,391	1,365	1,304	1,228
100	2,756	2,356	2,139	2,002	1,906	1,834	1,778	1,732	1,695	1,663	1,612	1,573	1,542	1,516	1,494	1,423	1,382	1,355	1,293	1,214
∞	2,706	2,303	2,084	1,945	1,847	1,774	1,717	1,670	1,632	1,599	1,546	1,505	1,471	1,444	1,421	1,342	1,295	1,263	1,185	1,000

	$1-\alpha = 0{,}950$																			
V_2	V_1: 1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	30	40	50	100	∞
1	161,4	199,5	215,8	224,7	230,4	234,2	237,0	239,1	240,8	242,1	244,2	245,6	246,8	247,6	248,3	250,4	251,5	252,1	253,4	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,44	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,69	8,67	8,66	8,62	8,59	8,58	8,54	8,53
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,912	5,873	5,844	5,821	5,803	5,746	5,717	5,699	5,658	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,678	4,636	4,604	4,579	4,558	4,496	4,464	4,444	4,403	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,206	4,147	4,099	4,060	4,000	3,956	3,922	3,896	3,874	3,808	3,773	3,753	3,711	3,669
7	5,591	4,737	4,347	4,120	3,971	3,866	3,787	3,726	3,677	3,636	3,574	3,529	3,494	3,467	3,444	3,375	3,340	3,318	3,274	3,230
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,284	3,237	3,202	3,173	3,150	3,079	3,042	3,020	2,974	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,073	3,025	2,989	2,960	2,936	2,864	2,826	2,803	2,755	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,913	2,865	2,828	2,798	2,774	2,699	2,661	2,637	2,588	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,788	2,739	2,701	2,671	2,646	2,570	2,531	2,507	2,456	2,404
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,687	2,637	2,599	2,568	2,544	2,466	2,426	2,401	2,350	2,296
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,604	2,554	2,515	2,484	2,459	2,380	2,339	2,314	2,261	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,534	2,484	2,445	2,413	2,388	2,308	2,266	2,240	2,187	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,475	2,424	2,385	2,353	2,328	2,247	2,204	2,178	2,123	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,425	2,373	2,333	2,302	2,276	2,194	2,151	2,124	2,068	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,381	2,329	2,289	2,257	2,230	2,148	2,104	2,077	2,020	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,342	2,290	2,250	2,217	2,191	2,107	2,063	2,035	1,978	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,308	2,256	2,215	2,182	2,155	2,071	2,026	1,999	1,940	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,278	2,225	2,184	2,151	2,124	2,039	1,994	1,966	1,907	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321	2,250	2,197	2,156	2,123	2,096	2,010	1,965	1,936	1,876	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,396	2,342	2,297	2,226	2,173	2,131	2,098	2,071	1,984	1,938	1,909	1,849	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,204	2,150	2,109	2,075	2,048	1,961	1,914	1,885	1,823	1,757
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,183	2,130	2,088	2,054	2,027	1,939	1,892	1,863	1,800	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,165	2,111	2,069	2,035	2,007	1,919	1,872	1,842	1,779	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,148	2,094	2,052	2,018	1,990	1,901	1,853	1,823	1,760	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,132	2,078	2,036	2,002	1,974	1,884	1,836	1,806	1,742	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,118	2,064	2,021	1,987	1,959	1,869	1,820	1,790	1,725	1,654
29	4,183	3,327	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,104	2,050	2,007	1,973	1,945	1,854	1,806	1,775	1,710	1,638
30	4,171	3,316	2,922	2,690	2,534	2,420	2,334	2,266	2,211	2,165	2,092	2,037	1,995	1,960	1,932	1,841	1,792	1,761	1,695	1,622
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,003	1,948	1,904	1,868	1,839	1,744	1,693	1,660	1,589	1,509
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026	1,952	1,895	1,850	1,814	1,784	1,687	1,634	1,599	1,525	1,438
60	4,001	3,150	2,758	2,525	2,368	2,254	2,166	2,097	2,040	1,993	1,917	1,860	1,815	1,778	1,748	1,649	1,594	1,559	1,481	1,389
70	3,978	3,127	2,735	2,503	2,346	2,231	2,143	2,074	2,017	1,969	1,893	1,836	1,790	1,753	1,722	1,622	1,566	1,530	1,450	1,353
80	3,960	3,110	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951	1,875	1,817	1,772	1,734	1,703	1,602	1,545	1,508	1,426	1,325
90	3,947	3,097	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938	1,861	1,803	1,757	1,720	1,688	1,586	1,528	1,491	1,407	1,302
100	3,936	3,087	2,695	2,463	2,305	2,191	2,102	2,032	1,975	1,927	1,850	1,792	1,746	1,708	1,676	1,573	1,515	1,477	1,392	1,283
∞	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,752	1,692	1,644	1,604	1,571	1,459	1,394	1,350	1,243	1,000

	$1-\alpha = 0{,}975$																			
ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	30	40	50	100	∞
1	647,5	799,5	864,3	899,8	922,0	937,3	948,5	956,9	963,5	968,9	977,0	982,8	987,2	990,6	993,4	1002	1006	1008	1014	1018
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,44	39,44	39,45	39,46	39,47	39,48	39,98	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,28	14,23	14,20	14,17	14,08	14,04	14,01	13,96	13,90
4	12,22	10,65	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,751	8,684	8,633	8,592	8,560	8,461	8,411	8,381	8,319	8,257
5	10,01	8,43	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,525	6,456	6,403	6,362	6,329	6,227	6,175	6,144	6,080	6,015
6	8,813	7,260	6,599	6,227	5,988	5,820	5,695	5,600	5,523	5,461	5,366	5,297	5,244	5,202	5,168	5,065	5,012	4,980	4,912	4,849
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,666	4,596	4,543	4,501	4,467	4,362	4,309	4,276	4,208	4,142
8	7,571	6,059	5,416	5,053	4,817	4,651	4,528	4,433	4,357	4,295	4,200	4,130	4,076	4,034	3,999	3,894	3,839	3,807	3,738	3,670
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,868	3,798	3,744	3,701	3,667	3,560	3,505	3,472	3,403	3,333
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,621	3,550	3,496	3,453	3,418	3,311	3,255	3,221	3,151	3,080
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,430	3,359	3,304	3,261	3,226	3,117	3,061	3,027	2,956	2,883
12	6,554	5,096	4,474	4,121	3,891	3,728	3,606	3,512	3,436	3,374	3,277	3,206	3,152	3,108	3,073	2,963	2,906	2,871	2,799	2,725
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,153	3,082	3,027	2,983	2,948	2,837	2,780	2,744	2,671	2,595
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,050	2,979	2,923	2,879	2,844	2,732	2,674	2,638	2,564	2,487
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	2,963	2,891	2,836	2,792	2,756	2,644	2,585	2,549	2,474	2,395
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049	2,986	2,889	2,817	2,761	2,717	2,681	2,568	2,508	2,472	2,396	2,316
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985	2,922	2,825	2,753	2,697	2,652	2,616	2,502	2,442	2,405	2,328	2,247
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929	2,866	2,769	2,696	2,640	2,596	2,559	2,444	2,384	2,347	2,269	2,187
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880	2,817	2,720	2,647	2,591	2,546	2,509	2,394	2,333	2,295	2,217	2,133
20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,676	2,603	2,547	2,501	2,464	2,349	2,287	2,249	2,170	2,085
21	5,827	4,420	3,819	3,475	3,250	3,089	2,969	2,874	2,798	2,735	2,637	2,564	2,507	2,462	2,425	2,308	2,246	2,208	2,128	2,042
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763	2,700	2,602	2,528	2,472	2,426	2,389	2,272	2,210	2,171	2,090	2,003
23	5,750	4,349	3,750	3,408	3,183	3,023	2,902	2,808	2,731	2,668	2,570	2,497	2,440	2,394	2,357	2,239	2,176	2,137	2,056	1,968
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703	2,640	2,541	2,468	2,411	2,365	2,327	2,209	2,146	2,107	2,024	1,935
25	5,686	4,291	3,694	3,353	3,129	2,968	2,848	2,753	2,677	2,613	2,515	2,441	2,384	2,338	2,300	2,182	2,118	2,079	1,996	1,906
26	5,659	4,265	3,670	3,329	3,105	2,945	2,824	2,729	2,653	2,590	2,491	2,417	2,360	2,314	2,276	2,157	2,093	2,053	1,969	1,878
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631	2,568	2,469	2,395	2,337	2,291	2,253	2,133	2,069	2,029	1,945	1,853
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611	2,547	2,448	2,374	2,317	2,270	2,232	2,112	2,048	2,007	1,922	1,829
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592	2,529	2,430	2,355	2,298	2,251	2,213	2,092	2,028	1,987	1,901	1,807
30	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,575	2,511	2,412	2,338	2,280	2,233	2,195	2,074	2,009	1,968	1,882	1,787
40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452	2,388	2,288	2,213	2,154	2,107	2,068	1,943	1,875	1,832	1,741	1,637
50	5,340	3,975	3,390	3,054	2,833	2,673	2,553	2,458	2,381	2,317	2,216	2,140	2,081	2,033	1,993	1,866	1,796	1,752	1,656	1,545
60	5,286	3,925	3,342	3,008	2,786	2,627	2,507	2,412	2,334	2,270	2,169	2,093	2,033	1,985	1,944	1,815	1,744	1,699	1,599	1,482
70	5,247	3,890	3,309	2,975	2,754	2,595	2,474	2,379	2,302	2,237	2,136	2,059	1,999	1,950	1,910	1,779	1,707	1,660	1,558	1,436
80	5,218	3,864	3,284	2,950	2,730	2,571	2,450	2,355	2,277	2,213	2,111	2,035	1,974	1,925	1,884	1,752	1,679	1,632	1,527	1,400
90	5,196	3,844	3,265	2,932	2,711	2,552	2,431	2,336	2,259	2,194	2,092	2,015	1,955	1,905	1,864	1,731	1,657	1,610	1,503	1,371
100	5,179	3,828	3,249	2,917	2,696	2,537	2,417	2,321	2,244	2,179	2,077	2,000	1,939	1,890	1,849	1,715	1,640	1,592	1,483	1,347
∞	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114	2,048	1,945	1,866	1,803	1,751	1,708	1,566	1,484	1,428	1,296	1,000

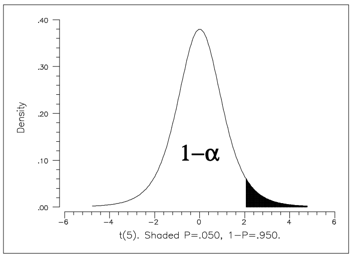
v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	30	40	50	100	∞
	$1-\alpha = 0{,}990$																			
1	4052	5000	5403	5625	5764	5859	5929	5981	6023	6056	6107	6143	6170	6192	6209	6261	6287	6303	6334	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,75	99,78	99,80	99,83	99,86	99,88	99,89	99,90	99,93	99,95	99,96	99,98	99,50
3	34,11	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,92	26,83	26,75	26,69	26,50	26,41	26,35	26,24	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,25	14,15	14,08	14,02	13,84	13,75	13,69	13,58	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,888	9,770	9,680	9,610	9,553	9,379	9,291	9,238	9,130	9,020
6	13,74	10,92	9,779	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,718	7,605	7,519	7,451	7,396	7,229	7,143	7,091	6,987	6,880
7	12,25	9,55	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,469	6,359	6,275	6,209	6,155	5,992	5,908	5,858	5,751	5,650
8	11,26	8,65	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,667	5,559	5,477	5,412	5,359	5,198	5,116	5,065	4,961	4,859
9	10,56	8,02	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257	5,111	5,005	4,924	4,860	4,808	4,649	4,567	4,516	4,414	4,311
10	10,04	7,56	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,706	4,601	4,520	4,457	4,405	4,246	4,165	4,115	4,013	3,909
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,631	4,539	4,397	4,293	4,213	4,150	4,099	3,941	3,859	3,809	3,707	3,602
12	9,330	6,927	5,952	5,412	5,064	4,820	4,639	4,499	4,387	4,296	4,155	4,052	3,972	3,909	3,858	3,701	3,619	3,569	3,466	3,361
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	3,960	3,857	3,778	3,715	3,664	3,507	3,425	3,375	3,272	3,165
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,800	3,697	3,619	3,556	3,505	3,347	3,266	3,215	3,112	3,004
15	8,683	6,359	5,417	4,893	4,556	4,318	4,141	4,004	3,895	3,805	3,666	3,564	3,485	3,423	3,372	3,214	3,132	3,081	2,977	2,868
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780	3,691	3,553	3,451	3,372	3,310	3,259	3,101	3,018	2,967	2,863	2,753
17	8,400	6,112	5,185	4,669	4,336	4,101	3,927	3,791	3,682	3,593	3,455	3,353	3,275	3,212	3,161	3,003	2,920	2,869	2,764	2,653
18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597	3,508	3,371	3,269	3,190	3,128	3,077	2,918	2,835	2,784	2,678	2,566
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,522	3,434	3,297	3,195	3,116	3,054	3,003	2,844	2,761	2,709	2,602	2,489
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,231	3,130	3,051	2,989	2,938	2,778	2,695	2,643	2,535	2,421
21	8,016	5,780	4,874	4,369	4,042	3,812	3,640	3,506	3,398	3,310	3,173	3,071	2,993	2,931	2,880	2,720	2,636	2,584	2,475	2,360
22	7,945	5,719	4,817	4,313	3,988	3,758	3,587	3,453	3,346	3,258	3,121	3,019	2,941	2,879	2,827	2,667	2,583	2,531	2,422	2,305
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299	3,211	3,074	2,973	2,894	2,832	2,780	2,620	2,535	2,483	2,373	2,256
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256	3,168	3,032	2,930	2,852	2,789	2,738	2,577	2,492	2,439	2,329	2,211
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129	2,993	2,892	2,813	2,751	2,699	2,538	2,453	2,400	2,289	2,169
26	7,721	5,526	4,636	4,140	3,818	3,591	3,421	3,288	3,182	3,094	2,958	2,857	2,778	2,715	2,664	2,503	2,417	2,364	2,252	2,131
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149	3,062	2,926	2,824	2,746	2,683	2,632	2,470	2,384	2,330	2,218	2,097
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120	3,032	2,896	2,795	2,716	2,653	2,602	2,440	2,354	2,300	2,187	2,064
29	7,598	5,420	4,538	4,045	3,725	3,499	3,330	3,198	3,092	3,005	2,868	2,767	2,689	2,626	2,574	2,412	2,325	2,271	2,158	2,034
30	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,067	2,979	2,843	2,742	2,663	2,600	2,549	2,386	2,299	2,245	2,131	2,006
40	7,314	5,179	4,312	3,828	3,514	3,291	3,124	2,993	2,888	2,801	2,665	2,563	2,484	2,421	2,369	2,203	2,114	2,058	1,938	1,805
50	7,170	5,056	4,199	3,720	3,408	3,186	3,020	2,890	2,785	2,698	2,563	2,461	2,382	2,318	2,265	2,098	2,007	1,949	1,825	1,683
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,496	2,394	2,315	2,251	2,198	2,028	1,936	1,877	1,749	1,601
70	7,011	4,922	4,074	3,600	3,291	3,071	2,906	2,777	2,672	2,585	2,450	2,348	2,268	2,204	2,150	1,980	1,886	1,826	1,695	1,540
80	6,962	4,881	4,036	3,563	3,255	3,036	2,871	2,742	2,637	2,551	2,415	2,313	2,233	2,169	2,115	1,944	1,849	1,788	1,655	1,494
90	6,925	4,849	4,007	3,535	3,228	3,009	2,844	2,715	2,611	2,524	2,389	2,286	2,206	2,142	2,088	1,916	1,820	1,759	1,623	1,457
100	6,895	4,824	3,983	3,513	3,206	2,988	2,823	2,694	2,590	2,503	2,368	2,265	2,185	2,120	2,067	1,893	1,797	1,735	1,598	1,427
∞	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407	2,321	2,185	2,082	2,000	1,934	1,878	1,696	1,592	1,523	1,358	1,000

$1-\alpha = 0{,}995$

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	30	40	50	100	∞
1	16210	20000	21615	22500	23056	23437	23715	23926	24091	24225	24427	24572	24682	24768	24836	25044	25149	25211	25338	25465
2	198,5	199,0	199,3	199,5	199,6	199,7	199,7	199,8	199,8	199,8	199,8	199,9	199,9	199,9	199,9	199,9	200,0	200,0	200,0	199,5
3	55,55	49,80	47,46	46,19	45,39	44,84	44,43	44,12	43,88	43,68	43,38	43,17	43,00	42,88	42,77	42,46	42,30	42,21	42,02	41,83
4	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	21,14	20,97	20,70	20,51	20,37	20,26	20,17	19,89	19,75	19,67	19,50	19,32
5	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,77	13,62	13,38	13,21	13,09	12,98	12,90	12,66	12,53	12,45	12,30	12,14
6	18,635	14,544	12,916	12,028	11,464	11,073	10,786	10,566	10,391	10,250	10,034	9,877	9,758	9,664	9,589	9,358	9,241	9,170	9,026	8,879
7	16,236	12,404	10,882	10,050	9,522	9,155	8,885	8,678	8,514	8,380	8,176	8,028	7,915	7,826	7,754	7,534	7,422	7,354	7,210	7,076
8	14,688	11,042	9,596	8,805	8,302	7,952	7,694	7,496	7,339	7,211	7,015	6,872	6,763	6,678	6,608	6,396	6,288	6,222	6,084	5,951
9	13,614	10,107	8,717	7,956	7,471	7,134	6,885	6,693	6,541	6,417	6,227	6,089	5,983	5,899	5,832	5,625	5,519	5,452	5,321	5,188
10	12,826	9,427	8,081	7,343	6,872	6,545	6,302	6,116	5,968	5,847	5,661	5,526	5,422	5,340	5,274	5,071	4,965	4,901	4,771	4,639
11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,537	5,418	5,236	5,103	5,001	4,921	4,855	4,654	4,550	4,487	4,358	4,226
12	11,754	8,510	7,226	6,521	6,071	5,757	5,525	5,345	5,202	5,085	4,906	4,774	4,674	4,594	4,530	4,331	4,228	4,165	4,036	3,904
13	11,373	8,186	6,926	6,233	5,791	5,482	5,253	5,076	4,935	4,820	4,643	4,513	4,413	4,334	4,270	4,073	3,970	3,908	3,779	3,647
14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,856	4,717	4,603	4,428	4,299	4,200	4,122	4,058	3,862	3,760	3,697	3,569	3,436
15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,536	4,423	4,250	4,122	4,024	3,946	3,882	3,687	3,585	3,522	3,394	3,260
16	10,575	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,384	4,272	4,099	3,972	3,875	3,797	3,734	3,539	3,437	3,375	3,246	3,112
17	10,384	7,354	6,155	5,497	5,075	4,779	4,559	4,389	4,253	4,142	3,971	3,844	3,747	3,670	3,607	3,412	3,311	3,248	3,119	2,984
18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,141	4,030	3,860	3,734	3,637	3,560	3,498	3,303	3,201	3,139	3,009	2,873
19	10,073	7,093	5,916	5,268	4,853	4,561	4,345	4,177	4,043	3,933	3,763	3,638	3,541	3,464	3,402	3,208	3,106	3,043	2,913	2,776
20	9,944	6,986	5,818	5,174	4,762	4,472	4,257	4,090	3,956	3,847	3,678	3,553	3,457	3,380	3,318	3,123	3,021	2,959	2,828	2,690
21	9,830	6,891	5,730	5,091	4,681	4,393	4,179	4,013	3,880	3,771	3,602	3,478	3,382	3,305	3,243	3,049	2,947	2,884	2,753	2,614
22	9,727	6,806	5,652	5,017	4,609	4,322	4,109	3,944	3,812	3,703	3,535	3,411	3,315	3,239	3,176	2,982	2,880	2,817	2,685	2,546
23	9,635	6,730	5,582	4,950	4,544	4,259	4,047	3,882	3,750	3,642	3,475	3,351	3,255	3,179	3,116	2,922	2,820	2,756	2,624	2,484
24	9,551	6,661	5,519	4,890	4,486	4,202	3,990	3,826	3,695	3,587	3,420	3,296	3,201	3,125	3,062	2,868	2,765	2,702	2,569	2,428
25	9,475	6,598	5,461	4,835	4,433	4,150	3,939	3,776	3,645	3,537	3,370	3,247	3,151	3,075	3,013	2,819	2,716	2,652	2,519	2,377
26	9,406	6,541	5,409	4,785	4,384	4,103	3,893	3,730	3,599	3,492	3,325	3,202	3,107	3,031	2,968	2,774	2,671	2,607	2,473	2,330
27	9,342	6,488	5,361	4,740	4,340	4,059	3,850	3,687	3,557	3,450	3,284	3,161	3,066	2,990	2,928	2,733	2,630	2,565	2,431	2,287
28	9,284	6,440	5,317	4,698	4,300	4,020	3,811	3,649	3,519	3,412	3,246	3,123	3,028	2,952	2,890	2,695	2,592	2,527	2,392	2,247
29	9,230	6,396	5,276	4,659	4,262	3,983	3,775	3,613	3,483	3,377	3,211	3,088	2,993	2,917	2,855	2,660	2,557	2,492	2,357	2,210
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,450	3,344	3,179	3,056	2,961	2,885	2,823	2,628	2,524	2,459	2,323	2,176
40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,222	3,117	2,953	2,831	2,737	2,661	2,598	2,401	2,296	2,230	2,088	1,932
50	8,626	5,902	4,826	4,232	3,848	3,578	3,376	3,219	3,092	2,988	2,825	2,703	2,609	2,533	2,470	2,272	2,164	2,097	1,951	1,786
60	8,494	5,795	4,729	4,140	3,760	3,492	3,291	3,134	3,008	2,904	2,742	2,620	2,526	2,450	2,387	2,187	2,079	2,010	1,861	1,689
70	8,402	5,720	4,661	4,076	3,698	3,431	3,232	3,075	2,950	2,846	2,684	2,563	2,468	2,392	2,329	2,128	2,019	1,949	1,797	1,618
80	8,334	5,665	4,611	4,029	3,652	3,387	3,188	3,032	2,907	2,803	2,641	2,520	2,425	2,349	2,286	2,084	1,974	1,903	1,748	1,563
90	8,282	5,623	4,573	3,992	3,617	3,352	3,154	2,999	2,874	2,770	2,609	2,487	2,393	2,316	2,253	2,051	1,939	1,868	1,711	1,520
100	8,240	5,589	4,542	3,963	3,589	3,325	3,127	2,972	2,847	2,744	2,583	2,461	2,367	2,290	2,227	2,024	1,912	1,840	1,681	1,485
∞	7,879	5,298	4,279	3,715	3,350	3,091	2,897	2,744	2,621	2,519	2,358	2,237	2,142	2,064	2,000	1,789	1,669	1,590	1,402	1,000

STUDENT (t)-Verteilung:

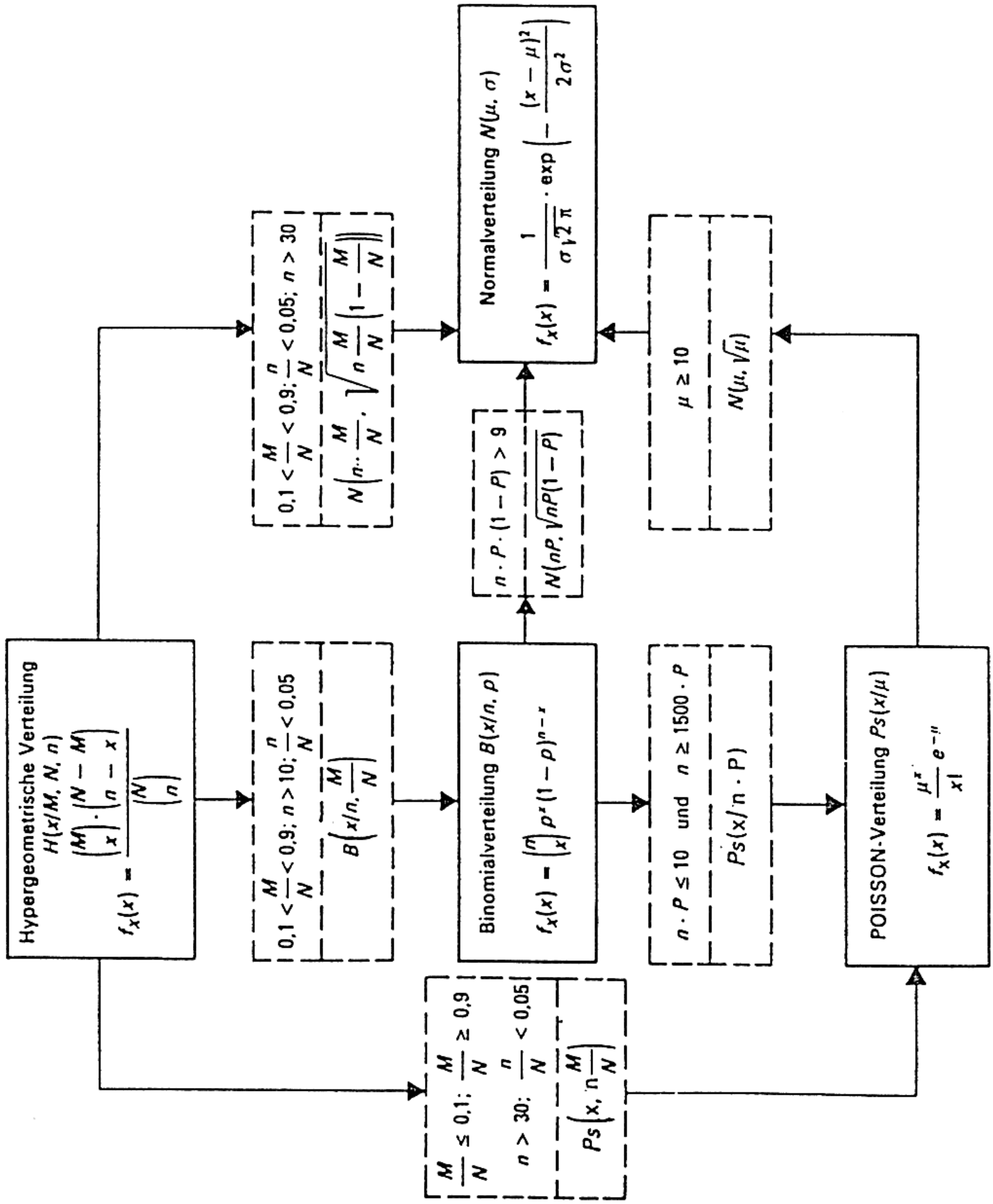
Für verschiedene Freiheitsgrade ν und für verschiedene Wahrscheinlichkeiten $P(T < t) = 1 - \alpha$ sind die entsprechenden t-Werte tabelliert (Quelle: J. Schwarze, Grundlagen der Statistik II, 1993, S. 308 - 309)



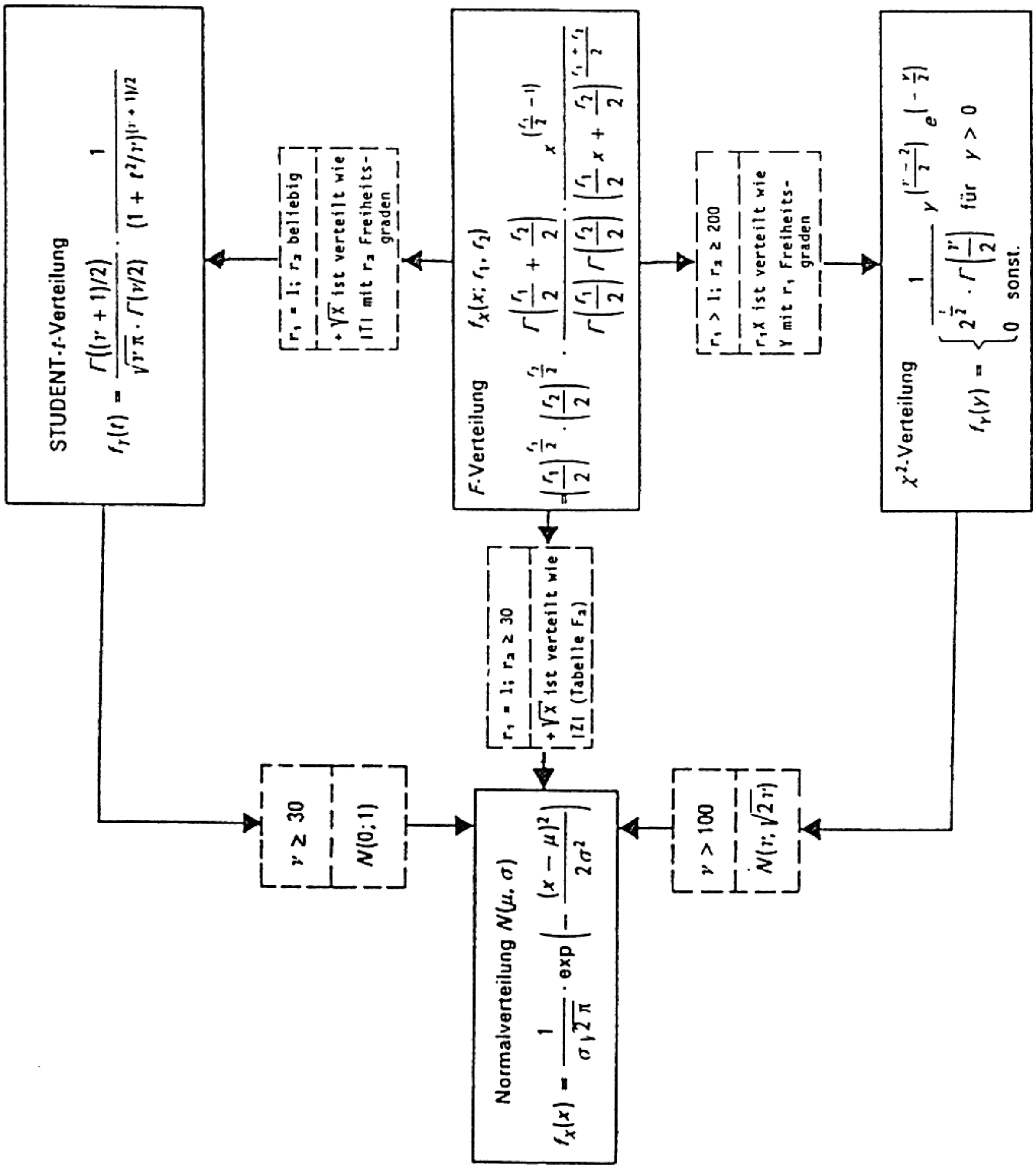
ν	1 − α: 0,6000	0,7000	0,8000	0,9000	0,9500	0,9750	0,9900	0,9950	0,9990	0,9995
1	0,3249	0,7265	1,3764	3,0777	6,3137	12,7062	31,8205	63,6568	318,3127	636,5894
2	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9646	9,9248	22,3273	31,5983
3	0,2767	0,5844	0,9785	1,6377	2,3534	3,1825	4,5407	5,8409	10,2146	12,9238
4	0,2707	0,5686	0,9410	1,5332	2,1318	2,7763	3,7470	4,6041	7,1732	8,6102
5	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3648	4,0322	5,8934	6,8688
6	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1426	3,7074	5,2076	5,9588
7	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,7851	5,4079
8	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5007	5,0411
9	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2968	4,7808
10	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5868
11	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247	4,4369
12	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
14	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1404
15	0,2579	0,5357	0,8662	1,3406	1,7530	2,1314	2,6025	2,9467	3,7328	4,0727
16	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193
22	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251

ν	$1-\alpha$									
	0,6000	0,7000	0,8000	0,9000	0,9500	0,9750	0,9900	0,9950	0,9990	0,9995
26	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6459
31	0,2555	0,5298	0,8534	1,3095	1,6955	2,0395	2,4528	2,7440	3,3749	3,6334
32	0,2555	0,5297	0,8530	1,3086	1,6939	2,0369	2,4487	2,7385	3,3653	3,6218
33	0,2554	0,5295	0,8526	1,3077	1,6924	2,0345	2,4448	2,7333	3,3563	3,6109
34	0,2553	0,5294	0,8523	1,3070	1,6909	2,0322	2,4411	2,7284	3,3479	3,6007
35	0,2553	0,5292	0,8520	1,3062	1,6896	2,0301	2,4377	2,7238	3,3400	3,5911
36	0,2552	0,5291	0,8517	1,3055	1,6883	2,0281	2,4345	2,7195	3,3326	3,5821
37	0,2552	0,5289	0,8514	1,3049	1,6871	2,0262	2,4314	2,7154	3,3256	3,5737
38	0,2551	0,5288	0,8512	1,3042	1,6860	2,0244	2,4286	2,7116	3,3190	3,5657
39	0,2551	0,5287	0,8509	1,3036	1,6849	2,0227	2,4258	2,7079	3,3128	3,5581
40	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5509
41	0,2550	0,5285	0,8505	1,3025	1,6829	2,0195	2,4208	2,7012	3,3013	3,5442
42	0,2550	0,5284	0,8503	1,3020	1,6820	2,0181	2,4185	2,6981	3,2960	3,5377
43	0,2549	0,5283	0,8501	1,3016	1,6811	2,0167	2,4163	2,6951	3,2909	3,5316
44	0,2549	0,5282	0,8499	1,3011	1,6802	2,0154	2,4141	2,6923	3,2861	3,5258
45	0,2549	0,5281	0,8497	1,3006	1,6794	2,0141	2,4121	2,6896	3,2815	3,5202
46	0,2548	0,5281	0,8495	1,3002	1,6787	2,0129	2,4102	2,6870	3,2771	3,5149
47	0,2548	0,5280	0,8493	1,2998	1,6779	2,0117	2,4083	2,6846	3,2729	3,5099
48	0,2548	0,5279	0,8492	1,2994	1,6772	2,0106	2,4066	2,6822	3,2689	3,5051
49	0,2547	0,5278	0,8490	1,2991	1,6765	2,0096	2,4049	2,6800	3,2651	3,5004
50	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	0,2545	0,5272	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	0,2543	0,5268	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	0,2542	0,5265	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	0,2541	0,5263	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019
100	0,2540	0,5261	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1737	3,3905

Übersicht zur Approximation von Verteilungen mit Approximationskriterien (Quelle: J. Schwarze, Grundlagen der Statistik II, 1993, S. 116)



Übersicht zur Approximation von Verteilungen mit Approximationskriterien (Quelle: J. Schwarze, Grundlagen der Statistik II, 1993, S. 117)



Anhang II: Verzeichnis wichtiger Symbole

Übersicht über wichtige Symbole und das griechische Alphabet

Symbol	**Bedeutung**
$\Phi(x)$	Verteilungsfunktion der Standardnormalverteilung N(0,1)
$\varphi(x)$	Dichtefunktion der Standardnormalverteilung N(0,1)
$\sim$	„verteilt nach“, z.B. X ~ N(0,1) = „X ist standardnormalvetrteilt“
μ, *auch* μ_X *oder* $E(X)$	Erwartungswert der Zufallsvariablen X
σ^2, *auch* σ_X^2 *oder* $Var(X)$	Varianz der Zufallsvariablen X
$Cov(X,Y)$	Covarianz der Zufallsvariablen X und Y
$N(\mu,\sigma^2)$	Normalverteilung mit Mittelwert μ und Varianz σ^2
$N(0,1)$	Standardnormalverteilung mit Mittelwert 0 und Varianz 1
t_υ	(Student-) t-Verteilung mit υ Freiheitsgraden
χ_υ^2	Chiquadrat-Verteilung mit υ Freiheitsgraden
$F_{\upsilon_1,\upsilon_2}$	F-Verteilung mit υ_1 Zähler – und υ_2 Nennerfreiheitsgraden
$B(n, p)$	Binomialverteilung mit den Parametern n (Anzahl Versuche) und p (Wahrscheinlichkeit „Treffer“)
G, *auch* Ω	Ereignisraum (Menge aller möglichen Ergebnisse eines Zufallsexperiments)
$P(A)$	Wahrscheinlichkeit für den Eintritt des Erignisses A
$P(A \vert B)$	Bedingte Wahrscheinlichkeit für den Eintritt des Ereignisses A unter der Bedingung des Eintretens des Ereignisses B
$F(x)$	Verteilungsfunktion der Zufallsvariablen X
$f(x)$	Dichtefunktion der Zufallsvariablen X
$\rho_{x,y}$	Theoretische Korrelation der Zufallsvariablen X und Y
$\overline{x}$	aritmetisches Mittel (aus der Stichprobe)
s^2, s_x^2	empirische Varianz (aus der Stichprobe)
$s_{x,y}$	empirische Covarianz (aus der Stichprobe)
$r_{x,y}$	empirische Korrelation (aus der Stichprobe)
R^2	Bestimmtheitsmaß (der Regression)
ML	Maximum Likelihood
$F(x,y)$	gemeinsame Verteilungsfunktion der Zufallsvariablen X und Y
$f(x,y)$	gemeinsame Dichtefunktion der der Zufallsvariablen X und Y
$\ln x$	Natürlicher Logarithmus (zur Basis e = 2,7182818…)
$\log x$	Dekadische Logarithmus (zur Basis 10)

$e^x, \exp(x)$	e-Funktion , Exponentialfunktion zur Basis e
$n!$	„n Fakultät": n! = 1*2*…*(n-1)*n
$\binom{n}{k}$	Binomialkoeffizient („n über k") = $\frac{n!}{k!(n-k)!}$
$\lim\limits_{x\to\infty}$	Grenzwert (Limes) wenn x gegen unendlich geht
$a_n \underset{n\to\infty}{\to} a$	„a konvergiert gegen a, wenn n gegen unendlich läuft"
min	Minimum, kleinster Wert einer Menge
max	Maximum, größter Wert einer Menge
$\lvert x\rvert$	Absolutwert, Betrag von x
$\sum\limits_i x_i, \sum_i x_i$	Summe von x über alle i = $x_1 + x_2 + ... + x_{n-1} + x_n$
$\sum\limits_i \sum\limits_j x_{ij}, \sum_i \sum_j x_{ij}$	Doppelsumme über x und y = $x_{11} + x_{12} + ... + x_{1m} + x_{21} + x_{22} + ... + x_{2m} + ... + x_{n1} + x_{n2} + ... + x_{nm}$
$\prod\limits_i x_i$	Produkt von x über alle i = $x_1 * x_2 * ... * x_n$

Rechenzeichen

$=$	„gleich"
$\approx$	„ungefähr gleich", „rund"
$\triangleq$	„entspricht"
$\leq$	„kleiner gleich"
$\geq$	„größer gleich"
$\in$	Element von; z.B. $3 \in \{1;2;3;4\}$
$\notin$	ist nicht Element von; z.B. $3 \notin \{1;2;4\}$
$\subset$	„Teilmenge von" $\{1;3\} \subset \{1;2;3;4\}$
$\emptyset$	„unmögliches Ereignis", „leere Menge"
$\cup$	„vereinigt mit"
$\cap$	„geschnitten mit"

Griechisches Alphabet

$\mathrm{A}\alpha$	Alpha
$\mathrm{B}\beta$	Beta
$\Gamma\gamma$	Gamma
$\Delta\delta$	Delta
$\mathrm{E}\varepsilon$	Epsilon
$\mathrm{Z}\zeta$	Zeta
$\mathrm{H}\eta$	Eta
$\Theta\theta$	Theta
$\mathrm{I}\iota$	Jota
$\mathrm{K}\kappa$	Kappa
$\Lambda\lambda$	Lambda
$\mathrm{M}\mu$	My
$\mathrm{N}\nu$	Ny

$\Xi\xi$	Xi
$\mathrm{O}o$	Omikron
$\Lambda\pi$	Pi
$\mathrm{P}\rho$	Rho
$\Sigma\sigma$	Sigma
$\mathrm{T}\tau$	Tau
$\Upsilon\upsilon$	Ypsilon
$\Phi\phi\varphi$	Phi
$\mathrm{X}\chi$	Chi
$\Psi\psi$	Psi
$\Omega\omega$	Omega

Literaturverzeichnis

Ausgewählte Literatur, Weiterstöbern empfohlen

A EINIGE STANDARDWERKE

Anderson, Davi, R., Sweeney, Dennis, J., Williams, Thomas, A., Freeman, Jim und Essie Shoesmith (2007), Statistics for Business and Economics, Thomson Publisher, London (mit CD)

Anderson, O., Popp, W., Schaffranek, M., Steinmetz, D. und H. Stenger (1997), Schätzen und Testen, Springer Verlag, 2. Auflage, Berlin

Bamberg, G. und F. Baur (2002), Statistik, R. Oldenbourg Verlag, 12. Auflage, München

Basler, H. (1994), Grundbegriffe der Wahrscheinlichkeitsrechnung und statistische Methodenlehre, Physica Verlag, 11. Auflage, Heidelberg

Bleymüller, J., Gehlert, G. und H. Gülicher (2004), Statistik für Wirtschaftswissenschaftler, Vahlen, 14. Auflage, München

Bomsdorf, E. (2002), Induktive Statistik – Eine Einführung, Verlag Josel Eul Verlag, 8. Auflage, Bergisch Gladbach/Köln

Bortz, I. (1999), Statistik für Sozialwissenschaftler, 5. Auflage, Berlin-New York

Bosch, K. (1998), Statistik-Taschenbuch, R. Oldenbourg Verlag, 3. Auflage, München/Wien

Conover, W. J. (1998), Practical Nonparametric Statistics, John Wiley & Sons, 3. Auflage, New York

Dillmann, R. (1990), Statistik II: Induktive Statistik, Physica Verlag, Heidelberg

Edwards, A. W. F.(1972), Likelihood – An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference, Cambridge University Press, Cambridge/Melbourne

Gale, W. A. (1986), Artificial Intelligence and Statistics, Addison-Wesley Publishing Company, Reading (Mass.)/San Juan

Gonick, L. und Smith, W. (1993), The Cartoon Guide to Statistics, New York

Greene, W. H. (2003), Econometric Analysis, Prentice Hall, 5. Auflage, New York/London

Greene, W. H. (1992), ET – The Econometrics Toolkit, Version 3.0, User's Guide, New York

Hansen, G. (1985), Methodenlehre der Statistik, Vahlen, 3. Auflage, München

Hochstädter, D. (1996), Statistische Methodenlehre, Verlag Harri Deutsch, 8. Auflage, Frankfurt am Main

Hujer, R. (1991), Statistik – Manuskript zur Vorlesung, Frankfurt

Kreyszig, E. (1979), Statistische Methoden und ihre Anwendungen, Vandenhoeck & Ruprecht, 7. Auflage, Göttingen

Kriz, J. (1983), Statistik in den Sozialwissenschaften, 4. Auflage, studium rororo, Reinbek bei Hamburg

Leiner, B. (1994), Stichprobentheorie, R. Oldenbourg Verlag, 3. Auflage, München

Malinvaud, E. (2000), Statistical Methods in Econometrics, 3rd ed., American Elsevier, New York

Merz, J. (2015), Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik, Skriptum zur Vorlesung, 11. Auflage, Lüneburg

Pfanzagl, J. (1974), Allgemeine Methodenlehre der Statistik – Teil II, Walter de Gruyter, Berlin/New York

Rinne, H. und G. Ickler (1986), Grundstudium Statistik, Verlag für Wirtschaftsskripten, 2. Auflage, München

Rüger, B. (2000), Induktive Statistik: Einführung für Wirtschafts- und Sozialwissenschaftler, R. Oldenbourg Verlag, München

Sachs, L (2004), Angewandte Statistik – Anwendung statistischer Methoden, Springer Verlag, 11. Auflage, Berlin

Sahner, H. (2002), Schließende Statistik, VS Verlag für Sozialwissenschaften, 5. Auflage, Stuttgart

Schaich, E. und A. Hamerle (1984), Verteilungsfreie statistische Prüfverfahren - Eine anwendungsorientierte Darstellung, Springer, Berlin

Scharnbacher, K. (2004), Statistik im Betrieb – Lehrbuch mit praktischen Beispielen, Gabler Verlag, 14. Auflage, Wiesbaden

Schira, J. (2009), Statistische Methoden der VWL und BWL Theorie und Praxis, 3. aktualisierte Auflage, Pearson Studium, München

Schwarze, J. (1977), Bibliographie zur Statistik in der Weiterbildung, Pädagogische Arbeitsstelle des Deutschen Volkshochschulverbandes, Holzhausenstr. 21, 6000 Frankfurt, 2. Auflage, Frankfurt/Bonn

Schwarze, J. (2009), Grundlagen der Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik, Verlag NWB, 9. Auflage, Herne/Berlin

Siegel, S. (1987), Nichtparametrische Statistische Methoden, Methoden in der Psychologie, Band 4, Fachbuchhandlung für Psychologie, 3. Auflage, Eschborn (Übersetzung aus dem amerikanischen von S. Siegel (1956), Nonparametric Statistics for the Behavioral Sciences, Mc Graw-Hill, New York)

Spanos, A. (1986), Statistical Foundations of Econometric Modelling, Cambridge University Press, Cambridge/Sydney

Wetzel, W. (1973), Statistische Grundausbildung für Wirtschaftswissenschaftler – Teil II, Walter de Gruyter, Berlin

Yamane, T. (1987), Statistik – Ein einführendes Lehrbuch, Fischer Taschenbuch Verlag, Frankfurt, Teil I und II

Yang, M. C. K. und D. Robinson (1986), Understanding and Learning Statistics by Computer, World Scientific, Singapore

B BÜCHER MIT ÜBUNGSAUFGABEN

Bamberg, G. und F. Baur (2004), Statistik Arbeitsbuch: Übungsaufgaben, Fallstudien, Lösungen, R. Oldenbourg Verlag, 7. Auflage, München

Bihn, W. R. und K. A. Schäffer (1986), Übungsaufgaben zur Grundausbildung in Statistik für Wirtschaftswissenschaftler, J.C. Witsch Nachf., Köln

Hartung, J., Elpelt, B. und K.-H. Klösner (2002), Statistik – Lehrbuch mit Übungsaufgaben, R. Oldenbourg Verlag, München

Hartung, J. und B. Heine (2004), Statistik Übungen: Induktive Statistik, R. Oldenbourg Verlag, 4. Auflage, München

Hochstädter, D. (1993), Aufgaben mit Lösungen zur statistischen Methodenlehre, Verlag Harri Deutsch, Frankfurt am Main

Lippe von der, P. (2004), Klausurtraining in Statistik, R. Oldenbourg Verlag, 6. Auflage, München

Merz, J. (1997), Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik, Übungs- und Klausuraufgaben mit Lösungen, Lüneburg

Spiegel, M. R. (1999), Statistik, Mc Graw Hill, Düsseldorf

Vogel, F. (2000), Beschreibende und schließende Statistik – Aufgaben und Beispiele, R. Oldenbourg Verlag, 12. Auflage, München

C FORMEL- UND TABELLENWERKE

Bihn, E. R. und K. A. Schäffer (1987), Formeln und Tabellen zur Grundausbildung in Statistik für Wirtschaftswissenschaftler, J. C. Witsch Nachf., Köln

Bleymüller, J. und G. Gehlert (1985), Statistische Formeln, Tabellen und Programme, Verlag Franz Vahlen, München

Bohley, P. (1991), Formeln, Rechenregeln und Tabellen zur Statistik, R. Oldenbourg Verlag, 4. Auflage, München

Rinne, H. (1988), Statistische Formelsammlung, Verlag Harri Deutsch, Frankfurt am Main

Vogel, F. (1991), Beschreibende und schließende Statistik – Formeln, Definitionen, Erläuterungen, Stichwörter und Tabellen, R. Oldenbourg Verlag, 6. Auflage, München