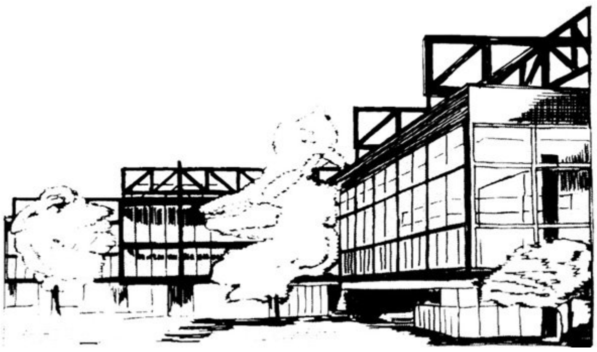
HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

FÄCHERGRUPPE MATHEMATIK / STATISTIK

Formelsammlung Stochastik bzw. Statistik

Beschreibende Statistik
Wahrscheinlichkeitsrechnung
Schließende Statistik



5. Februar 2020

Inhaltsverzeichnis

1 Beschreibende Statistik **4**
1.1 Grundbegriffe 4
1.2 Maßzahlen 7
1.2.1 Lagemaße 7
1.2.2 Streuungsmaße 8
1.3 Verhältniszahlen und Indizes 10
1.4 Messung wirtschaftlicher Konzentration 11
1.4.1 Lorenzkurve und Ginikoeffizient 11
1.4.2 Konzentrationskurve und Konzentrationsmaße 12
1.5 Zusammenhangs- bzw. Assoziationsmaße 12
1.5.1 Nominalskala 12
1.5.2 Metrische Skala 13
1.5.3 Ordinalskala (oder besser) 13
1.5.4 Konventionen bzgl. Stärke des Zusammenhangs 14
1.6 Empirische lineare Regression 14

2 Wahrscheinlichkeitsrechnung **16**
2.1 Grundlagen 16
2.1.1 Mengen, Ereignisse, σ-Algebra 16
2.1.2 Wahrscheinlichkeit(smaß) 17
2.1.3 Bedingte, totale Wahrscheinlichkeit, Unabhängigkeit, Bayes 17
2.1.4 Kombinatorik 18
2.2 Zufallsvariable und Verteilungen 19
2.2.1 Zufallsvariable (ZV) 19
2.2.2 Diskrete Zufallsvariable 20
2.2.3 Stetige Zufallsvariable 20
2.2.4 Zweidimensionale Verteilungsfunktionen 21
2.2.5 Erwartungswert einer Zufallsvariable 22
2.2.6 Varianz 23
2.2.7 Kovarianz 23
2.2.8 Tschebyschoff-Ungleichung 24
2.2.9 Zentraler Grenzwertsatz 24
2.2.10 Grenzwertsatz von Poisson 25
2.2.11 Approximationen 25
2.2.12 Summen von Zufallsvariablen 25

3 Schließende Statistik **26**
3.1 Zufallsstichproben 26
3.1.1 Stichprobenfunktionen bzw. Statistiken 26
3.1.2 Spezielle Stichprobenfunktionen 26
3.2 Punktschätzverfahren 27
3.2.1 Eigenschaften von Schätzern 27

3.2.2 Konstruktion von Schätzern . . . 27
3.3 Intervallschätzung (Konfidenzintervalle) . . . 28
3.3.1 Definitionen . . . 28
3.3.2 Spezielle Konfidenzintervalle . . . 28
3.4 Hypothesentests . . . 28
3.4.1 Konstruktion eines Tests . . . 29
3.4.2 Verwerfungsbereiche von Tests auf Parameter ϑ . . . 29
3.4.3 p-Wert . . . 29
3.5 Klassische Parametertests . . . 30
3.6 Tests für Abhängigkeit bzw. Korrelationsmaße . . . 31
3.6.1 Normalverteilung: Test auf lineare Korrelation . . . 31
3.6.2 Rangdaten: Test auf monotonen Zusammenhang . . . 32
3.6.3 Nominaldaten: χ^2-Unabhängigkeitstest von Pearson . . . 32
3.6.4 Vierfeldertafeln: Exakter Test von Fisher (1970) . . . 33
3.7 Anpassungstests . . . 34
3.7.1 Allgemeines . . . 34
3.7.2 Kolmogoroff-Smirnoff-Test . . . 34
3.7.3 χ^2-Anpassungstest von Pearson . . . 35
3.7.4 QQ-Plots . . . 35
3.7.5 Shapiro-Wilk-Test (für Normalverteilung) . . . 36
3.8 Tests und mehr im Regressionsmodell . . . 37
3.8.1 Grundlagen . . . 37
3.8.2 Tests bzgl. Absolutglied β_0 und Anstieg β_1 . . . 38
3.8.3 F-Test . . . 39
3.8.4 Prognose und Prognoseintervall . . . 39

4 Appendix **40**
4.1 Diskrete Verteilungen . . . 41
4.2 Stetige Verteilungen . . . 42
4.3 Übersicht Konfidenzintervalle . . . 44
4.4 Übersicht Einstichprobentests . . . 45
4.5 Übersicht Zweistichprobentests . . . 46
4.6 Verteilungsfunktion der Standardnormalverteilung . . . 47
4.7 Quantile der Standardnormalverteilung . . . 49
4.8 Quantile der t-Verteilung . . . 50
4.9 Quantile der χ^2-Verteilung . . . 51
4.10 Quantile der F-Verteilung . . . 53
4.11 Kolmogoroff-Smirnoff-Test . . . 58

1 Beschreibende Statistik

1.1 Grundbegriffe

- Untersuchung: Vorbereitung, Planung, Ausführung
- Grundgesamtheit Ω: Die Menge der Objekte ω (statistische Einheiten), auf die sich die Untersuchung bezieht, geschrieben: $\omega \in \Omega$
- Merkmal $X\colon \Omega \to S$ mit Merkmalsraum S, Realisierung $x = X(\omega)$.
- Skalen:
  - *Voraussetzung:* x und y seien Realisationen des Merkmals.
  - *Nominalskala (klassifikatorische/kategoriale Skala):*

    Lediglich $x = y$ bzw. $x \neq y$ sind interpretierbar.
  - *Ordinalskala (Rangskala):*

    Es existiert Ordnungsrelation: $x = y$ oder $x > y$ oder $x < y$.
  - *Intervallskala:* Ordinalskala + die absolute Differenz zwischen 2 Ausprägungen hat eine inhaltliche Bedeutung.
  - *Verhältnisskala:* Intervallskala + fester Nullpunkt
  - *metrische Skala:* Intervall- oder Verhältnisskala.
- Mächtigkeit:
  - Merkmal heißt *diskret*, falls Ω höchstens abzählbar (endlich oder gleichmächtig zu $\mathbb{N}$) ist.
  - Merkmal heißt *stetig (kontinuierlich)*, falls Ω überabzählbar ist.
- Untersuchungseinheiten: $\omega_1, \omega_2, \ldots, \omega_n$
- Ausprägungen bzw. Originalwerte: $x_1, x_2, \ldots, x_n$, wobei $x_j = X(\omega_j)$.
- Häufigkeiten der verschiedenen Ausprägungen $\{a_i\}_{i=1}^k$ mit $a_i \neq a_j$ für $i \neq j$:
  - *absolute Häufigkeit* von $a_i : n_i = n(a_i) =$ Anzahl der Fälle, in denen a_i auftritt.
  - *relative Häufigkeit* von $a_i : h_i = h(a_i) = n_i/n$.
  - Graphische Darstellung: *Stab-, Balken-, Kreisdiagramm, Häufigkeitspolygon*.

$\downdownarrows$ **Voraussetzung:** Merkmal sei nun wenigstens *ordinalskaliert* (*metrisch* für Hist.), d. h. Sortierung $a_1 < a_2 < \ldots < a_k$

- *Summenhäufigkeit:*
  - *absolute* $\sim$ von a_i: $N(a_i) = \sum_{j=1}^{i} n(a_j) = \sum_{j=1}^{i} n_j$.
  - *relative* $\sim$ von a_i: $H(a_i) = \sum_{j=1}^{i} h(a_j) = \sum_{j=1}^{i} h_j = N(a_i)/n$.

- Histogramm

  Voraussetzung: Metrisches Merkmal mit Merkmalsraum $S \subset \mathbb{R}$

  Klassenbildung: Zerlege S in r (meist gleichbreite) disjunkte Intervalle (Klassen) K_i mit

$$
\begin{aligned}
K_i &= (x_i^{\mathrm{u}}, x_i^{\mathrm{o}}], \quad \text{Klassenbreite} \quad |K_i| = x_i^{\mathrm{o}} - x_i^{\mathrm{u}}, \\
n(K_i) &= \text{Anzahl der Ausprägungen in } K_i, \\
h(K_i) &= n(K_i)/n.
\end{aligned}
$$

  Das *Histogramm (empirische Dichte)* ist dann

$$
f_n(x) = \frac{\text{relative Klassenhäufigkeit}}{\text{Klassenbreite}} = \frac{h(K_i)}{|K_i|} \quad \text{für} \quad x \in K_i.
$$

- Empirische Verteilungsfunktion

  Die *empirische Verteilungsfunktion* $F_n(x)$ ist gleich der relativen Anzahl der Realisationen, die $\leq x$ sind, d. h.

$$
F_n(x) = \begin{cases} 0 & , \; x < a_1 \\ H(a_i) = \sum_{j=1}^{i} h(a_j) & , \; a_i \leq x < a_{i+1} \\ 1 & , \; a_k \leq x \end{cases}
$$

  Eigenschaften:

  1. $F_n(x)$ monoton steigend (alias nichtfallend), $F_n(x)$ rechtsseitig stetig.
  2. $F_n(a_j) - F_n(a_j - 0) = h(a_j)$, $\quad F_n(a_j - 0) := \lim_{\varepsilon \downarrow 0} F_n(a_j - \varepsilon)$.

  Empirische Verteilungsfunktion für klassierte Daten (lineare Interpolation):

$$
F_n(x) = \begin{cases} 0 & , \; x < x_1^u \\ F_n(x_i^u) + \frac{x - x_i^u}{x_i^o - x_i^u}\left(F_n(x_i^o) - F_n(x_i^u)\right) & , \; x \in (x_i^u, x_i^o] \\ 1 & , \; x \geq x_r^o \end{cases},
$$

  oder auch für $x \in (x_i^u, x_i^o]$:

$$
F_n(x) = \sum_{j<i} h(K_j) + (x - x_i^u)\frac{h(K_i)}{|K_i|} = F_n(x_i^u) + (x - x_i^u) f_n(x_i^o)
$$

- Bivariate Häufigkeit

  Voraussetzung: Zwei Merkmale X, Y mit beobachteten Ausprägungspaaren

  $(x_1, y_1), \ldots, (x_n, y_n)$

  und verschiedenen Ausprägungen

  X: $a_1 < a_2 < \ldots < a_k$
  Y: $b_1 < b_2 < \ldots < b_l$

  - *absolute Häufigkeit von (a_i, b_j):* $n_{ij} = n(X = a_i, Y = b_j) =$ Anzahl der Fälle, in denen das Paar (a_i, b_j) auftritt.
  - *absolute Randhäufigkeit von a_i:* $n_{i\bullet} =$ Anzahl der Fälle, in denen die Ausprägung a_i in $x_1, \ldots, x_n$ auftritt, analog ist $n_{\bullet j}$ für b_j definiert.
  - *Bivariate Häufigkeitstabelle/Kontingenztafel/Kreuztabelle:*

X	Y				
	b_1	b_2	$\cdots$	b_l	$\sum_j$
a_1	n_{11}	n_{12}	$\cdots$	n_{1l}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	$\cdots$	n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}	$n_{k\bullet}$
$\sum_i$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet l}$	n

  - Entsprechende *relative* Häufigkeiten $h(\cdot)$ ergeben sich durch Division mit Gesamtanzahl n.
  - *bedingte relative Häufigkeit von $X = a_i$, falls $Y = b_j$:*

$$h(X = a_i \,|\, Y = b_j) = \frac{n(X = a_i, Y = b_j)}{n(Y = b_j)} = \frac{n_{ij}}{n_{\bullet j}}.$$

  - *bedingte relative Häufigkeit von $Y = b_j$, falls $X = a_i$:*

$$h(Y = b_j \,|\, X = a_i) = \frac{n(X = a_i, Y = b_j)}{n(X = a_i)} = \frac{n_{ij}}{n_{i\bullet}}.$$

- Bivariates Histogramm

  - Klassenbildung je Komponente (X, Y): K_i^X, K_j^Y.
  - $f_n(x, y) = \dfrac{h\big((X, Y) \in K_i^X \times K_j^Y\big)}{|K_i^X|\,|K_i^Y|}$ für $x \in K_i^X$ und $y \in K_j^Y$, sonst 0.
  - Aus Rechtecken werden Quader.

- Empirische bivariate Verteilungsfunktion

  Beginnen mit der Kreuztabelle der *relativen* Häufigkeiten:

$$
\begin{array}{c|cccc|c}
X & & Y & & & \\
 & b_1 & b_2 & \cdots & b_l & \sum_j \\
\hline
a_1 & h_{11} & h_{12} & \cdots & h_{1l} & h_{1\bullet} \\
a_2 & h_{21} & h_{22} & \cdots & h_{2l} & h_{2\bullet} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
a_k & h_{k1} & h_{k2} & \cdots & h_{kl} & h_{k\bullet} \\
\hline
\sum_i & h_{\bullet 1} & h_{\bullet 2} & \cdots & h_{\bullet l} & 1
\end{array}
$$

Empirische bivariate Verteilungsfunktion von (X, Y):

$$F_n(x,y) = \sum_{i:a_i \leq x} \; \sum_{j:b_j \leq y} h_{ij}$$

Vorgehensweise bei *klassierten* Daten: Bilineare Interpolation zwischen den Ecken; diese ergeben sich durch Aufsummierung der relativen Klassenhäufigkeiten.

1.2 Maßzahlen

1.2.1 Lagemaße

Arithmetisches Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{j=1}^{k} n_j a_j = \sum_{j=1}^{k} h_j a_j \qquad \text{(Bei klassierten Daten ist } a_j \text{ die } \textbf{Klassenmitte}\text{!)}\,.$$

Geometrisches Mittel

$$\bar{x}_G = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} = \left(\prod_{j=1}^{k} a_j^{n_j} \right)^{\frac{1}{n}} = \prod_{j=1}^{k} a_j^{h_j}\,.$$

Harmonisches Mittel

$$\bar{x}_H = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^{k} \frac{n_j}{a_j} \right)^{-1} = \left(\sum_{j=1}^{k} \frac{h_j}{a_j} \right)^{-1}\,.$$

$\downdownarrows$ **Vorarbeit geordnete Stichprobe:** Sortierte Werte $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

α-getrimmtes Mittel $\big(\alpha \in (0, 0.5)\big)$

$$\bar{x}_\alpha = \frac{1}{n-2h} \sum_{i=h+1}^{n-h} x_{(i)} \quad \text{mit } h = \lfloor n\alpha \rfloor \; \Big(h \in \mathbb{N}\,,\; h \leq n\alpha < h+1 \;\widehat{=}\; \text{Abrunden} \Big)\,.$$

α-Quantile $\big(\alpha \in (0, 1)\big)$

$$x_\alpha \begin{cases} = x_{(\lfloor n\alpha \rfloor + 1)} & n\alpha \notin \mathbb{N} \text{ oder Rangskala} \\ \in \left[x_{(n\alpha)},\, x_{(n\alpha+1)}\right] \text{ bzw. } \Big(x_{(n\alpha)} + x_{(n\alpha+1)} \Big) /2 & n\alpha \in \mathbb{N},\ \text{metrische (stetige) Skala} \end{cases}$$

- α-Quantile für klassierte Größen $(\alpha \in (0,1))$

$$i = \text{Klassennummer mit } F_n(x_i^u) \leq \alpha < F_n(x_i^o)\,,$$
$$x_\alpha = x_i^u + \frac{\alpha - F_n(x_i^u)}{f_n(x_i^o)} \quad \left(\text{Histogramm } f_n(x_i^o) = \frac{F_n(x_i^o) - F_n(x_i^u)}{x_i^o - x_i^u} = \frac{h(K_i)}{|K_i|}\right)\,.$$

- Spezielle Quantile (Quartile)

$$\begin{array}{lll} x_{0.25} & & \text{1. bzw. unteres Quartil} \\ x_{0.5} = median(x) = x_{\text{med}} = \tilde{x} & \text{heißt} & \text{Median (2. Quartil)} \\ x_{0.75} & & \text{3. bzw. oberes Quartil} \end{array}$$

Modus

$$x_d = \begin{cases} \text{argmax}\{n(x_1), n(x_2), \ldots, n(x_k)\} \\ \quad \text{für diskrete Merkmale, d. h. häufigster Wert} \\ \text{Mitte von } \text{argmax}\{f_n(K_1), f_n(K_2), \ldots, f_n(K_r)\} \\ \quad \text{für klassierte (stetige) Merkmale, d. h. Mitte der dichtesten Klasse} \\ \quad \text{(Klassenmitte des höchsten Histogrammrechtecks)} \end{cases}$$

1.2.2 Streuungsmaße

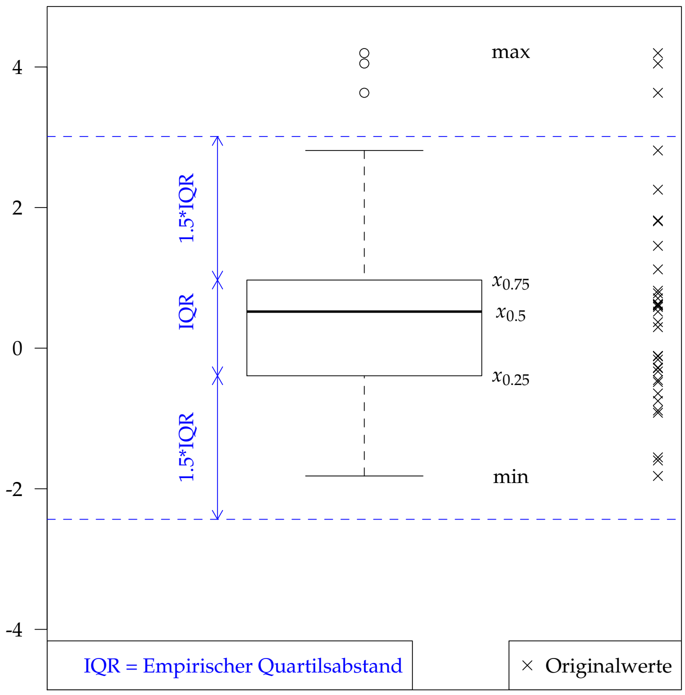
Empirische Spannweite

$$R = x_{(n)} - x_{(1)}\,.$$

Empirischer Quartilsabstand (IQR)

$$d_Q = x_{0.75} - x_{0.25}\,.$$

Boxplot



Median der absoluten Abweichungen vom Median

$$MAD = median\big(|x - x_{0.5}|\big)\,.$$

Mittlere absolute Abweichung vom Median

$$d_X = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_{0.5}| = \frac{1}{n}\sum_{j=1}^{k} n_j|a_j - x_{0.5}| = \sum_{j=1}^{k} h_j|a_j - x_{0.5}|\,.$$

Empirische Varianz

$$\begin{aligned}\tilde{s}^2 &= \frac{1}{n}\sum_{i=1}^{n}\big(x_i - \bar{x}\big)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2\\ &= \frac{1}{n}\sum_{j=1}^{k} n_j(a_j - \bar{x})^2 = \frac{1}{n}\sum_{j=1}^{k} n_j a_j^2 - \bar{x}^2 = \sum_{j=1}^{k} h_j a_j^2 - \bar{x}^2\,.\end{aligned}$$

Bemerkung: Bei klassierten Daten setze a_j gleich der Klassenmitte von K_j.

Stichprobenvarianz

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(x_i - \bar{x}\big)^2 = \frac{n}{n-1}\tilde{s}^2\,.$$

Bemerkung: $\tilde{s}^2$ wird in der Beschreibenden und s^2 in der Schließenden Statistik bevorzugt.

(Empirische) Standardabweichung

$$\tilde{s} = \sqrt{\tilde{s}^2}\,.$$

Empirischer Variationskoeffizient

$$V = \frac{\tilde{s}}{\bar{x}}\,.$$

Mittel und Varianz bei Zerlegung in Teilgesamtheiten

Gegeben seien $i = 1, 2, \ldots, r$ Teilgesamtheiten mit jeweils n_i Beobachtungen x_{ij}:

$$\begin{aligned} n &= \sum_{i=1}^{r} n_i\,, \quad \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}\,, \quad \tilde{s}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}\big(x_{ij} - \bar{x}_i\big)^2,\\ \bar{x} &= \frac{1}{n}\sum_{i=1}^{r} n_i\bar{x}_i \quad \left(= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij}\right),\\ \tilde{s}^2 &= \underbrace{\frac{1}{n}\sum_{i=1}^{r} n_i\tilde{s}_i^2}_{\text{interne Varianz}} + \underbrace{\frac{1}{n}\sum_{i=1}^{r} n_i\big(\bar{x}_i - \bar{x}\big)^2}_{\text{externe Varianz}} \quad \left(= \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2\right).\end{aligned}$$

Effekt von Lineartransformation auf Mittel und Varianz:

Sei $y_i = a + bx_i$ für $i = 1, 2, \ldots, n$.

Dann folgt für die zugehörigen Mittelwerte und Varianzen

$$\bar{y} = a + b\bar{x}\,,$$

$$\tilde{s}_Y^2 = b^2\tilde{s}_X^2 \quad \text{bzw.} \quad \tilde{s}_Y = |b|\tilde{s}_X\,.$$

1.3 Verhältniszahlen und Indizes[1]

Indexzahlen

Voraussetzung: z_t bezeichne eindimensionale Zahlenreihe, Index t steht für Zeit

Messzahlen (in %) zum Basiszeitpunkt 0:

$$I_{0,t} = \frac{z_t}{z_0} \cdot 100\,.$$

Umbasieren (von 0 auf r):

$$I_{r,t} = \frac{I_{0,t}}{I_{0,r}} \cdot 100\,.$$

Verketten (von r auf 0):

$$I_{0,t} = \frac{I_{0,r} I_{r,t}}{100}\,.$$

Preis- und Mengenindizes $q_{i,t}$ u. $p_{i,t}$ Menge bzw. Preis von Gut i zum Zeitpunkt t

Wertindex (in %):

$$V_{0,t} = \frac{\sum\limits_{i=1}^{n} q_{i,t} p_{i,t}}{\sum\limits_{i=1}^{n} q_{i,0} p_{i,0}} \cdot 100\,.$$

Preisindex nach *Paasche* bzw. *Laspeyres* (in %; Mengen zum Zeitpunkt t bzw. 0):

$$P^P_{0,t} = \frac{\sum\limits_{i=1}^{n} q_{i,t} p_{i,t}}{\sum\limits_{i=1}^{n} q_{i,t} p_{i,0}} \cdot 100\,, \qquad P^L_{0,t} = \frac{\sum\limits_{i=1}^{n} q_{i,0} p_{i,t}}{\sum\limits_{i=1}^{n} q_{i,0} p_{i,0}} \cdot 100\,.$$

Mengenindex nach *Paasche* bzw. *Laspeyres* (in %; Preise zum Zeitpunkt t bzw. 0):

$$Q^P_{0,t} = \frac{\sum\limits_{i=1}^{n} q_{i,t} p_{i,t}}{\sum\limits_{i=1}^{n} q_{i,0} p_{i,t}} \cdot 100\,, \qquad Q^L_{0,t} = \frac{\sum\limits_{i=1}^{n} q_{i,t} p_{i,0}}{\sum\limits_{i=1}^{n} q_{i,0} p_{i,0}} \cdot 100\,.$$

Deflationierung:

$$Q^P_{0,t} = \frac{V_{0,t}}{P^L_{0,t}} \cdot 100 \qquad \text{bzw.} \qquad Q^L_{0,t} = \frac{V_{0,t}}{P^P_{0,t}} \cdot 100\,.$$

Wertgewichtsmethode für Preis- und Mengenindex nach *Laspeyres*

$$P^L_{0,t} = \sum_{i=1}^{n} w_{i0} \frac{p_{i,t}}{p_{i,0}} 100\,, \quad Q^L_{0,t} = \sum_{i=1}^{n} w_{i0} \frac{q_{i,t}}{q_{i,0}} 100 \quad \text{mit} \quad w_{i0} = \frac{q_{i,0} p_{i,0}}{\sum_{j=1}^{n} q_{j,0} p_{j,0}}$$

[1] Nur relevant für BWL, POL, VWL.

mit *Wertgewicht* w_{i0} (Umsatzanteil der Ware i in der Basisperiode 0)

$$\frac{p_{i,t}}{p_{i,0}}100 \quad \text{und} \quad \frac{q_{i,t}}{q_{i,0}}100 \quad \textit{Preismessziffern} \text{ und } \textit{Mengenmessziffern}$$

Subindizes

$$P^L_{0,t} = w_{A0}\, P^A_{0,t} + w_{B0}\, P^B_{0,t} \quad \text{bzw.} \quad Q^L_{0,t} = w_{A0}\, Q^A_{0,t} + w_{B0}\, Q^B_{0,t}$$

w_{A0} bzw. w_{B0} Umsatzanteil der Warengruppe A bzw. B am Gesamtumsatz

$P^A_{0,t}$ bzw. $P^B_{0,t}$ *Laspeyres*-Preisindizes, $Q^A_{0,t}$ bzw. $Q^B_{0,t}$ *Laspeyres*-Mengenindizes für Warengruppe A bzw. B.

1.4 Messung wirtschaftlicher Konzentration[2]

1.4.1 Lorenzkurve und Ginikoeffizient

Voraussetzung: Nichtnegative (metrische) Mermalsausprägungen
$0 \leq a_1 < a_2 < \ldots < a_k$ bzw. sortierte Originalwerte $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Dem jeweiligen Wert H_i der relativen Summenhäufigkeit ordnet man die kumulierten Anteile L_i der geordneten Merkmalsanteile zu:

$$L_i = \sum_{j=1}^{i} \ell_j \qquad \text{mit} \quad \ell_j = \frac{a_j n_j}{\sum\limits_{j=1}^{k} a_j n_j}\,,\ i = 1, 2, \ldots, k$$

bzw.

$$L_i = \sum_{j=1}^{i} \ell_j \qquad \text{mit} \quad \ell_j = \frac{x_{(j)}}{\sum\limits_{j=1}^{n} x_{(j)}}\,,\ i = 1, 2, \ldots, n$$

Lorenzkurve: Streckenzug, der ausgehend von $(0, 0)$ die Punkte (H_i, L_i) verbindet.

Ginikoeffizient G ist

$$G = \sum_{i=1}^{k} \underbrace{(L_i - L_{i-1})}_{=l_i}(H_{i-1} + H_i) \ - \ 1 \quad \big(\text{mit } H_0 = 0\big)$$

$$\text{bzw.} \quad G = \frac{1}{n^2 \bar{x}} \sum_{i=1}^{n} (2i - 1) x_{(i)} \ - \ 1\,.$$

Das normierte Gini-Maß ist

$$G^* = \frac{n}{n-1} G$$

und erfüllt $0 \leq G^* \leq 1$, während für das ursprüngliche G noch $0 \leq G \leq \frac{n-1}{n}$ gilt.

[2]Nur relevant für BWL, POL, VWL.

1.4.2 Konzentrationskurve und Konzentrationsmaße

Der kumulierte Anteil C_i der i größten Merkmalsanteile

$$C_i = \ell_n + \ell_{n-1} + \ldots + \ell_{n-i+1} = \sum_{j=1}^{i} c_j \quad \text{mit} \quad c_j = \ell_{n-j+1} = x_{(n-j+1)} \Big/ \sum_{j=1}^{n} x_{(j)}$$

$$\text{bzw.} \quad C_i = \sum_{j=n-i+1}^{n} x_{(j)} \Big/ \sum_{j=1}^{n} x_{(j)} \qquad \text{heißt } \textit{Konzentrationsrate.}$$

Konzentrationskurve: Streckenzug, der von $(0,0)$ ausgehend die Punkte (i, C_i) bzw. (N_i, C_i) verbindet.

Konzentrationsindizes

- **Rosenbluth-Index:** $K_R = \dfrac{1}{2\sum_{i=1}^{n} i c_i \ - \ 1} = \dfrac{1}{\frac{2}{n\bar{x}} \sum_{i=1}^{n} i x_{(n-i+1)} \ - \ 1} = \frac{1}{n(1-G)}.$
- **Herfindahl-Index:** $K_H = \sum_{i=1}^{n} \ell_i^2 = \sum_{i=1}^{n} c_i^2 \qquad \Big(\ell_i = x_{(i)} \Big/ \sum_{j=1}^{n} x_{(j)}\Big).$
- **Exponentialindex:** $K_E = \prod_{i=1}^{n} \ell_i^{\ell_i} = \prod_{i=1}^{n} c_i^{c_i} \qquad \Big(\ell_i = x_{(i)} \Big/ \sum_{j=1}^{n} x_{(j)}\Big).$
- *Bemerkung:* Bei $x_{(1)} = \ldots = x_{(n)}$ sind $K_R = K_H = K_E = 1/n$
und bei $x_{(1)} = \ldots = x_{(n-1)} = 0\,,\ x_{(n)} > 0$ sind $K_R = K_H = K_E = 1$.

1.5 Zusammenhangs- bzw. Assoziationsmaße

1.5.1 Nominalskala

Voraussetzung: Daten als Kontingenztafel gegeben.

Unabhängigkeit

X und Y sind genau dann unabhängig, wenn für alle i, j

$$h(X = a_i, Y = b_j) = h(X = a_i)\, h(Y = b_j) \quad \text{bzw.} \quad n_{ij} = n_{i\bullet} n_{\bullet j}/n\,.$$

Chi-Quadrat-Assoziationsmaß:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(n_{ij} - n_{i\bullet} n_{\bullet j}/n\right)^2}{n_{i\bullet} n_{\bullet j}/n} = n\Big(\sum_{i=1}^{k} \sum_{j=1}^{l} \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} \ - \ 1\Big)$$

(vorausgesetzt, dass $n_{i\bullet} \neq 0$, $n_{\bullet j} \neq 0$ für alle i und j).

Spezialfall $k = l = 2$ (Vierfeldertafel):

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}}\,.$$

Kontingenzkoeffizient von Cramér:

$$V = \sqrt{\frac{\chi^2}{n(\min\{k,l\} - 1)}} \qquad \text{mit} \quad 0 \le V \le 1\,.$$

Kontingenzkoeffizient von Pearson:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad \text{mit} \quad 0 \le C \le C_{max} = \sqrt{\frac{\min\{k,l\} - 1}{\min\{k,l\}}}\,.$$

Korrigierter Kontingenzkoeffizient von Pearson:

$$C^* = \frac{C}{C_{max}} \qquad \text{mit} \quad 0 \le C^* \le 1\,.$$

1.5.2 Metrische Skala

Voraussetzung: Wertepaare auf metrischer Skala: $(x_1, y_1), \ldots, (x_n, y_n)$

Empirische Kovarianz:

$$\tilde{s}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l}(a_i - \bar{x})(b_j - \bar{y}) \cdot n_{ij} = \overline{xy} - \bar{x}\bar{y}\,.$$

Korrelationskoeffizient nach Bravais-Pearson:

$$r_{XY} = \frac{\tilde{s}_{XY}}{\tilde{s}_X\tilde{s}_Y} \qquad \text{mit} \quad -1 \le r_{XY} \le 1\,.$$

Achtung: Es wird der lineare Zusammenhang evaluiert!

1.5.3 Ordinalskala (oder besser)

Voraussetzung: Werte auf ordinaler bzw. metrischer Skala: $(x_1, y_1), \ldots, (x_n, y_n)$

Vorarbeit: Ermittlung der Ränge, d. h. jeder Beobachtung x_i bzw. y_i wird die jeweilige Position innerhalb der geordneten Werte zugeordnet:

$$R(x_i) = v \quad \Leftrightarrow \quad x_i = x_{(v)}\,.$$

Gibt es geteilte Plätze (Bindungen), d. h. $x_i = x_j\,,\ i \neq j$, so werden die Positionen aufgeteilt (Konstruktion eines mittleren Ranges): Bei bspw. zwei „Erstplatzierten“ wird $(1+2)/2 = 1.5$ an jeweils beide Beobachtungen vergeben.

Rangkorrelationskoeffizient nach Spearman $R_{XY} = r_{R(X),R(Y)}$.

Allgemein (mit $\bar{R} = (n+1)/2$):

$$R_{XY} = \frac{\sum_{i=1}^{n} \big(R(x_i) - \bar{R}\big)\big(R(y_i) - \bar{R}\big)}{\sqrt{\sum_{i=1}^{n} \big(R(x_i) - \bar{R}\big)^2 \sum_{i=1}^{n} \big(R(y_i) - \bar{R}\big)^2}} \qquad \Big(\text{mit} \quad -1 \le R_{XY} \le 1\Big)$$

$$= \frac{\sum_{i=1}^{n} R(x_i)\, R(y_i) - n\bar{R}^2}{\sqrt{\left(\sum_{i=1}^{n} R(x_i)^2 - n\bar{R}^2\right)\left(\sum_{i=1}^{n} R(y_i)^2 - n\bar{R}^2\right)}} .$$

Im Fall keiner Bindungen:

$$R_{XY} = 1 - \frac{6 \sum_{i=1}^{n} \big(R(x_i) - R(y_i)\big)^2}{n\,(n^2 - 1)} .$$

Kontingenztafel (viele Bindungen):

$$R_{XY} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} R(a_i) R(b_j) - n\bar{R}^2}{\sqrt{\sum_{i=1}^{k} n_{i\bullet} R(a_i)^2 - n\bar{R}^2}\sqrt{\sum_{j=1}^{l} n_{\bullet j} R(b_j)^2 - n\bar{R}^2}} .$$

1.5.4 Konventionen[3] bzgl. Stärke des Zusammenhangs

$\{C^*, \lvert r_{XY}\rvert, \lvert R_{XY}\rvert\}$	Interpretation	
≈ 0	keine	
$(\ldots, 0.5)$	schwache	
$[0.5, 0.8)$	mittlere	Korrelation bzw. Zusammenhang
$[0.8, 1)$	starke	
1	perfekte	

1.6 Empirische lineare Regression

Idee: Modelliere linearen Zusammenhang zwischen zwei oder mehr Merkmalen.

Einfache lineare Regression

Voraussetzung: Werte auf metrischer Skala: $(x_1, y_1), \ldots, (x_n, y_n)$

[3] Im Rahmen der Lehrveranstaltungen. Weicht in der statistischen Praxis zuweilen davon ab.

Modell: $y_i = \beta_0 + \beta_1 x_i + u_i$

Schätzung von β_0 und β_1 über Methode der kleinsten Quadrate (OLS):

$$SoS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \to \min_{\beta_0, \beta_1} .$$

Nullsetzen der beiden Ableitungen usw. liefert

$$\hat{\beta}_1 = \frac{\tilde{s}_{XY}}{\tilde{s}_X^2} \quad \text{und} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} .$$

Andere Schreibweise des Modells:

$$y_i = \underbrace{\bar{y} + \hat{\beta}_1 (x_i - \bar{x})}_{=\hat{\mu}_i} + u_i .$$

Varianzzerlegung:

$$\begin{aligned} \sum \left(y_i - \bar{y}\right)^2 &= \sum \left(y_i - \hat{\mu}_i\right)^2 + \sum \left(\hat{\mu}_i - \bar{y}\right)^2 , \\ \tilde{s}^2 &= \tilde{s}_{\hat{u}}^2 + \tilde{s}_{\hat{\mu}}^2 \\ \text{Varianz der } y &= \text{Varianz der emp. Residuen } \hat{u} + \text{Erklärte Varianz} . \end{aligned}$$

Bestimmtheitsmaß:

$$R^2 = \frac{\tilde{s}_{\hat{\mu}}^2}{\tilde{s}_y^2} = \frac{\sum \left(\hat{\mu}_i - \bar{y}\right)^2}{\sum \left(y_i - \bar{y}\right)^2}$$

... und bei **einfachem** linearen Regressionsmodell:

$$R^2 = r_{XY}^2 .$$

Multiple lineare Regression

Modell:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \qquad (k > 1) .$$

Matrixschreibweise:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$$\boldsymbol{y} = \mathbb{X}\,\boldsymbol{\beta} + \boldsymbol{u} .$$

Methode der kleinsten Quadrate usw.:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbb{X}'\,\mathbb{X})^{-1}\mathbb{X}'\,\boldsymbol{y} \\ \hat{\boldsymbol{\mu}} &= \mathbb{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbb{X}(\mathbb{X}'\,\mathbb{X})^{-1}\mathbb{X}'}_{\text{Hat-Matrix } \mathbb{H}}\boldsymbol{y} \quad , \mathbb{H} - \text{Projektion auf Spalten von } \mathbb{X}. \end{aligned}$$

Die Matrix $\mathbb{X}$ muss vollen Rang haben, d. h. die Spalten von $\mathbb{X}$ müssen linear unabhängig sein.

2 Wahrscheinlichkeitsrechnung

2.1 Grundlagen

2.1.1 Mengen, Ereignisse, σ-Algebra

- *Grundgesamtheit:* Ω – Menge aller Elementarereignisse ω,
- *Ereignis*: geeignete Teilmenge von Ω,
- Operationen auf Ereignissen wie auf Mengen, $A, B \subseteq \Omega$:
  - $A \cap B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\}$ – „A und B treten ein" (**Durchschnitt**).
  - $A \cup B = \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\}$ – „mindestens eines der beiden Ereignisse tritt ein, d. h. A oder B oder beide" (inklusives „oder") (**Vereinigung**).
  - $A \setminus B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \notin B\}$ – „A aber nicht B tritt ein" (**Differenz**).
  - $\bar{A}\ \big(= \sim A = A^c \big) = \{\omega \in \Omega : \omega \notin A\}$ – „A tritt nicht ein" – **komplementäres Ereignis (Negation)**.
- Rechenregeln, $A, B \subseteq \Omega$:

$$
\begin{aligned}
A \cup B &= B \cup A\,, \\
A \cap B &= B \cap A\,, \\
(A \cup B) \cup C &= A \cup (B \cup C)\,, \\
(A \cap B) \cap C &= A \cap (B \cap C)\,, \\
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C)\,, \\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)\,, \\
\overline{(A \cap B)} &= \bar{A} \cup \bar{B}\,, \\
\overline{(A \cup B)} &= \bar{A} \cap \bar{B}\,, \\
A \setminus B &= A \cap \bar{B} \\
\overline{\overline{A}} &= A.
\end{aligned}
$$

- Zwei Ereignisse A und B heißen *disjunkt (unvereinbar)*, falls $A \cap B = \emptyset$ ist.
- $\emptyset$ bzw. Ω heißen das **unmögliche** bzw. das **sichere** Ereignis.
- Ein Mengensystem (Menge aus Mengen) $\mathcal{A}$ in Ω heißt eine **σ-Algebra (ein Ereignissystem)** in Ω, wenn es folgende Eigenschaften hat:
  1. $\Omega \in \mathcal{A}$.
  2. $A \in \mathcal{A} \;\Rightarrow\; \bar{A} \in \mathcal{A}\,.$

3. $A_n \in \mathcal{A}\,,\ n = 1,2,\ldots \ \Rightarrow\ A_1 \cup A_2 \cup \ldots = \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}\,.$

Dann gilt auch:

- $\emptyset \in \mathcal{A}$.
- $A_n \in \mathcal{A}\,,\ n = 1,2,\ldots \ \Rightarrow\ A_1 \cap A_2 \cap \ldots = \bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{A}\,,$
- $A, B \in \mathcal{A} \ \Rightarrow\ A \setminus B \in \mathcal{A}\,.$

- Das Paar $(\Omega, \mathcal{A})$ heißt **messbarer Raum (Messraum)**.

2.1.2 Wahrscheinlichkeit(smaß)

- **Wahrscheinlichkeitsmaß** P – Abbildung, die allen Ereignissen $A \in \mathcal{A}$ eine Zahl, $P(A)$, zuordnet und die die folgenden Bedingungen erfüllt:
  1. $0 \le P(A) \le 1,$
  2. $P(\Omega) = 1,$
  3. $P\Big(\bigcup\limits_{n=1}^{\infty} A_n\Big) = \sum\limits_{n=1}^{\infty} P(A_n)$ für disjunkte $A_n \in \mathcal{A}$ ($A_i \cap A_j = \emptyset$ für $i \neq j$).
- Das Tripel $(\Omega, \mathcal{A}, P)$ heißt **Wahrscheinlichkeitsraum**.
- Rechenregeln, $A, B \in \mathcal{A}$:
  1. $P(\bar{A}) = 1 - P(A)\,,$
  2. $P(\emptyset) = 0\,,$
  3. $P(A) = P(A \cap B) + P(A \cap \bar{B})\,,$
  4. Ist $B \subseteq A$, so ist $P(B) \le P(A)\,,$
  5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)\,,$
  6. Ist Ω endlich, so gilt für $A \in \mathcal{A}$: $\quad P(A) = \sum\limits_{a \in A} P(\{a\})\,.$
- LAPLACE-Wahrscheinlichkeit (Ω endlich und alle Elementarereignisse besitzen dieselbe Wahrscheinlichkeit):

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ „günstigen Fälle“}}{\text{Anzahl der möglichen Fälle}},$$

wobei $|A|$ die Anzahl der Elemente von A bezeichne, analog $|\Omega|$.

2.1.3 Bedingte, totale Wahrscheinlichkeit, Unabhängigkeit, Bayes

- *Bedingte Wahrscheinlichkeit:* Für Ereignisse A und B mit $P(B) > 0$ ist

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

die **bedingte Wahrscheinlichkeit** für A, wenn bereits bekannt ist, dass das Ereignis B bereits eingetreten ist.

- *Unabhängigkeit:* Ereignisse A und B mit $P(A) > 0$ und $P(B) > 0$ sind genau dann **(stochastisch) unabhängig**, wenn
$$P(A \cap B) = P(A) \cdot P(B) \quad \Leftrightarrow \quad P(A \mid B) = P(A) \quad \Leftrightarrow \quad P(B \mid A) = P(B)\,.$$
Ereignisse $A_i, i = 1, 2, \ldots, n$ heißen **(stochastisch) unabhängig**, wenn für jede endliche Teilmenge $\{i_1, i_2, \ldots, i_m\} \subseteq \{1, 2, \ldots, n\}$ gilt, dass
$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \ldots \cdot P(A_{i_m})\,.$$

- *Multiplikationssatz für bedingte Wahrscheinlichkeiten:*
Für $n \geq 2$ Ereignisse $A_1, \ldots, A_n$ mit $P(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$ gilt
$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_n) &\\ = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots &\\ \cdot P(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1})\,. \end{aligned}$$

- *Satz von der totalen Wahrscheinlichkeit:* Für das Ereignis A und paarweise disjunkte Ereignisse $C_1, \ldots, C_n$ mit $C_1 \cup C_2 \cup \ldots \cup C_n = \Omega$ ist
$$P(A) = \sum_{i=1}^{n} P(A \mid C_i) P(C_i) \qquad \Big(= \sum_{i=1}^{n} P(A \cap C_i) \Big)$$
die **totale Wahrscheinlichkeit** des Ereignisses A.

- *Satz von Bayes:* Für diesselben Ereignisse wie im vorigen Absatz gilt die **Formel von Bayes**
$$P(C_j \mid A) = \frac{P(A \mid C_j) P(C_j)}{\sum\limits_{i=1}^{n} P(A \mid C_i) P(C_i)}\,.$$

2.1.4 Kombinatorik

Modell: Urne mit von 1 bis n nummerierten Kugeln, Ziehung von k Kugeln.

Man unterscheidet

a) Ziehungen in Reihenfolge **(Variationen)**

   a_1) mit Zurücklegen

   a_2) ohne Zurücklegen

b) Ziehungen ohne (Berücksichtigung der) Reihenfolge **(Kombinationen)**

   b_1) mit Zurücklegen

   b_2) ohne Zurücklegen

Die jeweiligen Anzahlen verschiedener Möglichkeiten sind hier zusammengefasst:

Reihenfolge	Zurücklegen: ja	Zurücklegen: nein
ja	n^k	$\dfrac{n!}{(n-k)!}$
nein	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Dabei bedeuten

$$n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n \qquad \text{(Fakultät)},$$

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!} \qquad \textbf{(Binomialkoeffizient)}.$$

Rechenregeln für Binomialkoeffizienten ($0 \le k \le n$):

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n, \quad \binom{n}{k} = \binom{n}{n-k},$$

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \qquad \text{(Pascalsches Dreieck)}.$$

Verallgemeinerung von b$_2$): Anzahl der Möglichkeiten, n verschiedene Objekte in k verschiedene Mengen vom Umfang $n_1, \ldots, n_k$ einzuteilen ($n_1 + \ldots + n_k = n$):

$$\frac{n!}{n_1! \ldots n_k!} = \binom{n}{n_1 \ldots n_k}.$$

Diese Größe heißt **Multinomialkoeffizient**.

2.2 Zufallsvariable und Verteilungen

2.2.1 Zufallsvariable (ZV)

Voraussetzung: Zufallsexperiment, beschrieben durch Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$

Eine Abbildung $X : \Omega \to \mathbb{R}$ von der Grundgesamtheit Ω in den Bildraum $S = \mathbb{R}$ heißt **(reelle) Zufallsvariable**.

Der Vektor $\boldsymbol{X} = (X_1, \ldots, X_k)$ mit Zufallsvariablen $X_1, \ldots, X_k$ heißt
k-dim. Zufallsvariable (Zufallsvektor).

Verteilungsfunktion von Zufallsvariable X

$$F_X(x) = P\Big(\{\omega \in \Omega : X(\omega) \le x\}\Big), \; x \in \mathbb{R} \quad \textit{kurz:} \quad F(x) = P(X \le x)$$

- $0 \le F(x) \le 1$ für alle x, $\;F(\infty) := \lim\limits_{x\to\infty} F(x) = 1$, $\;F(-\infty) := \lim\limits_{x\to-\infty} F(x) = 0$
- $F(x)$ ist monoton wachsend in x
- F ist rechtsseitig stetig, d. h. für alle x gilt $\underbrace{\lim\limits_{\varepsilon\downarrow 0} F(x+\varepsilon)}_{=:\,F(x+0)} = F(x)$

Berechnung von Wahrscheinlichkeiten Es gilt für $a < b$:

- $P(a < X \leq b) = F(b) - F(a)$
- $P(a \leq X \leq b) = F(b) - \underbrace{\lim_{\varepsilon \downarrow 0} F(a - \varepsilon)}_{=: F(a-0)}$
- $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$,
  $P(X \geq a) = 1 - P(X < a) = 1 - F(a - 0)$
- $P(X = a) = F(a) - F(a - 0)$

2.2.2 Diskrete Zufallsvariable

Nimmt X höchstens abzählbar viele verschiedene Werte an, so heißt X eine **diskrete Zufallsvariable** und F_X eine **diskrete Verteilungsfunktion**.

Werte von X: $\quad \ldots < x_{i-1} < x_i < x_{i+1} < \ldots$

Wahrscheinlichkeitsfunktion von X ist

$$f_X(x) = \begin{cases} p_i & \text{falls } x = x_i \\ 0 & \text{falls } x \neq x_i \text{ für alle } i \end{cases} \quad \text{mit} \quad p_i = P(X = x_i)\,.$$

Verteilungsfunktion ist

$$F_X(x) = \sum_{j=1}^{i} f_X(x_j) = P(X = x_1) + \cdots + P(X = x_i) \quad \text{für} \quad x_i \leq x < x_{i+1}\,.$$

Insbesondere ist $F(x) = 0$ für $x < x_1$ und $F(x) = 1$ für $x > x_n$.

Beispiele für diskrete Verteilungen siehe Seite 41.

2.2.3 Stetige Zufallsvariable

Definition

X heißt **stetige** Zufallsvariable, falls für alle $x \in \mathbb{R}$ gilt:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt\,, \quad f(t) \geq 0 \quad \text{für alle } t \in \mathbb{R}\,.$$

f heißt **Dichte (Wahrscheinlichkeitsdichte)** von X.

Eigenschaften ($a \leq b \in \mathbb{R}$)

- $\int_{-\infty}^{\infty} f(t)\,dt = 1$
- $P(a < X \leq b) = \int_a^b f(t)\,dt\,, \quad P(X > a) = \int_a^\infty f(t)\,dt$
- $P(a < X < b) = P(a \leq X \leq b)\,, \quad P(X = a) = 0$
- Wenn F differenzierbar, dann gilt $f(x) = F'(x)$.

Beispiele für stetige Verteilungen siehe Seite 42 f.

2.2.4 Zweidimensionale Verteilungsfunktionen

Definitionen

Voraussetzung: (X, Y) 2-dim. Zufallsvariable:

$$F(x, y) = P\Big(\{\omega \in \Omega : X(\omega) \leq x,\, Y(\omega) \leq y\}\Big) = P(X \leq x,\, Y \leq y)$$

ist die **zweidimensionale Verteilungsfunktion** des Zufallsvektors (X, Y)

- $F(x, \infty)$ ist die **Randverteilung von** X,
- $F(\infty, y)$ ist die **Randverteilung von** Y

mit

$$\begin{aligned} F(x, \infty) &= P(X \leq x) =: F_X(x)\,, \\ F(\infty, y) &= P(Y \leq y) =: F_Y(y)\,. \end{aligned}$$

Falls (X, Y) **diskret**, d. h. nimmt höchstens abzählbar viele verschiedene Wertepaare an, heißt $f(x_i, y_j) = P(X = x_i, Y = y_j)$ **Wahrscheinlichkeitsfunktion** von (X, Y).

(X, Y) heißt **stetig**, falls

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t)\, ds\, dt$$

mit $f(x, y) \geq 0$ für alle x, y. f heißt *Dichte* von (X, Y).

Eigenschaften einer zweidimensionalen Verteilungsfunktion

- $0 \leq F(x, y) \leq 1$,
- $\lim\limits_{x \to -\infty} F(x, y) = 0\,, \quad \lim\limits_{y \to -\infty} F(x, y) = 0\,, \quad \lim\limits_{\substack{x \to \infty \\ y \to \infty}} F(x, y) = 1$,
- Monotonie: $x_1 \leq x_2\,,\ y_1 \leq y_2 \ \Rightarrow\ F(x_1, y_1) \leq F(x_2, y_2)$,
- rechtsseitig stetig: $F(x, y) = F(x + 0, y) = F(x, y + 0) = F(x + 0, y + 0)$,
- $x_1 < x_2\,,\ y_1 < y_2$:

$$P(x_1 < X \leq x_2,\, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0.$$

Bedingte Verteilungen

Voraussetzung: f Wahrscheinlichkeitsfunktion/Dichte von (X, Y), f_X, f_Y jeweils die Wahrscheinlichkeitsfunktion/Dichte von X, Y (Randverteilung):

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)} \quad \text{für} \quad f_Y(y) > 0$$

heißt „bedingte Wahrscheinlichkeitsfunktion/Dichte von X, wenn Y den Wert y annimmt“, analog $f(y \mid x)$.

Unabhängigkeit von Zufallsvariablen

Definition: Zufallsvariablen $X_1, \ldots, X_n$ heißen **unabhängig**, falls für alle $x_1, \ldots, x_n$

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i).$$

Äquivalente Definitionen:

- diskrete ZV: für alle $x_1, \ldots, x_n$ gilt
  $$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$
- stetige ZV: für alle $x_1, \ldots, x_n$ gilt
  $$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$
- 2-dim. Zufallsvariable, f_X, f_Y Wahrscheinlichkeitsfunktion/Dichte:
  $$\text{für alle } x, y \quad \text{mit} \quad f_Y(y) > 0 \quad \text{gilt} \quad f(x \mid y) = f_X(x)$$

Summen von Zufallsvariable bei Unabhängigkeit

Voraussetzung: X, Y unabhängige Zufallsvariable

- diskret: $P(X + Y = z) = \sum_t P(X = z - t) P(Y = t)$
- stetig: $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z - t)\, f_Y(t)\, dt$

2.2.5 Erwartungswert einer Zufallsvariable

Definition:

$$E(X) = \begin{cases} \sum_i x_i\, P(X = x_i) & X \text{ diskrete Zufallsvariable} \\ \int_{-\infty}^{\infty} x\, f(x)\, dx & X \text{ stetige Zufallsvariable} \end{cases}$$

Eigenschaften:

- f sei Wahrscheinlichkeits- oder Dichtefunktion: Ist f symmetrisch bzgl. m, d. h. $f(m + x) = f(m - x)$ für alle x, so ist $E(X) = m$, falls existent.
- Es sei $Y = h(X)$.

  Ist X diskret, so gilt (falls existent) $E(Y) = E\big(h(X)\big) = \sum_i h(x_i)\, f(x_i)\,.$

  Ist X stetig, so gilt (falls existent) $E(Y) = E\big(h(X)\big) = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx\,.$

  Beispiel: $Y = a\,X + b$: $\qquad E\big(a\,X + b\big) = a\,E(X) + b\,.$
- Es sei $Y = h(X)$ und $h(\cdot)$ konvexe Funktion. Dann ist $E\big(h(X)\big) \geq h\big(E(X)\big)$ – folgt über die JENSENsche Ungleichung.

  Beispiel: $E(X^2) \geq E(X)^2\,.$

- Sind $X_1, \ldots, X_n$ Zufallsvariable mit jeweils existierendem Erwartungswert, so ist

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)\,.$$

  Sind die Zufallsvariablen $X_1, \ldots, X_n$ zusätzlich noch unabhängig, so ist auch

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)\,.$$

2.2.6 Varianz

Voraussetzung: Zufallsvariable X mit Verteilungsfunktion F

Definition: Varianz von X ist

$$Var(X) = E\big([X - E(X)]^2\big)\,.$$

Man schreibt auch $\sigma^2 = Var(X)$.

$\sigma = \sqrt{Var(X)}$ heißt **Standardabweichung.**

Eigenschaften

- Verschiebungssatz (mit $\mu := E(X)$):

$$Var(X) = E\big([X - \mu]^2\big) = E(X^2) - \mu^2\,.$$

- Für alle $a, b \in \mathbb{R}$ ist

$$Var(aX + b) = a^2 Var(X)\,.$$

- Sind die Zufallsvariablen $X_1, \ldots, X_n$ unkorreliert und existieren jeweils die Varianzen, so ist

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)\,.$$

- Für alle $a, b \in \mathbb{R}$ und zwei **beliebige** Zufallsvariablen X und Y gilt

$$Var\big(aX + bY\big) = a^2 Var(X) + 2ab\, Cov(X, Y) + b^2 Var(Y)\,.$$

  Für $Cov()$ siehe unten. Für unkorrelierte X und Y ist natürlich $Cov(X, Y) = 0$.

2.2.7 Kovarianz

Voraussetzung: Zufallsvariablen X und Y mit Verteilungsfunktionen F_X und F_Y

Definition: Kovarianz von X und Y ist

$$Cov(X, Y) = E\big([X - E(X)][Y - E(Y)]\big)\,.$$

Man schreibt auch $\sigma_{XY} = Cov(X, Y)$.

Eigenschaften

- „Verschiebungssatz“:

$$Cov(X,Y) = E\big([X - E(X)][Y - E(Y)]\big) = E(XY) - E(X)\,E(Y)\,.$$

- Für alle $a, b, c, d \in \mathbb{R}$ ist

$$Cov(aX + b, cY + d) = acCov(X,Y)\,.$$

- Cauchy-Schwarzsche Ungleichung:

$$Cov(X,Y)^2 \leq Var(X) \cdot Var(Y)\,.$$

- Lineare Korrelation (siehe auch Bravais-Pearsonscher Korrelationskoeffizient):

$$\varrho = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}\,.$$

- Für $\varrho = 0$ sagt man, dass X und Y unkorreliert sind.

2.2.8 Tschebyschoff-Ungleichung

Voraussetzung: Zufallsvariable X mit der Verteilungsfunktion F mit endlicher Varianz, d. h. $0 < Var(X) < \infty$.

Es gilt für alle $\varepsilon > 0$:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

bzw. äquivalent

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}\,.$$

2.2.9 Zentraler Grenzwertsatz

Zentraler Grenzwertsatz (ZGWS): Sei $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen, die alle unabhängig und identisch verteilt sind, d. h. es gilt insbesondere $E(X) = \mu$ und $Var(X) = \sigma^2$ (mit $0 < \sigma^2 < \infty$) für alle X_i, $i = 1, 2, \ldots$. Sei weiter $\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$ das Stichprobenmittel (arithmetisches Mittel). Dann gilt:

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z) \qquad \text{(oder auch } \lim_{n\to\infty} P\left(\frac{\sum X_i - n\,\mu}{\sqrt{n}\,\sigma} \leq z\right) = \Phi(z)\text{)}\,,$$

wobei $\Phi(z)$ die Verteilungsfunktion der Standard-Normalverteilung ist (siehe Normalverteilung allgemein auf Seite 43 und Bedeutung sowie Tabellierung von Φ auf 47 f.).

Spezialfall ZGWS Moivre-Laplace

Vor.: unabhängige $X_i \sim B(1,p)$, $\sum_{i=1}^{n} X_i =: Y_n \sim B(n,p)$ sowie $a, b \in \mathbb{N}$ mit $a < b$ und n groß

$$P(a < Y_n \leq b) \approx \Phi\left(\frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a + 0.5 - np}{\sqrt{np(1-p)}}\right).$$

Bemerkung: Der Term „+0.5“ heißt *Stetigkeitskorrektur* und dient der Erhöhung der Approximationsgüte.

2.2.10 Grenzwertsatz von Poisson

$Y_n \sim B(n,p)$: Mit $p \to 0$ und $np \to \lambda > 0$ für $n \to \infty$ konvergiert Y_n gegen $Z \sim Pois(\lambda)$.

2.2.11 Approximationen

Verteilung	Approximation durch	Bedingung (Faustregel)
$H(N,M,n)$	$B(n,p)$, $p = \frac{M}{N}$	$\frac{n}{N} < 0.05$
$H(N,M,n)$	$Pois(\lambda)$, $\lambda = n\frac{M}{N}$	$n > 10$, $\frac{M}{N} < 0.05$, $\frac{n}{N} < 0.05$
$B(n,p)$	$\mathcal{N}(\mu,\sigma^2)$, $\mu = np$, $\sigma^2 = np(1-p)$	$np(1-p) \geq 9$
$B(n,p)$	$Pois(\lambda)$, $\lambda = np$	$n \geq 20$, $p \leq 0.1$, $np \leq 10$
$Pois(\lambda)$	$\mathcal{N}(\mu,\sigma^2)$, $\mu = \lambda$, $\sigma^2 = \lambda$	$\lambda \geq 9$
χ^2_n	$\left(\mathcal{N}(\mu,\sigma^2)\right)^2/2$, $\mu = \sqrt{2n-1}$, $\sigma^2 = 1$	$n > 50$
χ^2_n	$\mathcal{N}(\mu,\sigma^2)$, $\mu = n$, $\sigma^2 = 2n$	$n > 200$
t_n	$\mathcal{N}(\mu,\sigma^2)$, $\mu = 0$, $\sigma^2 = n/(n-2)$	$n > 30$

2.2.12 Summen von Zufallsvariablen

Vor.: $X_1, \ldots, X_n$ unabhängige Zufallsvariable.

Verteilung Summand X_i	**Verteilung Summe $\sum X_i$**
$\mathcal{N}(\mu_i,\sigma_i^2)$	$\mathcal{N}\left(\sum \mu_i, \sum \sigma_i^2\right)$
für $\mathcal{N}(\mu,\sigma^2)$	speziell $\bar{X} \sim \mathcal{N}(\mu,\sigma^2/n)$
$B(1,p)$	$B(n,p)$
$B(n_i,p)$	$B\left(\sum n_i, p\right)$
$Exp(\lambda)$	$Erlang(n,\lambda) = \Gamma(n,\lambda)$

3 Schließende Statistik

3.1 Zufallsstichproben

Ein Merkmal sei auf einer Grundgesamtheit gemäß einer Zufallsvariablen X verteilt. Aus dieser Grundgesamtheit wird eine Stichprobe von n Elementen zufällig ausgewählt. Die Ergebnisse $(x_1, x_2, \ldots, x_n)$ sind dann Realisierungen der n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$. Alle Zufallsvariablen X_i, $i = 1, 2, \ldots, n$, haben dieselbe Verteilung wie X. Man spricht daher auch von einer Stichprobe aus (der Verteilungsfunktion) F_X. Da alle Stichprobenvariablen X_i, $i = 1, 2, ..., n$, identisch verteilt sind, haben sie insbesondere auch alle den Erwartungswert $E(X)$ und die Varianz $Var(X)$. Weiter sind wegen der zufälligen Auswahl der Elemente aus der Grundgesamtheit die Zufallsvariablen $X_1, X_2, \ldots, X_n$ alle unabhängig voneinander. Eine solche n-dimensionale Zufallsvariable $(X_1, X_2, \ldots, X_n)$ nennen wir eine **Zufallsstichprobe**.

3.1.1 Stichprobenfunktionen bzw. Statistiken

Eine Funktion g einer Stichprobe, d. h. eine Abbildung

$$(X_1, X_2, \ldots, X_n) \longrightarrow g(X_1, X_2, \ldots, X_n)$$

nennt man Stichprobenfunktion bzw. **Statistik**.

3.1.2 Spezielle Stichprobenfunktionen

Sei $(X_1, X_2, \ldots, X_n)$ eine Stichprobe aus X mit $E(X) = \mu$ und $Var(X) = \sigma^2$:

Stichprobenfunktion	Erwartungswert	Varianz
Summe $\sum_{i=1}^{n} X_i$	$n\mu$	$n\sigma^2$
Stichprobenmittel $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	μ	$\frac{\sigma^2}{n}$
Standardisiertes Stichprobenmittel $\frac{\bar{X}-\mu}{\sigma}\sqrt{n}$	0	1
Empirische Varianz $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$	$\frac{n-1}{n}\sigma^2$	$\frac{(n-1)\,2\sigma^4}{n^2}$
Stichprobenvarianz $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$	σ^2	$\frac{2\sigma^4}{n-1}$

3.2 Punktschätzverfahren

3.2.1 Eigenschaften von Schätzern

Seien $\hat{\vartheta}$, $\hat{\vartheta}_1$, $\hat{\vartheta}_2$ Schätzer für einen Parameter ϑ und n die Größe der Stichprobe. Dann heißt:

$\hat{\vartheta}$ erwartungstreu (unverzerrt) für ϑ,	wenn $E(\hat{\vartheta}) = \vartheta$,
$\hat{\vartheta}$ asymptotisch erwartungstreu für ϑ,	wenn $\lim\limits_{n\to\infty} E(\hat{\vartheta}) = \vartheta$,
$\hat{\vartheta}_1$ wirksamer als $\hat{\vartheta}_2$ ($E(\hat{\vartheta}_1) = E(\hat{\vartheta}_2) = \vartheta$),	wenn $Var(\hat{\vartheta}_1) < Var(\hat{\vartheta}_2)$,
$\hat{\vartheta}_1$ wirksamst bzw. effizient ($E(\hat{\vartheta}_1) = \vartheta$),	wenn $Var(\hat{\vartheta}_1) = \inf\limits_{\hat{\vartheta}:E(\hat{\vartheta})=\vartheta} Var(\hat{\vartheta})$,
$\hat{\vartheta}$ konsistent für ϑ,	wenn für $\varepsilon > 0$ gilt: $\lim\limits_{n\to\infty} P\big(\lvert\hat{\vartheta} - \vartheta\rvert > \varepsilon\big) = 0$.

Insbesondere ist $\hat{\vartheta}$ konsistent, wenn $E(\hat{\vartheta}) \to \vartheta$ und $Var(\hat{\vartheta}) \to 0$ für $n \to \infty$.

3.2.2 Konstruktion von Schätzern

Maximum-Likelihood-Methode

Sei $(X_1, X_2, \dots, X_n)$ eine Zufallsstichprobe aus X, d. h. alle X_i haben dieselbe Wahrscheinlichkeitsfunktion bzw. Dichtefunktion $f_{\boldsymbol{\vartheta}}(x)$, wobei $\boldsymbol{\vartheta} = (\vartheta_1, ..., \vartheta_k)$ der Vektor k unbekannter Verteilungsparameter ist. Dann heißt die Funktion

$$L\big(x_1, \dots, x_n; \boldsymbol{\vartheta}\big) = \prod_{i=1}^{n} f_{\boldsymbol{\vartheta}}(x_i) = f_{\boldsymbol{\vartheta}}(x_1) \cdot f_{\boldsymbol{\vartheta}}(x_2) \cdot \ldots \cdot f_{\boldsymbol{\vartheta}}(x_n)$$

Likelihoodfunktion und ihr (natürlicher) Logarithmus

$$\ln L\big(x_1, \dots, x_n; \boldsymbol{\vartheta}\big) = \sum_{i=1}^{n} \ln f_{\boldsymbol{\vartheta}}(x_i)$$

Loglikelihoodfunktion. Als Schätzer $\hat{\boldsymbol{\vartheta}}$ für $\boldsymbol{\vartheta}$ wird dann der Vektor

$$\hat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}}{\operatorname{argmax}}\ \ln L\big(x_1, \dots, x_n; \boldsymbol{\vartheta}\big) \qquad \text{verwendet.}$$

Momenten-Schätzmethode

Unter der Voraussetzung, dass die theoretischen Momente Funktionen der Parameter ϑ sind, werden die empirisch beobachteten Momente mit den theoretischen Momenten gleichgesetzt und, falls möglich, nach den Parametern aufgelöst:

$$E\big(X^k\big) = \overline{X^k}.$$

Beispiel: Gleichverteilung auf $[a, b]$ – Schätzung von a und b.

$$E(X) = \frac{a+b}{2}\,, \quad E(X^2) = Var(X) + E(X)^2 = \ldots = \frac{a^2 + ab + b^2}{3}\,,$$
$$\bar{X} = \frac{\hat{a} + \hat{b}}{2}\,, \quad \overline{X^2} = \frac{\hat{a}^2 + \hat{a}\hat{b} + \hat{b}^2}{3}\,,$$
$$\hat{b} = \bar{X} + \sqrt{3\big(\overline{X^2} - \bar{X}^2\big)} = \bar{X} + \sqrt{3}\,\tilde{S} \quad , \quad \hat{a} = \bar{X} - \sqrt{3}\,\tilde{S}\,.$$

3.3 Intervallschätzung (Konfidenzintervalle)

3.3.1 Definitionen

Aus einer Zufallsstichprobe $(X_1, X_2, \dots, X_n)$ werden zwei Stichprobenfunktionen $\hat{\vartheta}_u = T_1(X_1, X_2, \dots, X_n)$ und $\hat{\vartheta}_o = T_2(X_1, X_2, \dots, X_n)$ so bestimmt, dass gilt

$$P_\vartheta\big(\hat{\vartheta}_u \leq \vartheta \leq \hat{\vartheta}_o\big) \geq 1 - \alpha\,,$$

wobei $0 < \alpha < 1$. Solch ein Intervall $[\hat{\vartheta}_u, \hat{\vartheta}_o]$ nennt man **zweiseitiges** $(1-\alpha)$-Konfidenzintervall, α die Fehlerwahrscheinlichkeit und $(1-\alpha)$ das Konfidenzniveau.

$[\hat{\vartheta}_u, \infty) = \big[T_1(X_1, X_2, \dots, X_n), \infty\big)$ heißt **einseitiges unteres Konfidenzintervall** zum Niveau $1 - \alpha$, falls gilt

$$P_\vartheta\big(\hat{\vartheta}_u \leq \vartheta\big) \;\geq\; 1 - \alpha \quad \forall\, \vartheta \in \Theta\,.$$

$(-\infty, \hat{\vartheta}_o] = \big(-\infty, T_2(X_1, X_2, \dots, X_n)\big]$ heißt **einseitiges oberes Konfidenzintervall** zum Niveau $1 - \alpha$, falls gilt

$$P_\vartheta\big(\vartheta \leq \hat{\vartheta}_o\big) \;\geq\; 1 - \alpha \quad \forall\, \vartheta \in \Theta\,.$$

3.3.2 Spezielle Konfidenzintervalle

Ein Überblick ist auf Seite 44 gegeben. Einseitige Konfidenzintervalle lassen sich aus den dort gegebenen zweiseitigen ableiten.

3.4 Hypothesentests

Ein Verfahren, das auf der Basis einer Stichprobe eine Entscheidung über Annahme oder Ablehnung einer Hypothese erlaubt, heißt statistischer Test bzw. (kurz) Test. Zur genauen Formulierung eines Testproblems gehören eine sogenannte Nullhypothese (H_0) und eine Alternativhypothese (H_1).

Bei parametrischen Tests geht es um die Entscheidung, ob ein Parameter ϑ einer Verteilung in einem gegebenen Parameterbereich Θ_0 liegt oder eben in einem anderen Parameterbereich Θ_1. Dabei gilt, natürlich, $\Theta_0 \cap \Theta_1 = \emptyset$ und häufig (aber nicht immer) $\Theta_0 \cup \Theta_1 = \Theta$. Ein Testproblem beginnt mit

$$H_0 : \vartheta \in \Theta_0 \quad \text{gegen} \quad H_1 : \vartheta \in \Theta_1\,.$$

Terminologie des Testens

- Folgende Fehler bei der Entscheidung können begangen werden:

  Fehler 1. Art: Entscheide mich für H_1, aber H_0 ist richtig.

  Fehler 2. Art: Entscheide mich für H_0, aber H_1 ist richtig.

Entscheidung	Realität	
	$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
H_0 nicht verwerfen	kein Fehler	Fehler 2.Art
H_1 annehmen	Fehler 1.Art	kein Fehler

- Gütefunktion $G : \Theta \longrightarrow [0,1]$ gibt die Wahrscheinlichkeit $G(\vartheta)$ an, H_1 anzunehmen, falls der Parameter den Wert ϑ besitzt

  Anforderung: Gütefunktion für $\vartheta \in \Theta_0$ klein und für $\vartheta \in \Theta_1$ groß.

- **Teststrategie:** Gebe obere Schranke α für die maximale Wahrscheinlichkeit des Fehlers 1. Art vor – übliche Werte sind $\alpha \in \{0.01, 0.05, 0.1\}$. α nennt man auch das nominale **Signifikanzniveau** des anzuwendenden Tests,

  Bezeichnung: **α-Niveau-Test für** H_0, falls die Wahrscheinlichkeit für einen Fehler 1. Art stets kleiner oder gleich α ist.

  Achtung: Bei einem α-Niveau-Test wird nur der Fehler 1. Art kontrolliert, nicht der Fehler 2. Art!

- **Hinweis:** Diejenige Hypothese, die bestätigt werden soll, muss als H_1 gewählt werden. Aber **Vorsicht:** Tests der Form $H_0 : \mu \neq \mu_0 \quad \leftrightarrow \quad H_1 : \mu = \mu_0$ sprengen den Rahmen der Grundausbildung und sollten daher vorerst vermieden werden.

3.4.1 Konstruktion eines Tests

(i) Hypothese(n) aufstellen.

(ii) Fehlerwahrscheinlichkeit 1. Art α festlegen.

(iii) Teststatistik T identifizieren.

(iv) Ablehnungsbereich C so festlegen, dass $P_{H_0}(T \in C) \leq \alpha$.

(v) Entscheidung treffen: H_0 wird verworfen, wenn der Wert der Teststatistik T bei der vorliegenden Stichprobe im Ablehnungsbereich C liegt.

3.4.2 Verwerfungsbereiche von Tests auf Parameter ϑ

Bezeichne T die Teststatistik eines Tests auf einen Parameter ϑ und q_α das α-Quantil der Verteilung der Teststatistik.

H_0	H_1	H_0 ablehnen, wenn
$\vartheta \leq \vartheta_0$	$\vartheta > \vartheta_0$	$T > c = q_{1-\alpha}$
$\vartheta \geq \vartheta_0$	$\vartheta < \vartheta_0$	$T < c = q_\alpha$
$\vartheta = \vartheta_0$	$\vartheta \neq \vartheta_0$	$T < c_1 = q_{\alpha/2}$ oder $T > c_2 = q_{1-\alpha/2}$

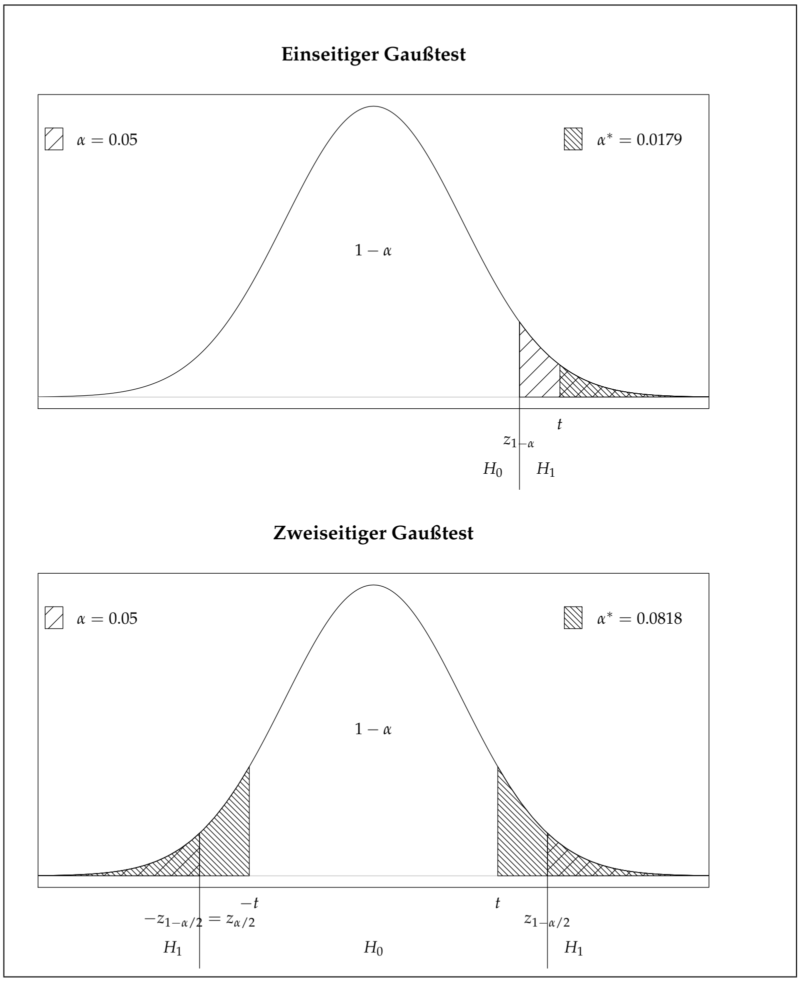
3.4.3 p-Wert

Begriffsbestimmung: p-Wert bzw. empirisches Signifikanzniveau α^* ist die Wahrscheinlichkeit, mit der die Prüfgröße T einen Wert annimmt, der unter H_0 genauso unplausibel wie oder noch unplausibler ist als der beobachtete Wert t.

$\rightsquigarrow$ Je kleiner α^*, desto unplausibler ist H_0, d. h. desto „signifikanter" ist die Abweichung der Stichprobe von H_0.

Vorgehen (bei Statistiksoftware):

1. α festlegen,
2. Software angewandt auf Daten liefert α^*,
3. $\alpha^* < \alpha \Rightarrow H_0$ verwerfen bzw. $\alpha^* \geq \alpha \Rightarrow H_0$ nicht verwerfen.



3.5 Klassische Parametertests

Übersichten über Ein- und Zweistichprobentests sind auf den Seiten 45 und 46 gegeben. Gewisse Spezialfälle seien hier behandelt.

Zweistichprobentest bei verbundenen Stichproben

Voraussetzung: X und Y sind abhängig, z. B. X_i und Y_i werden am selben Merkmalsträger gemessen, Normalvert., Variablen innerhalb der Teilstichproben $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$ unabhängig.

Hypothesen: $H_0\colon \mu_X = \mu_Y \quad \leftrightarrow \quad H_1\colon \mu_X \neq \mu_Y$

Vorgehen: Betrachten $D_i = X_i - Y_i$ für $i = 1, 2, \dots, n$

Daher:

$$T = \frac{\bar{D}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}}\sqrt{n} \sim t_{n-1}$$

Bemerkung: $(n-1)\widehat{Var}(D) = (n-1)\left(\widehat{Var}(X) + \widehat{Var}(Y) - 2\widehat{Cov}(X,Y)\right)$

$\rightsquigarrow$ Wären die Stichproben unabhängig ($\rightarrow Cov(X,Y) = 0$), so wäre $\widehat{Cov}(X,Y)$ rund 0 und damit wäre man beim unverbundenen t-Test (siehe Seite 46). Würde man fälschlich den unverbundenen t-Test benutzen, so würde man beim Schätzen der Varianz einen systematischen Fehler in der Größenordnung von $2\widehat{Cov}(X,Y)$ machen.

Zweistichprobentest mit Nichtstandardhypothesen

In der Liste auf Seite 46 werden Hypothesen der Form

$$H_0\colon \mu_X = \mu_Y \quad \leftrightarrow \quad H_1\colon \mu_X \neq \mu_Y\,,$$
$$H_0\colon \sigma_X^2 = \sigma_Y^2 \quad \leftrightarrow \quad H_1\colon \sigma_X^2 \neq \sigma_Y^2$$

inkl. der jeweiligen einseitigen Varianten behandelt. Etwas allgemeiner sind

$$H_0\colon \mu_X - \mu_Y = \Delta \quad \leftrightarrow \quad H_1\colon \mu_X - \mu_Y \neq \Delta\,,$$
$$H_0\colon \sigma_X^2/\sigma_Y^2 = \delta \quad \leftrightarrow \quad H_1\colon \sigma_X^2/\sigma_Y^2 \neq \delta$$

mit vorgegebenen Werten Δ bzw. δ. Mit den neuen Prüfgrößen

$$T = \frac{(\bar{X} - \bar{Y}) - \Delta}{\text{„Nenner“}} \quad \text{bzw.} \quad T = \frac{S_X^2}{\delta S_Y^2}$$

sind die Spalten „Verteilung von T unter H_0“ und „kritische Werte, d. h. H_0 ablehnen falls“ wieder anwendbar. „Nenner“ steht hier stellvertretend für die Version des jeweiligen Tests. Für $\Delta = 0$ bzw. $\delta = 1$ ergeben sich jeweils die Standardfälle der Tabelle auf Seite 46.

3.6 Tests für Abhängigkeit bzw. Korrelationsmaße

Voraussetzung: Es seien X und Y zwei Merkmale.

Testproblem: $H_0\colon$ X und Y unabhängig $\quad\leftrightarrow\quad$ $H_1\colon$ X und Y nicht unabhängig

Gegeben: Realisierungen (x_i, y_i), $i = 1, \ldots, n$ einer unabhängigen Zufallsstichprobe (X_i, Y_i).

3.6.1 Normalverteilung: Test auf lineare Korrelation

Voraussetzung: (X,Y) 2-dim. normalverteilt mit

$$E(X) = \mu_X\,,\ Var(X) = \sigma_X^2\,,\ E(Y) = \mu_Y\,,\ Var(Y) = \sigma_Y^2\,,\ Corr(X,Y) = \varrho\,.$$

Man *beachte*, dass (nur) bei Normalverteilung gilt: X und Y unabhängig $\quad\Leftrightarrow\quad \varrho = 0$

Schätzer für ϱ: empirischer Korrelationskoeffizient von Bravais-Pearson $\hat{\varrho} = \frac{S_{XY}}{S_X S_Y}$, wobei S_{XY}, S_X, S_Y jeweils Stichprobenkovarianz und Stichprobenstandardabweichung für X bzw. Y sind.

Hypothesen: $H_0\colon\ \varrho = 0 \quad\leftrightarrow\quad H_1\colon\ \varrho \neq 0$

Prüfgröße $T = \sqrt{n-2}\frac{\hat{\varrho}}{\sqrt{1-\hat{\varrho}^2}}$ $\quad$ unter H_0 ist $T \sim t_{n-2}$, d. h. für $|T| > t_{n-2;1-\alpha/2}$ wird H_0 verworfen.

Bemerkung Für die einseitigen Tests

$H_0\colon\ \varrho = 0 \ \leftrightarrow\ H_1\colon\ \varrho > 0$ und $\quad H_0\colon\ \varrho = 0 \ \leftrightarrow\ H_1\colon\ \varrho < 0$

wird H_0 bei $T > t_{n-2;1-\alpha}$ bzw. $T < t_{n-2;\alpha}$ verworfen.

3.6.2 Rangdaten: Test auf monotonen Zusammenhang

Bezeichnung: $R(X_i)$ – Rang der i-ten Beobachtung innerhalb $X_1, \ldots, X_n$

Voraussetzung: Es gibt keine Bindungen, d. h. jede Beobachtungen tritt genau **ein** Mal auf – siehe auch Seite 14.

Dann ist der Rangkorrelationskoeffizient gegeben über

$$\hat{R} = 1 - \frac{6 \sum\limits_{i=1}^{n} \big(R(X_i) - R(Y_i)\big)^2}{n^3 - n}$$

Hypothesen

$$\begin{array}{llll} H_0: & X \text{ und } Y \text{ unabhängig} & \leftrightarrow & H_1: \text{ zwischen } X \text{ und } Y \text{ monotoner Zusammenhang} \\ H_0: & R = 0 & \leftrightarrow & H_1: R \neq 0 \end{array}$$

Unter H_0 ist $\hat{R}$ approximativ normalverteilt, wobei $E(\hat{R}) = 0$ und $Var(\hat{R}) = 1/(n-1)$.

Vergleich von $T = \sqrt{n-1} \cdot \hat{R}$ für große n (≥ 100) mit den Quantilen der Standardnormalverteilung, sonst entsprechende Tabellen.

Hypothesen	H_0 verwerfen, wenn
$H_0\colon R = 0 \leftrightarrow H_1\colon R \neq 0$	$\lvert T\rvert > z_{1-\alpha/2}$
$H_0\colon R = 0 \leftrightarrow H_1\colon R > (<)\, 0$	$T > z_{1-\alpha}$ $(T < z_\alpha)$

3.6.3 Nominaldaten: χ^2-Unabhängigkeitstest von Pearson

Voraussetzung: X und Y diskret mit nur endlich vielen Ausprägungen $a_1, \ldots, a_k$ bzw. $b_1, \ldots, b_l$.

Hypothesen: H_0: X und Y unabhängig $\leftrightarrow$ H_1: X und Y abhängig

n_{ij} gebe an, wie oft das Paar (a_i, b_j) in der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ auftritt (siehe S. 6 o. 12).

Kontingenztafel der absoluten Häufigkeiten:

X	Y				
	b_1	b_2	$\cdots$	b_l	$\sum_j$
a_1	n_{11}	n_{12}	$\cdots$	n_{1l}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	$\cdots$	n_{2l}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}	$n_{k\bullet}$
$\sum_i$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet l}$	n

Schätzer für $P(X = a_i, Y = b_j)$, $P(X = a_i)$, $P(Y = b_j)$ sind $\dfrac{n_{ij}}{n}$, $\dfrac{n_{i\bullet}}{n}$ bzw. $\dfrac{n_{\bullet j}}{n}$.

Bezeichnung: $\hat{n}_{ij} := \frac{n_{i\bullet} n_{\bullet j}}{n}$

Indifferenztabelle: Ersetze in obiger Kontingenztafel die n_{ij} durch $\hat{n}_{ij}$ bei fixierten Randhäufigkeiten.

Unter H_0 ist $\hat{n}_{ij} \approx n_{ij}$ (vgl. Seite 12).

Prüfgröße: $T = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{l} \dfrac{\left(n_{ij} - \hat{n}_{ij}\right)^2}{\hat{n}_{ij}}$.

Unter H_0 ist $\lim_{n\to\infty} P(\chi^2 \leq y) = F_{\chi^2_{(k-1)(l-1)}}(y)$

mit $F_{\chi^2_{(k-1)(l-1)}}(y)$ als Verteilungsfunktion einer χ^2-Verteilung mit $(k-1)(l-1)$ Freiheitsgraden.

Entscheidungsregel: Ist $T > \chi^2_{(k-1)(l-1);1-\alpha}$, so wird H_0 abgelehnt.

Bemerkungen

1. Bequemere Schreibweise: $T = n\left(\sum_{i=1}^{k}\sum_{j=1}^{l} \frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1\right)$

2. Vereinfachte Berechnung für Vierfeldertafeln: $\chi^2 = \frac{n\left(n_{11}n_{22}-n_{12}n_{21}\right)^2}{n_{1\bullet}n_{\bullet 1}n_{2\bullet}n_{\bullet 2}}$

Faustregel: Es sollte $\frac{n_{i\bullet}n_{\bullet j}}{n} \geq 5$ für **alle** i, j sein, d. h. $\hat{n}_{ij} \geq 5$. Ist für ein oder mehrere (i, j) dies verletzt, so sollten Zeilen bzw. Spalten der Kontingenztafel geeignet zusammengelegt werden, so dass die Bedingung dann erfüllt ist.

Mit Klasseneinteilung kann dieser Test auch auf stetige Zufallsvariable angewendet werden.

3.6.4 Vierfeldertafeln: Exakter Test von Fisher (1970)

Voraussetzung: Merkmale X und Y haben nur jeweils zwei Ausprägungen

Bemerkung: Es läßt sich auch einseitig testen, d. h.

Hypothesen H_0: X und Y unabhängig $\leftrightarrow$ H_1: X und Y abhängig
$\leftrightarrow$ H_1': $P(Y = y_1|X = x_1) > (<)\ P(Y = y_1)$

X	Y		
	y_1	y_2	
x_1	m	$n-m$	n
x_2	$M-m$	$N-M-(n-m)$	$N-n$
	M	$N-M$	N

Teststatistik: Anzahl der Paare (x_1, y_1), d. h. $T = m$

Verteilung unter H_0 (Hypergeometrische Verteilung!)

$$P_{H_0}(T = m) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

Wertebereich der Teststatistik T: $\{\underline{m}, \underline{m}+1, \ldots, \overline{m}\}$ mit $\underline{m} = \max\{0, n-(N-M)\}$ und $\overline{m} = \min\{n, M\}$.

„unkritischer" Bereich: $\{k_1, k_1+1, \ldots, k_2\}$ mit $\underline{m} < k_1 < k_2 < \overline{m}$

kritischer Bereich: $K := \{\underline{m}, \underline{m}+1, \ldots, k_1 - 1\} \cup \{k_2+1, k_2+2, \ldots, \overline{m}\}$

$P_{H_0}(T < k_1) \leq \alpha/2$, $P_{H_0}(T > k_2) \leq \alpha/2 \rightsquigarrow$ $P_{H_0}(k_1 \leq T \leq k_2) \geq 1-\alpha$ (Gleichheit lässt sich nur schwerlich erreichen, da Verteilung diskret).

Bestimmung von k_1 und k_2, wobei $H(m)$ die Verteilungsfunktion der Hypergeometrischen Verteilung bezeichnet:

$$\begin{aligned} k_1 = k_1(\alpha) &= \max\left\{m \in \mathbb{N} : P_{H_0}(T < m) \le \alpha/2\right\} \\ &= \max\left\{m \in \mathbb{N} : H(m-1) \le \alpha/2\right\}, \\ k_2 = k_2(\alpha) &= \min\left\{m \in \mathbb{N} : P_{H_0}(T > m) \le \alpha/2\right\} \\ &= \min\left\{m \in \mathbb{N} : H(m) \ge 1-\alpha/2\right\}. \end{aligned}$$

Entscheidungsregel: Ist $T \in K$, so wird H_0 abgelehnt.

Im einseitigen Fall besteht K nur aus $\left\{\underline{m}, \underline{m}+1, \ldots, \tilde{k}_1 - 1\right\}$ oder $\left\{\tilde{k}_2+1, \tilde{k}_2+2, \ldots, \overline{m}\right\}$.

Bemerkung: Im einseitigen Fall ist es bequemer, sofort den p-Wert $\alpha^* = P_{H_0}\big(T \le m\big)$ zu berechnen und diesen dann wie üblich mit dem vorab festgelegten Fehlerniveau α zu vergleichen.

3.7 Anpassungstests

(engl. *goodness-of-fit tests*)

3.7.1 Allgemeines

Voraussetzung: $x_1, \ldots, x_n$ seien die Realisierung einer unabhängigen Zufallsstichprobe $X_1, \ldots, X_n$ mit $X_i \sim F$ (F sei die unbekannte Verteilungsfunktion) für alle $i = 1, \ldots, n$

Testproblem

$$H_0: \; F = F_0 \quad \leftrightarrow \quad H_1: \; F \neq F_0$$

Dabei ist F_0 eine bekannte Verteilungsfunktion.

Achtung: Es wird nicht (bei einem Irrtumsniveau α) bewiesen, dass F_0 die „richtige“ Verteilungsfunktion ist, sondern es wird (nur) geprüft, ob die Daten signifikant gegen F_0 sprechen!

3.7.2 Kolmogoroff-Smirnoff-Test

Voraussetzung: X ist **stetige** Zufallsvariable.

Bezeichnung: empirische Verteilungsfunktion $\hat{F}_n$, d. h. $\hat{F}_n(x)$ = relative Anzahl von Beobachtungen aus $x_1, \ldots, x_n$, die kleiner oder gleich x sind.

Teststatistik: $D = \sup\limits_{x \in \mathbb{R}} \left|\hat{F}_n(x) - F_0(x)\right|$

Die asymptotische Verteilung von D unter H_0 ist tabelliert. **Sie hängt nicht von F_0 ab (wenn F_0 stetig ist)!** c wird als $(1-\alpha)$-Quantil der Verteilung tabelliert (siehe Seite 58).

Testentscheidung: Ist $D > c_{n;1-\alpha}$, so wird H_0 abgelehnt.

Bemerkungen:

1. Kolmogoroff bewies 1933: $\lim\limits_{n \to \infty} P\left(D < \frac{x}{\sqrt{n}}\right) \underset{x>0}{=} \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}$
2. Smirnoff übertrug das 1939 auf zwei emp. Verteilungsfunktionen (Zweistichprobenfall).

Praktische Berechnung von D

$$D = \max_{1 \le i \le n} \{\hat{F}_n(x_i) - F_0(x_i), F_0(x_i) - \hat{F}_n(x_{i-1})\} \ , \ \hat{F}_n(x_0) := 0 \, .$$

R	Excel	SPSS	STATA
`ks.test(x, ...)`	`<addon only>`	`[Analysieren->NP Tests->K-S]`	`ksmirnov`

3.7.3 χ^2-Anpassungstest von Pearson

Voraussetzung: X habe r verschiedene Ausprägungen $t_1, \ldots, t_r$.

Testproblem:

$$H_0 : \ P(X = t_i) = p_i \quad \text{für } i = 1, \ldots, r \quad \leftrightarrow \quad H_1 : \ P(X = t_i) \neq p_i \quad \text{für mindestens ein } i$$

Bezeichnung: N_i = Anzahl der Beobachtungen an, die gleich t_i; $E(N_i) = np_i$ unter H_0.

Prüfgröße: $Q = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i}$.

Es gilt unter der Nullhypothese (Stichprobenumfang n):

$$\lim_{n \to \infty} P(Q \le x) = F_{\chi^2_{r-1}}(x) \, ,$$

wobei $F_{\chi^2_{r-1}}(x)$ die Verteilungsfunktion einer χ^2-Verteilung mit $r - 1$ Freiheitsgraden bezeichnet.

Testentscheidung: Ist $Q > \chi^2_{r-1;1-\alpha}$, so wird H_0 abgelehnt.

Beachte: Damit die asymptotische Verteilungsannahme gerechtfertigt ist, sollte $np_i \ge 5$ (unter H_0) für **alle** i gelten. Falls für ein i die Schranke 5 unterschritten wird $\rightsquigarrow$ Klassenbildung (siehe unten).

Erweiterung auf stetige Verteilungen: (bzw. diskrete Verteilung mit einer großen Zahl von Ausprägungen wie z. B. Poisson-, geometrische Verteilung etc.) Zerlegung des Wertebereichs von X in r disjunkte Klassen $T_1, \ldots, T_r$. Unter H_0 sei $p_i = P(X \in T_i) > 0, \ i = 1, \ldots, r$. N_i gebe die Anzahl der Stichprobenelemente an, die in der Klasse T_i liegen. Die Prüfgröße Q wird wie oben konstruiert. Erneut ist Q unter H_0 approximativ χ^2-verteilt mit $r - 1$ Freiheitsgraden (hier ist r die Klassenanzahl). Die Klassen $T_1, \ldots, T_r$ dürfen nicht von den Daten abhängen.

ACHTUNG: Hängt die Verteilung F_0 von l unbekannten Parametern ab, so werden diese **(mittels der gruppierten Daten)** geschätzt. In der Prüfgröße wird p_i durch $\hat{p}_i$ ersetzt. Die asymptotische Verteilung der Teststatistik Q unter H_0 ist dann χ^2_{r-1-l}.

3.7.4 QQ-Plots

(Quantil-Quantil-Plots)

Voraussetzung: geordnete Zufallsstichprobe $X_{(1)}, \ldots, X_{(n)}$ bzw. Stichprobe $x_{(1)}, \ldots, x_{(n)}$.

Ansatz: Vermutete Basisverteilungsfunktion $F_0(y)$ mit $E(Y) = 0$ und $Var(Y) = 1$, d. h. vermutlich ist $F_X(x) = F_0\big((x - \mu)/\sigma\big)$ mit gegebenenfalls noch zu schätzenden μ und σ.

(i) Ordnungsstatistiken

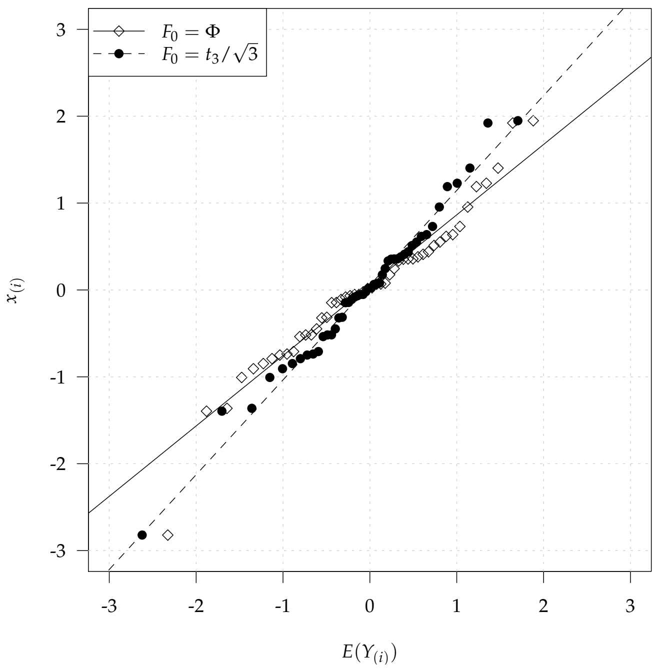
Idee: Vergleich $x_{(i)}$ mit $E(Y_{(i)})$ für alle i

Approximation: $E(Y_{(i)}) \approx F_0^{-1}\big((i - 0.5)/n\big)$

(ii) Quantile

Kollektion von empirischen Quantilen, wie z. B. die 9 Dezile $\tilde{x}_{0.1}, \ldots, \tilde{x}_{0.9}$, und Vergleich mit den theoretischen (also vermuteten), wie z. B. $F_0^{-1}(0.1), \ldots, F_0^{-1}(0.9)$ usw.

Beispiel: *QQ-Plot für die tägliche Aktienrendite eines Monats, F_0 ist entweder die Standardnormalverteilung bzw. eine standardisierte t-Verteilung mit 3 Freiheitsgraden*



Bemerkung: Je näher die Datenpunkte an der Regressionsgeraden liegen, desto besser passt die Verteilung $F_X(x) = F_0\big((x-\mu)/\sigma\big)$ zu den Daten. Darauf basieren auch einige Tests.

3.7.5 Shapiro-Wilk-Test (für Normalverteilung)

Voraussetzung: geordnete Zufallsstichprobe $X_{(1)}, \ldots, X_{(n)}$

Hypothesen: $H_0 \colon X \sim \mathcal{N}(\mu, \sigma^2) \quad \leftrightarrow \quad H_1 \colon X \not\sim \mathcal{N}(\mu, \sigma^2)$

Bemerkung: μ und σ^2 **müssen nicht** bekannt sein.

Technik, Bezeichnungen usw.:

$$\begin{aligned}
\text{geordnete Stichprobe} \quad & z_1 < z_2 < \ldots < z_n \\
& \qquad\qquad \text{standardnormalverteilter Zufallsvariabler}\,, \\
E(z_i) = m_i \quad & \text{– Erwartungswert der } i\text{-ten „Ordnungsstatistik"}\,, \\
Cov(z_i, z_j) = v_{ij} \quad & \text{– Kovarianz zweier solcher}\,, \\
\mathbb{V} = \begin{pmatrix} v_{11} & \ldots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \ldots & v_{nn} \end{pmatrix}\,, \; & \boldsymbol{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}\,, \; \boldsymbol{X} = \begin{pmatrix} X_{(1)} \\ \vdots \\ X_{(n)} \end{pmatrix}\,. \\
\text{Für } X \sim \mathcal{N}(\mu, \sigma^2) \text{ ist} \quad & X_{(i)} = \mu + \sigma m_i + \text{kleiner Fehler}\,,
\end{aligned}$$

Verallgemeinerte MKQ-Schätzung:

$$\hat{\mu} = \bar{X}\,, \; \hat{\sigma} = \frac{\boldsymbol{m}'\,\mathbb{V}^{-1}\boldsymbol{X}}{\boldsymbol{m}'\,\mathbb{V}^{-1}\boldsymbol{m}}\,,$$

$$\text{Hilfsgrößen} \quad \boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \frac{\boldsymbol{m}'\,\mathbb{V}^{-1}}{\left(\boldsymbol{m}'\,\mathbb{V}^{-1}\mathbb{V}^{-1}\boldsymbol{m}\right)^{1/2}}\,,$$

$$\text{Teststatistik} \quad W = \frac{\left(\sum\limits_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum\limits_{i=1}^{n} \left(X_{(i)} - \bar{X}\right)^2} \qquad \left(\approx \frac{\hat{\sigma}^2}{S^2}\right),$$

$$\text{kritischer Bereich} \quad (0, w_\alpha) \quad \rightsquigarrow \quad W < w_\alpha \; \Rightarrow \; H_0 \text{ ist zu verwerfen.}$$

Bemerkung: Die Konstanten m_i, v_{ij} und a_i lassen sich nur numerisch ermitteln. Diese gibt es sowohl tabelliert als auch innerhalb von Statistiksoftware.

R	Excel	SPSS	STATA
`shapiro.test(x)`	`<addon only>`	`[Analysieren->NP Tests->S-W]`	`swilk`

3.8 Tests und mehr im Regressionsmodell

3.8.1 Grundlagen

Regressionsmodell:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i\,, \; i = 1, \ldots, n\,, \\ E(\varepsilon_i) &= 0\,, \\ Var(\varepsilon_i) &= \sigma^2 \quad \text{– Varianz hängt nicht von } i \text{ ab (Homoskedastizität)}\,, \\ Cov(\varepsilon_i, \varepsilon_j) &\underset{i \neq j}{=} 0 \quad \text{– Residuen sind nicht (auto)korreliert} \\ &\left(\varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \text{– normalverteilte Residuen}\right). \end{aligned}$$

MKQ-Schätzer für β_0 und β_1: (siehe auch Seite 15)

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad , \; \hat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} = \begin{cases} S_{xY}/s_x^2 \\ \tilde{S}_{xY}/\tilde{s}_x^2 \end{cases}$$

mit

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{bzw.} \quad \tilde{s}_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2\,,$$

$$S_{xY} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}) \quad \text{bzw.} \quad \tilde{S}_{xY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})\,.$$

Schätzer für σ^2: $\hat{\sigma}^2 = \frac{1}{n-2} \sum\limits_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 = \frac{n-1}{n-2} \left(S_Y^2 - \frac{(S_{xY})^2}{s_x^2}\right) = S_Y^2 (1 - R^2) \frac{n-1}{n-2}\,.$

bzw. $\hat{\sigma}^2 = \frac{n}{n-2} \left(\tilde{S}_Y^2 - \frac{(\tilde{S}_{xY})^2}{\tilde{s}_x^2}\right) = \frac{n}{n-2} (\tilde{S}_Y^2 - \hat{\beta}_1 \tilde{S}_{xY}) = \tilde{S}_Y^2 (1 - R^2) \frac{n}{n-2}\,.$

$\hat{\beta}_0, \hat{\beta}_1$ und $\hat{\sigma}^2$ sind erwartungstreu.

Varianzen von $\hat{\beta}_0, \hat{\beta}_1$ und $\hat{\sigma}^2$:

$$\begin{aligned}
Var(\hat{\beta}_0) &= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right) = \frac{\sigma^2}{n}\left(1 + \frac{\bar{x}^2}{\tilde{s}_x^2}\right),\\
Var(\hat{\beta}_1) &= \frac{\sigma^2}{(n-1)s_x^2} = \frac{\sigma^2}{n\tilde{s}_x^2},\\
Cov(\hat{\beta}_0, \hat{\beta}_1) &= -\frac{\sigma^2\bar{x}}{(n-1)s_x^2} = -\frac{\sigma^2\bar{x}}{n\tilde{s}_x^2},\\
Var(\hat{\sigma}^2) &= \frac{2\sigma^4}{n-2}.
\end{aligned}$$

Somit sind $\hat{\beta}_0, \hat{\beta}_1$ und $\hat{\sigma}^2$ auch konsistent.

3.8.2 Tests bzgl. Absolutglied β_0 und Anstieg β_1

Bemerkung: Um die Varianzen der Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ zu schätzen, ersetzt man in den obigen Formeln einfach σ^2 durch den Schätzer $\hat{\sigma}^2$.

(I) Absolutglied β_0

Hypothesen: $H_0\colon\ \beta_0 = a \quad \leftrightarrow \quad H_1\colon\ \beta_0 \neq a$

Prüfgröße:

$$T_{\beta_0} = \frac{\hat{\beta}_0 - a}{\sqrt{\widehat{Var}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - a}{\hat{\sigma}\sqrt{1/n + \bar{x}^2/\left[(n-1)s_x^2\right]}} = \frac{\hat{\beta}_0 - a}{\hat{\sigma}\sqrt{1 + \bar{x}^2/\tilde{s}_x^2}}\sqrt{n}$$

Bei normalverteilten Residuen, d. h. $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, gilt unter der Nullhypothese $T_{\beta_0} \sim t_{n-2}$.

$\rightsquigarrow \ |T_{\beta_0}| > t_{n-2;1-\alpha/2} \quad \Rightarrow \quad H_0$ wird abgelehnt.

Bemerkung: Sind die Residuen ε_i nicht normalverteilt, so ist T_1 unter H_0 approximativ normalverteilt (meist nutzt man auch hier die t-Verteilung, da diese zu kleineren Ablehnungsbereichen führt). Dabei sollte $\sum_{i=1}^{n}(x_i - \bar{x})^2 \to \infty$ für $n \to \infty$ gelten.

(II) Anstieg β_1

Hypothesen: $H_0\colon\ \beta_1 = b \quad \leftrightarrow \quad H_1\colon\ \beta_1 \neq b$

Prüfgröße:

$$T_{\beta_1} = \frac{\hat{\beta}_1 - b}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - b}{\hat{\sigma}}\sqrt{n-1}s_x = \frac{\hat{\beta}_1 - b}{\hat{\sigma}}\sqrt{n}\tilde{s}_x$$

Bei $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ gilt unter der Nullhypothese $T_{\beta_1} \sim t_{n-2}$.

$\rightsquigarrow \ |T_{\beta_1}| > t_{n-2;1-\alpha/2} \quad \Rightarrow \quad H_0$ wird abgelehnt.

Bemerkung: Wie bei T_{β_0} gilt das für $\varepsilon_i \not\sim \mathcal{N}(0, \sigma^2)$ nur approximativ.

3.8.3 F-Test

Typischer Output von Statistik-Software beim Regressionsmodell ist die **Varianzanalysetabelle** (engl. *ANOVA[analysis of variance] table*).

Streuungstyp	Summe	Freiheitsgrade	Mittlerer quadratischer Fehler	Prüfgröße
Erklärte Streuung	$SQE = \sum(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{Y})^2$	1	$MQE = \frac{SQE}{1}$	$F = \frac{MQE}{MQR}$
Reststreuung	$SQR = \sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$	$n-2$	$MQR = \frac{SQR}{n-2} = \hat{\sigma}^2$	
Gesamtstreuung	$SQT = \sum(Y_i - \bar{Y})^2$	$n-1$	S_Y^2	

Die Tabelle illustriert, welchen Beitrag das Regressionsmodell bei der Streuungszerlegung leistet. Die „Gesamtstreuung“ SQT wird um den Beitrag der „Erklärten Streuung“ SQE zur sogenannten „Reststreuung“ SQR reduziert. Von SQR führt dann der Weg direkt zur Schätzung der Residualvarianz $\sigma^2 = Var(\varepsilon_i)$.Der Quotient F vergleicht unmittelbar Erklärte und Reststreuung. Es lässt sich nun zeigen, dass

$$\begin{aligned}
&F = T_{\beta_1}^2 \quad \text{(siehe Prüfgröße in 3.8.2(II) für } b = 0\text{)}\,, \\
&F = \frac{R^2}{1-R^2}(n-2) \\
\text{und} \quad &F \sim F_{1,n-2} \quad \text{unter } H_0\colon \beta_1 = 0 \ \Leftrightarrow \ \text{einfache lineare Regression nicht nötig.}
\end{aligned}$$

Mittels dieses F-Tests lässt sich prüfen, ob das lineare Regressionsmodell „signifikant“ ist. Dies funktioniert auch bei multiplen Regressionsmodellen (siehe dazu die Matrizenschreibweise auf Seite 15) mit mehr als einem Regressanden . Dann steigt die Zahl der Freiheitsgrade im Zähler auf eben diese Anzahl, während die Freiheitsgrade im Nenner um eben diesen Betrag fallen. Die neue Nullhypothese lautet dann $H_0\colon \beta_1 = \beta_2 = \ldots = \beta_k = 0$ ($k > 1$, $k = 1$ – ein Regressand). Die Teststatistik F ist dann unter H_0 $F_{k,n-k-1}$-verteilt.

3.8.4 Prognose und Prognoseintervall

Prognose für $Y_T = f(x_T)$: $\quad \hat{Y}_T = \hat{\beta}_0 + \hat{\beta}_1 x_T$

Konstruktion eines Prognoseintervalls:

Prognosefehler:

$$\begin{aligned}
\hat{\varepsilon}_T &= Y_T - \hat{Y}_T = \beta_0 + \beta_1 x_T + \varepsilon_T - \left(\hat{\beta}_0 + \hat{\beta}_1 x_T\right) \\
&= \beta_0 - \hat{\beta}_0 + (\beta_1 - \hat{\beta}_1) x_T + \varepsilon_T\,.
\end{aligned}$$

Varianz des Prognosefehlers ($E(\hat{\varepsilon}_T) = 0$):

$$\begin{aligned}
Var(\hat{\varepsilon}_T) &= Var(\hat{\beta}_0) + x_T^2 Var(\hat{\beta}_1) + 2x_T Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\varepsilon_T) \\
Var(\hat{\varepsilon}_T) &= \sigma^2 \left(\frac{1}{n}\left(1 + \frac{\bar{x}^2}{\tilde{s}_x^2}\right) + \frac{x_T^2}{n\tilde{s}_x^2} - 2x_T \frac{\bar{x}}{n\tilde{s}_x^2} + 1 \right) \\
&= \sigma^2 \left(1 + \frac{1}{n}\left(1 + \frac{\left(x_T - \bar{x}\right)^2}{\tilde{s}_x^2}\right)\right)
\end{aligned}$$

$\rightsquigarrow$ *Prognoseintervall für Y_T* bei normalverteilten Residuen ε_i (sonst approximativ):

$$\hat{Y}_T - \hat{\sigma}\sqrt{1 + \frac{1}{n}\left(1 + \frac{(x_T - \bar{x})^2}{\tilde{s}_x^2}\right)}\, t_{n-2;1-\alpha/2} < Y_T < \hat{Y}_T + \hat{\sigma}\sqrt{1 + \frac{1}{n}\left(1 + \frac{(x_T - \bar{x})^2}{\tilde{s}_x^2}\right)}\, t_{n-2;1-\alpha/2}$$

$t_{n-2;1-\alpha/2}$ ist das Quantil einer Studentverteilung mit $n-2$ Freiheitsgraden.

4 Appendix

Im Anhang finden Sie eine Reihe von Tabellen, so

1. einen Überblick über einige Verteilungen,
2. Konfidenzintervalle,
3. Tests,
4. die Verteilungsfunktion der Standardnormalverteilung,
5. die Quantile der Standardnormal-, der studentischen *t*-, der χ^2- und der (Fisher) *F*-Verteilung,
6. kritische Werte des Kolmogoroff-Smirnoff-Tests.

4.1 Diskrete Verteilungen

Verteilung	Wahrscheinlichkeits-funktion $f(m)$	Parameter-raum	Erwartungs-wert $\mu = E(X)$	Varianz $\sigma^2 = E\big([X-\mu]^2\big)$
Binomial[§] $B(n,p)$	$\binom{n}{m} p^m (1-p)^{n-m}$ $m \in \{0,1,\ldots,n\}$	$0 < p < 1$ $n \in \{1,2,\ldots\}$	np	$np(1-p)$
Hyper-geometrisch $H(N,M,n)$	$\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$ $m \in \{m_{min}, m_{min}+1, \ldots, m_{max}\}$, $m_{min} := \max\{0, n-(N-M)\}$, $m_{max} := \min\{n, M\}$	$N \in \{1,2,\ldots\}$, $M \in \{0,1,\ldots,N\}$, $n \in \{1,2,\ldots N\}$	$n\frac{M}{N}$	$n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$ $(N > 1)$
Poisson $Pois(\lambda)$	$\frac{\lambda^m}{m!}e^{-\lambda}$ $m \in \{0,1,\ldots\}$	$\lambda > 0$	λ	λ
Geometrisch $Geom(p)$	$p(1-p)^{m-1}$ $m \in \{1,2,\ldots\}$	$0 < p < 1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

§: Für $n = 1$ nennt man die Binomialverteilung auch Bernoulliverteilung mit $P(X=1) = p = 1 - P(X=0)$, $E(X) = p$, $Var(p) = p(1-p)$.

4.2 Stetige Verteilungen

Verteilung	Wahrscheinlichkeits-dichte $f(x)$	Verteilungs-funktion $F(x)$	Parameter-raum	Erwartungs-wert $\mu = E(X)$	Varianz $\sigma^2 = E\big([X-\mu]^2\big)$
Gleich (Uniform)	$\dfrac{1}{b-a}\,,\ x \in [a,b]$	$\dfrac{x}{b-a}\,,\ x \in [a,b]$	$-\infty < a < b < \infty$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Exponential $Exp(\lambda)$	$\lambda e^{-\lambda x}\,,\ x \geq 0$	$1 - e^{-\lambda x}\,,\ x \geq 0$	$\lambda > 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Gamma	$\dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}\,,\ x \geq 0$	$\dfrac{\gamma(r, \lambda x)}{\Gamma(r)}\,,\ x \geq 0$	$\lambda > 0, r > 0$	$\dfrac{r}{\lambda}$	$\dfrac{r}{\lambda^2}$
Cauchy	$\dfrac{1}{\pi\beta\{1 + [(x-\alpha)/\beta]^2\}}$	$\frac{1}{\pi}\text{atan}\left(\frac{x-\alpha}{\beta}\right) + \frac{1}{2}$	$\beta > 0\,,\ \alpha \in \mathbb{R}$	-	-
Laplace	$\dfrac{\lambda}{2} e^{-\lambda\lvert x-\mu\rvert}$	$\begin{cases} \frac{1}{2} e^{\lambda(x-\mu)} & , x < \mu \\ 1 - \frac{1}{2} e^{-\lambda(x-\mu)} & , x \geq \mu \end{cases}$	$\lambda > 0\,,\ \mu \in \mathbb{R}$	μ	$\dfrac{2}{\lambda^2}$

- Gamma-Funktion $\Gamma(x)$: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$ mit $\Gamma(x+1) = x\Gamma(x)\,,\quad \Gamma(n) = (n-1)!$ für $n \in \mathbb{N}\,,\quad \Gamma(0.5) = \sqrt{\pi}\,.$
- unvollständige Gamma-Funktion der oberen Grenze $\gamma(x,s)$: $\gamma(x,s) = \int_0^s t^{x-1} e^{-t}\, dt$ ... eher von anekdotischem Interesse.

4.2 Stetige Verteilungen II

Verteilung	Wahrscheinlichkeits-dichte $f(x)$	Parameter-raum	Erwartungs-wert $\mu = E(X)$	Varianz $\sigma^2 = E\big([X-\mu]^2\big)$
Normal $\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ ,\ x \in \mathbb{R}$	$\mu \in \mathbb{R}\,,\ \sigma > 0$	μ	σ^2
χ^2_n *	$\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\ ,\ x \geq 0$	$n \in \{1,2,\ldots\}$	n	$2n$
Student** t_n	$\frac{\Gamma\left(\frac{n+1}{2}\right)\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi\, n}}\ ,\ x \in \mathbb{R}$	$n \in \{1,2,\ldots\}$	$0 \quad (n > 1)$	$\frac{n}{n-2} \quad (n > 2)$
Fisher*** $F_{m,n}$	$\frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}\,x\right)^{-\frac{m+n}{2}}\ ,\ x \geq 0$	$m,n \in \{1,2,\ldots\}$	$\frac{n}{n-2}$ $(n > 2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n > 4)$

*: $\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$ mit unabhängigen $X_i \sim \mathcal{N}(0,1)$, $i = 1,2,..,n$ — äquivalent zur Gamma-Verteilung $\Gamma(1/2,n)$.

**: $X/\big(\sqrt{Y/n}\,\big) \sim t_n$ mit unabhängigen $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi^2_n$

***: $\big(X/m\,\big)/\big(Y/n\,\big) \sim F_{m,n}$ mit unabhängigen $X \sim \chi^2_m$ und $Y \sim \chi^2_n$

- $\Gamma()$: Siehe Seite 42.

4.3 Übersicht Konfidenzintervalle

Verteilung, Voraussetzungen	ϑ	$\hat{\vartheta}$	Konfidenzintervall
n groß	μ	$\bar{X}$	$\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$ (oder S statt σ bei unbekanntem σ^2)
$B(1,p)$, $n\hat{p}(1-\hat{p}) > 9$	p	$\hat{p} = \bar{X}$	$\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{1-\alpha/2} < p < \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{1-\alpha/2}$
$B(1,p)$, $n\hat{p}(1-\hat{p}) > 5$	p	$\hat{p} = \bar{X}$	$\frac{n\left[\hat{p} + \frac{1}{2n} z_{1-\alpha/2}^2 - B\right]}{n + z_{1-\alpha/2}^2} < p < \frac{n\left[\hat{p} + \frac{1}{2n} z_{1-\alpha/2}^2 + B\right]}{n + z_{1-\alpha/2}^2}$ mit $B = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \left(\frac{1}{2n} z_{1-\alpha/2}\right)^2}$
$\mathcal{N}(\mu, \sigma^2)$, σ^2 bekannt	μ	$\bar{X}$	$\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$
$\mathcal{N}(\mu, \sigma^2)$, σ^2 unbekannt	μ	$\bar{X}$	$\bar{X} - \frac{S}{\sqrt{n}} t_{n-1;1-\alpha/2} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{n-1;1-\alpha/2}$
$\mathcal{N}(\mu, \sigma^2)$, μ bekannt	σ^2	$\mathring{S}^2$	$\frac{n\mathring{S}^2}{\chi^2_{n;1-\alpha/2}} < \sigma^2 < \frac{n\mathring{S}^2}{\chi^2_{n;\alpha/2}}$, $\mathring{S}^2 = 1/n \sum_{i=1}^{n} \left(X_i - \mu\right)^2$
$\mathcal{N}(\mu, \sigma^2)$, μ unbekannt	σ^2	S^2	$\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}$, $S^2 = 1/(n-1) \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
$\mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right)$	ϱ	r	$r - \frac{1-r^2}{\sqrt{n}} z_{1-\alpha/2} < \varrho < r + \frac{1-r^2}{\sqrt{n}} z_{1-\alpha/2}$, $r \mathrel{\widehat{=}}$ Korrelationskoeffizient nach Bravais-Pearson
			oder $\left[\tanh\left(\tanh^{-1}(r) - \frac{z_{1-\alpha/2}}{\sqrt{n-3}}\right), \tanh\left(\tanh^{-1}(r) + \frac{z_{1-\alpha/2}}{\sqrt{n-3}}\right)\right]$, Z-Transformation nach Fisher

4.4 Übersicht Einstichprobentests

Verteilung, Voraussetzungen	H_0	H_1	Prüfgröße T	Verteilung von T unter H_0	kritische Werte, d. h. H_0 ablehnen, falls
Normalverteilung	$\mu = \mu_0$	$\mu \neq \mu_0$			$T \notin \left[-z_{1-\alpha/2}, z_{1-\alpha/2}\right]$
σ bekannt	$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	Φ	$T > z_{1-\alpha}$
	$\mu \geq \mu_0$	$\mu < \mu_0$			$T < -z_{1-\alpha}$
σ unbekannt	$\mu = \mu_0$	$\mu \neq \mu_0$		t_{n-1}	$T \notin \left[-t_{n-1;1-\alpha/2}, t_{n-1;1-\alpha/2}\right]$
	$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$	bzw. für $n > 30$ approx. Φ	$T > t_{n-1;1-\alpha}$
	$\mu \geq \mu_0$	$\mu < \mu_0$			$T < -t_{n-1;1-\alpha}$
beliebige Verteilung	$\mu = \mu_0$	$\mu \neq \mu_0$			$T \notin \left[-z_{1-\alpha/2}, z_{1-\alpha/2}\right]$
mit Erwartungswert μ	$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$	approx. Φ	$T > z_{1-\alpha}$
	$\mu \geq \mu_0$	$\mu < \mu_0$			$T < -z_{1-\alpha}$
Bernoulliverteilung	$p = p_0$	$p \neq p_0$	Anzahl		$T \notin [c_{\alpha/2}, c_{1-\alpha/2}]$
n klein	$p \leq p_0$	$p > p_0$		$B(n, p_0)$	$T > c_{1-\alpha}$
	$p \geq p_0$	$p < p_0$	der Erfolge		$T < c_\alpha$ ... c_α ist das α-Quantil von $B(n, p_0)$
	$p = p_0$	$p \neq p_0$			$T \notin [-z_{1-\alpha/2}, z_{1-\alpha/2}]$
$np_0(1-p_0) > 9$	$p \leq p_0$	$p > p_0$	$\dfrac{(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)/n}}$	approx. Φ	$T > z_{1-\alpha}$
	$p \geq p_0$	$p < p_0$			$T < -z_{1-\alpha}$
Normalverteilung	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$(n-1)S^2/\sigma_0^2$	χ^2_{n-1}	$T \notin [\chi^2_{n-1;\alpha/2}, \chi^2_{n-1;1-\alpha/2}]$
	$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$			$T > \chi^2_{n-1;1-\alpha}$
	$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(n\mathring{S}^2/\sigma_0^2)^*$	$(\chi^2_n)^*$	$T < \chi^2_{n-1;\alpha}$ $\quad$ (n statt $n-1$)*

*: bei bekanntem Erwartungswert μ , $\quad S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $\quad \mathring{S}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

4.5 Übersicht Zweistichprobentests

Verteilung, Voraussetzungen	H_0	H_1	Prüfgröße T	Verteilung von T unter H_0	kritische Werte, d. h. H_0 ablehnen, falls
Normalverteilung	$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$			$T \notin \left[-z_{1-\alpha/2}, z_{1-\alpha/2}\right]$
σ_X und σ_Y	$\mu_X \leq \mu_Y$	$\mu_X > \mu_Y$	$\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}$	Φ	$T > z_{1-\alpha}$
bekannt	$\mu_X \geq \mu_Y$	$\mu_X < \mu_Y$			$T < -z_{1-\alpha}$
σ_X und σ_Y unbekannt	$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$	$\left(\bar{X}-\bar{Y}\right)/\sqrt{\hat{\sigma}^2\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)}$		$T \notin \left[-t_{n_X+n_Y-2;1-\alpha/2}, t_{n_X+n_Y-2;1-\alpha/2}\right]$
$\sigma_X = \sigma_Y$,	$\mu_X \leq \mu_Y$	$\mu_X > \mu_Y$	mit $\hat{\sigma}^2 = \frac{n_X-1}{n_X+n_Y-2}S_X^2 + \frac{n_Y-1}{n_X+n_Y-2}S_Y^2$	$t_{n_X+n_Y-2}$	$T > t_{n_X+n_Y-2;1-\alpha}$
Stichproben klein	$\mu_X \geq \mu_Y$	$\mu_X < \mu_Y$			$T < -t_{n_X+n_Y-2;1-\alpha}$
σ_X und σ_Y unbekannt	$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$		t_m,	$T \notin \left[-t_{m;1-\alpha/2}, t_{m;1-\alpha/2}\right]$
$\sigma_X \neq \sigma_Y$,	$\mu_X \leq \mu_Y$	$\mu_X > \mu_Y$	$\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}$	$m = \frac{(1+R)^2}{R^2/(n_X-1)+1/(n_Y-1)}$,	$T > t_{m;1-\alpha}$
Stichproben klein	$\mu_X \geq \mu_Y$	$\mu_X < \mu_Y$		$R = \frac{n_Y S_X^2}{n_X S_Y^2}$	$T < -t_{m;1-\alpha}$
Beliebige Verteilungen	$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$			$T \notin \left[-z_{1-\alpha/2}, z_{1-\alpha/2}\right]$
mit σ_X, σ_Y unbekannt,	$\mu_X \leq \mu_Y$	$\mu_X > \mu_Y$	$\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}$	Φ	$T > z_{1-\alpha}$
Stichproben groß ($n_X, n_Y > 30$)	$\mu_X \geq \mu_Y$	$\mu_X < \mu_Y$			$T < -z_{1-\alpha}$
Bernoulliverteilung	$p_X = p_Y$	$p_X \neq p_Y$	$\left(\hat{p}_X-\hat{p}_Y\right)/\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)}$		$T \notin \left[-z_{1-\alpha/2}, z_{1-\alpha/2}\right]$
Stichproben groß ($n_X, n_Y > 30$)	$p_X \leq p_Y$	$p_X > p_Y$	mit $\hat{p} = \frac{n_X\hat{p}_X+n_Y\hat{p}_Y}{n_X+n_Y}$	Φ	$T > z_{1-\alpha}$
	$p_X \geq p_Y$	$p_X < p_Y$			$T < -z_{1-\alpha}$
Normalverteilung	$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$			$T \notin \left[1/F_{n_Y-1,n_X-1;1-\alpha/2}, F_{n_X-1,n_Y-1;1-\alpha/2}\right]$
	$\sigma_X^2 \leq \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$\frac{S_X^2}{S_Y^2}$	F_{n_X-1,n_Y-1}	$T > F_{n_X-1,n_Y-1;1-\alpha}$
	$\sigma_X^2 \geq \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$			$T < 1/F_{n_Y-1,n_X-1;1-\alpha}$

- $S_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ usw. – es wird also stets die Stichprobenvarianz benutzt.
- Siehe auch Seite 31 für sogenannte Nichtstandardhypothesen (z. B. $H_0\colon \mu_X - \mu_Y = \Delta$).

4.6 Verteilungsfunktion der Standardnormalverteilung

$X \sim \mathcal{N}(0,1)$

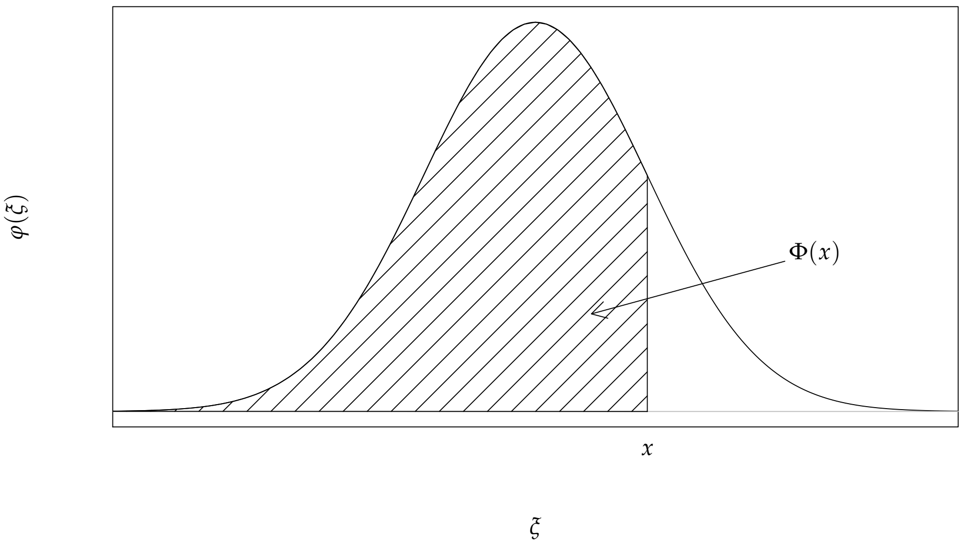
$$P(X \leq x) = \Phi(x) = \int_{-\infty}^{x} \overbrace{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\xi^2}{2}\right)}^{\varphi(\xi)} d\xi\,,$$

$$Y \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad \frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)\,,$$

$$X \sim \mathcal{N}(0,1) \quad \Rightarrow \quad \mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)\,.$$

Rechenregel für negative x:

$$\Phi(x) = 1 - \Phi(-x)$$



Beispiele:

$$\Phi(1.3) = 0.9032\,,$$
$$\Phi(2.64) = 0.9959\,,$$
$$\Phi(-0.83) = 1 - \Phi(0.83) = 1 - 0.7967 = 0.2033\,.$$

Ab $x \geq 3.9$ ist $\Phi(x) \approx 1$ bzw. $x \leq -3.9$ ist $\Phi(x) \approx 0$.

R	Excel	SPSS	STATA
`pnorm(x)`	`NORMVERT(x; 0; 1; WAHR)`	`CDF.NORMAL(x, 0, 1)`	`normal(x)`

x	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
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

R	Excel	SPSS	STATA
`pnorm(x)`	`NORMVERT(x; 0; 1; WAHR)`	`CDF.NORMAL(x, 0, 1)`	`normal(x)`

4.7 Quantile der Standardnormalverteilung

$$X \sim \mathcal{N}(0,1) : z_\alpha = \Phi^{-1}(\alpha) \quad \Leftrightarrow \quad \alpha = \Phi(z_\alpha) = \int_{-\infty}^{z_\alpha} \overbrace{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)}^{\varphi(x)} dx\,,$$

$$Y \sim \mathcal{N}(\mu, \sigma^2) : y_\alpha = \mu + \sigma z_\alpha \quad \rightsquigarrow \quad P(Y \leq y_\alpha) = \alpha\,.$$

Rechenregel für $\alpha < 0.5$:

$$z_\alpha = -z_{1-\alpha}$$

α	z_α	α	z_α	α	z_α	α	z_α
0.9999	3.7190	**0.9975**	**2.8070**	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

R	Excel	SPSS	STATA
`qnorm(alpha)`	`NORMINV(alpha; 0; 1)`	`IDF.NORMAL(alpha, 0, 1)`	`invnorm(alpha)`

4.8 Quantile der t-Verteilung

$$t_{n;\alpha} = F_{t_n}^{-1}(\alpha) \quad \Leftrightarrow \quad \alpha = F_{t_n}(t_{n;\alpha}) = \int_{-\infty}^{t_{n;\alpha}} f_n(\tau)\,d\tau \quad , \; f_n(\tau) = \frac{\Gamma\left(\frac{n+1}{2}\right)\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi\,n}}$$

Rechenregel für $\alpha < 0.5$:

$$t_{n;\alpha} = -t_{n;1-\alpha}$$

	α							
n	0.9	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
1	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	1.638	2.353	3.182	4.541	5.841	7.453	10.22	12.92
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
500	1.283	1.648	1.965	2.334	2.586	2.820	3.107	3.310
∞	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Für große $n > 30$ ist $t_{n;\alpha} \approx \sqrt{\frac{n}{n-2}}z_\alpha$; z_α ist das entsprechende Quantil der Standardnormalverteilung.

z. B. 40	1.315	1.688	2.011	2.387	2.643	2.880	3.171	3.376

R	Excel	SPSS	STATA
`qt(alpha, n)`	`TINV(2*(1-alpha); n)`	`IDF.T(alpha, n)`	`invttail(m, 1-alpha)`

4.9 Quantile der χ^2-Verteilung

$$\chi^2_{n;\alpha} = F^{-1}_{\chi^2_n}(\alpha) \quad \Leftrightarrow \quad \alpha = F_{\chi^2_n}(\chi^2_{n;\alpha}) = \int_{-\infty}^{\chi^2_{n;\alpha}} f_n(c)\, dc \quad , \; f_n(c) = \frac{c^{\frac{n}{2}-1} e^{-c/2}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

$$n = 1, 2, \ldots, 50$$

	α									
n	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.60
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.81	37.16
19	6.844	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.434	8.260	9.591	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.034	8.897	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.643	9.542	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.260	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.886	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
31	14.46	15.66	17.54	19.28	21.43	41.42	44.99	48.23	52.19	55.00
32	15.13	16.36	18.29	20.07	22.27	42.58	46.19	49.48	53.49	56.33
33	15.82	17.07	19.05	20.87	23.11	43.75	47.40	50.73	54.78	57.65
34	16.50	17.79	19.81	21.66	23.95	44.90	48.60	51.97	56.06	58.96
35	17.19	18.51	20.57	22.47	24.80	46.06	49.80	53.20	57.34	60.27
36	17.89	19.23	21.34	23.27	25.64	47.21	51.00	54.44	58.62	61.58
37	18.59	19.96	22.11	24.07	26.49	48.36	52.19	55.67	59.89	62.88
38	19.29	20.69	22.88	24.88	27.34	49.51	53.38	56.90	61.16	64.18
39	20.00	21.43	23.65	25.70	28.20	50.66	54.57	58.12	62.43	65.48
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
41	21.42	22.91	25.21	27.33	29.91	52.95	56.94	60.56	64.95	68.05
42	22.14	23.65	26.00	28.14	30.77	54.09	58.12	61.78	66.21	69.34
43	22.86	24.40	26.79	28.96	31.63	55.23	59.30	62.99	67.46	70.62
44	23.58	25.15	27.57	29.79	32.49	56.37	60.48	64.20	68.71	71.89
45	24.31	25.90	28.37	30.61	33.35	57.51	61.66	65.41	69.96	73.17
46	25.04	26.66	29.16	31.44	34.22	58.64	62.83	66.62	71.20	74.44
47	25.77	27.42	29.96	32.27	35.08	59.77	64.00	67.82	72.44	75.70
48	26.51	28.18	30.75	33.10	35.95	60.91	65.17	69.02	73.68	76.97
49	27.25	28.94	31.55	33.93	36.82	62.04	66.34	70.22	74.92	78.23
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49

Quantile der χ^2-Verteilung – Fortsetzung

$n = 55, \ldots, 1000$

	α									
n	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
55	31.73	33.57	36.40	38.96	42.06	68.80	73.31	77.38	82.29	85.75
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
65	39.38	41.44	44.60	47.45	50.88	79.97	84.82	89.18	94.42	98.11
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
75	47.21	49.48	52.94	56.05	59.79	91.06	96.22	100.8	106.4	110.3
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
85	55.17	57.63	61.39	64.75	68.78	102.1	107.5	112.4	118.2	122.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
95	63.25	65.90	69.92	73.52	77.82	113.0	118.8	123.9	130.0	134.2
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2
110	75.55	78.46	82.87	86.79	91.47	129.4	135.5	140.9	147.4	151.9
120	83.85	86.92	91.57	95.70	100.6	140.2	146.6	152.2	159.0	163.6
130	92.22	95.45	100.3	104.7	109.8	151.0	157.6	163.5	170.4	175.3
140	100.7	104.0	109.1	113.7	119.0	161.8	168.6	174.6	181.8	186.8
150	109.1	112.7	118.0	122.7	128.3	172.6	179.6	185.8	193.2	198.4
160	117.7	121.3	126.9	131.8	137.5	183.3	190.5	196.9	204.5	209.8
170	126.3	130.1	135.8	140.8	146.8	194.0	201.4	208.0	215.8	221.2
180	134.9	138.8	144.7	150.0	156.2	204.7	212.3	219.0	227.1	232.6
190	143.5	147.6	153.7	159.1	165.5	215.4	223.2	230.1	238.3	244.0
200	152.2	156.4	162.7	168.3	174.8	226.0	234.0	241.1	249.4	255.3
220	169.7	174.2	180.8	186.7	193.6	247.3	255.6	263.0	271.7	277.8
240	187.3	192.0	199.0	205.1	212.4	268.5	277.1	284.8	293.9	300.2
260	205.0	209.9	217.2	223.7	231.2	289.6	298.6	306.6	316.0	322.5
280	222.8	227.9	235.5	242.2	250.1	310.7	320.0	328.2	338.0	344.7
300	240.7	246.0	253.9	260.9	269.1	331.8	341.4	349.9	359.9	366.8
320	258.6	264.1	272.3	279.6	288.0	352.8	362.7	371.4	381.8	388.9
340	276.6	282.3	290.8	298.3	307.0	373.8	384.0	393.0	403.6	410.9
360	294.6	300.5	309.3	317.0	326.1	394.8	405.2	414.5	425.3	432.9
380	312.7	318.8	327.9	335.8	345.1	415.7	426.5	435.9	447.1	454.8
400	330.9	337.2	346.5	354.6	364.2	436.6	447.6	457.3	468.7	476.6
450	376.5	383.2	393.1	401.8	412.0	488.8	500.5	510.7	522.7	531.0
500	422.3	429.4	439.9	449.1	459.9	540.9	553.1	563.9	576.5	585.2
550	468.3	475.8	486.9	496.6	507.9	592.9	605.7	616.9	630.1	639.2
600	514.5	522.4	534.0	544.2	556.1	644.8	658.1	669.8	683.5	693.0
650	560.9	569.1	581.2	591.9	604.2	696.6	710.4	722.5	736.8	746.6
700	607.4	615.9	628.6	639.6	652.5	748.4	762.7	775.2	790.0	800.1
750	654.0	662.9	676.0	687.5	700.8	800.0	814.8	827.8	843.0	853.5
800	700.7	709.9	723.5	735.4	749.2	851.7	866.9	880.3	896.0	906.8
850	747.6	757.0	771.1	783.3	797.6	903.2	918.9	932.7	948.8	960.0
900	794.5	804.3	818.8	831.4	846.1	954.8	970.9	985.0	1002	1013
950	841.5	851.5	866.5	879.5	894.6	1006	1023	1037	1054	1066
1000	888.6	898.9	914.3	927.6	943.1	1058	1075	1090	1107	1119

Für große $n > 50$ ist $\chi^2_{n;\alpha} \approx \left(z_\alpha + \sqrt{2\,n-1}\right)^2/2$; z_α – Quantil der Standardnormalverteilung.

z. B. 1000	887.7	898.2	913.8	927.3	943.0	1058	1074	1089	1106	1118

R	Excel	SPSS	STATA
`qchisq(alpha, n)`	`CHIINV(1-alpha; n)`	`IDF.CHISQ(alpha, n)`	`invchi2(m, alpha)`

4.10 Quantile der F-Verteilung

$$F_{m,n;1-\alpha} = F_{F_{m,n}}^{-1}(1-\alpha) \quad \Leftrightarrow \quad 1-\alpha = F_{F_{m,n}}\left(F_{m,n;1-\alpha}\right) = \int_0^{F_{m,n;1-\alpha}} f(x)\,dx$$

$$\text{mit } f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-(m+n)/2}$$

Bezeichnung:

m – Anzahl der Freiheitsgrade im Zähler, n – Anzahl der Freiheitsgrade im Nenner

Rechenregel zum Ermittlung von Quantilen kleiner Ordnung ($\alpha < 0.5$):

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

Auf den nächsten vier Seiten folgen die Quantile für

- $\alpha \in \{0.9, 0.95, 0.975, 0.99\}$ und
- $n, m \in \{1, 2, \ldots, 10, 15, 20, 30, \ldots, 100\}$.

R	Excel	SPSS	STATA
`qf(alpha, m, n)`	`FINV(1-alpha; m; n)`	`IDF.F(alpha, m, n)`	`invF(m, n, alpha)`

Quantile der F-Verteilung $F_{m,n;\alpha}$ mit $\alpha = 0.9$

m \ n	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	60	70	80	90	100
1	39.86	8.526	5.538	4.545	4.060	3.776	3.589	3.458	3.360	3.285	3.073	2.975	2.881	2.835	2.809	2.791	2.779	2.769	2.762	2.756
2	49.50	9.000	5.462	4.325	3.780	3.463	3.257	3.113	3.006	2.924	2.695	2.589	2.489	2.440	2.412	2.393	2.380	2.370	2.363	2.356
3	53.59	9.162	5.391	4.191	3.619	3.289	3.074	2.924	2.813	2.728	2.490	2.380	2.276	2.226	2.197	2.177	2.164	2.154	2.146	2.139
4	55.83	9.243	5.343	4.107	3.520	3.181	2.961	2.806	2.693	2.605	2.361	2.249	2.142	2.091	2.061	2.041	2.027	2.016	2.008	2.002
5	57.24	9.293	5.309	4.051	3.453	3.108	2.883	2.726	2.611	2.522	2.273	2.158	2.049	1.997	1.966	1.946	1.931	1.921	1.912	1.906
6	58.20	9.326	5.285	4.010	3.405	3.055	2.827	2.668	2.551	2.461	2.208	2.091	1.980	1.927	1.895	1.875	1.860	1.849	1.841	1.834
7	58.91	9.349	5.266	3.979	3.368	3.014	2.785	2.624	2.505	2.414	2.158	2.040	1.927	1.873	1.840	1.819	1.804	1.793	1.785	1.778
8	59.44	9.367	5.252	3.955	3.339	2.983	2.752	2.589	2.469	2.377	2.119	1.999	1.884	1.829	1.796	1.775	1.760	1.748	1.739	1.732
9	59.86	9.381	5.240	3.936	3.316	2.958	2.725	2.561	2.440	2.347	2.086	1.965	1.849	1.793	1.760	1.738	1.723	1.711	1.702	1.695
10	60.20	9.392	5.230	3.920	3.297	2.937	2.703	2.538	2.416	2.323	2.059	1.937	1.819	1.763	1.729	1.707	1.691	1.680	1.670	1.663
15	61.22	9.425	5.200	3.870	3.238	2.871	2.632	2.464	2.340	2.244	1.972	1.845	1.722	1.662	1.627	1.603	1.587	1.574	1.564	1.557
20	61.74	9.441	5.184	3.844	3.207	2.836	2.595	2.425	2.298	2.201	1.924	1.794	1.667	1.605	1.568	1.543	1.526	1.513	1.503	1.494
30	62.26	9.458	5.168	3.817	3.174	2.800	2.555	2.383	2.255	2.155	1.873	1.738	1.606	1.541	1.502	1.476	1.457	1.443	1.432	1.423
40	62.53	9.466	5.160	3.804	3.157	2.781	2.535	2.361	2.232	2.132	1.845	1.708	1.573	1.506	1.465	1.437	1.418	1.403	1.391	1.382
50	62.69	9.471	5.155	3.795	3.147	2.770	2.523	2.348	2.218	2.117	1.828	1.690	1.552	1.483	1.441	1.413	1.392	1.377	1.365	1.355
60	62.79	9.475	5.151	3.790	3.140	2.762	2.514	2.339	2.208	2.107	1.817	1.677	1.538	1.467	1.424	1.395	1.374	1.358	1.346	1.336
70	62.87	9.477	5.149	3.786	3.135	2.756	2.508	2.333	2.202	2.100	1.808	1.667	1.527	1.455	1.412	1.382	1.361	1.344	1.332	1.321
80	62.93	9.479	5.147	3.782	3.132	2.752	2.504	2.328	2.196	2.095	1.802	1.660	1.519	1.447	1.402	1.372	1.350	1.334	1.321	1.310
90	62.97	9.480	5.145	3.780	3.129	2.749	2.500	2.324	2.192	2.090	1.797	1.655	1.512	1.439	1.395	1.364	1.342	1.325	1.312	1.301
100	63.01	9.481	5.144	3.778	3.126	2.746	2.497	2.321	2.189	2.087	1.793	1.650	1.507	1.434	1.388	1.358	1.335	1.318	1.304	1.293

Quantile der F-Verteilung $F_{m,n;\alpha}$ mit $\alpha = 0.95$

m \ n	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	60	70	80	90	100
1	161.4	18.51	10.13	7.709	6.608	5.987	5.591	5.318	5.117	4.965	4.543	4.351	4.171	4.085	4.034	4.001	3.978	3.960	3.947	3.936
2	200.0	19.00	9.552	6.944	5.786	5.143	4.737	4.459	4.256	4.103	3.682	3.493	3.316	3.232	3.183	3.150	3.128	3.111	3.098	3.087
3	215.7	19.16	9.277	6.591	5.409	4.757	4.347	4.066	3.863	3.708	3.287	3.098	2.922	2.839	2.790	2.758	2.736	2.719	2.706	2.696
4	224.6	19.25	9.117	6.388	5.192	4.534	4.120	3.838	3.633	3.478	3.056	2.866	2.690	2.606	2.557	2.525	2.503	2.486	2.473	2.463
5	230.2	19.30	9.013	6.256	5.050	4.387	3.972	3.687	3.482	3.326	2.901	2.711	2.534	2.449	2.400	2.368	2.346	2.329	2.316	2.305
6	234.0	19.33	8.941	6.163	4.950	4.284	3.866	3.581	3.374	3.217	2.790	2.599	2.421	2.336	2.286	2.254	2.231	2.214	2.201	2.191
7	236.8	19.35	8.887	6.094	4.876	4.207	3.787	3.500	3.293	3.135	2.707	2.514	2.334	2.249	2.199	2.167	2.143	2.126	2.113	2.103
8	238.9	19.37	8.845	6.041	4.818	4.147	3.726	3.438	3.230	3.072	2.641	2.447	2.266	2.180	2.130	2.097	2.074	2.056	2.043	2.032
9	240.5	19.38	8.812	5.999	4.772	4.099	3.677	3.388	3.179	3.020	2.588	2.393	2.211	2.124	2.073	2.040	2.017	1.999	1.986	1.975
10	241.9	19.40	8.786	5.964	4.735	4.060	3.637	3.347	3.137	2.978	2.544	2.348	2.165	2.077	2.026	1.993	1.969	1.951	1.938	1.927
15	246.0	19.43	8.703	5.858	4.619	3.938	3.511	3.218	3.006	2.845	2.403	2.203	2.015	1.924	1.871	1.836	1.812	1.793	1.779	1.768
20	248.0	19.45	8.660	5.803	4.558	3.874	3.445	3.150	2.936	2.774	2.328	2.124	1.932	1.839	1.784	1.748	1.722	1.703	1.688	1.676
30	250.1	19.46	8.617	5.746	4.496	3.808	3.376	3.079	2.864	2.700	2.247	2.039	1.841	1.744	1.687	1.649	1.622	1.602	1.586	1.573
40	251.1	19.47	8.594	5.717	4.464	3.774	3.340	3.043	2.826	2.661	2.204	1.994	1.792	1.693	1.634	1.594	1.566	1.545	1.528	1.515
50	251.8	19.48	8.581	5.699	4.444	3.754	3.319	3.020	2.803	2.637	2.178	1.966	1.761	1.660	1.599	1.559	1.530	1.508	1.491	1.477
60	252.2	19.48	8.572	5.688	4.431	3.740	3.304	3.005	2.787	2.621	2.160	1.946	1.740	1.637	1.576	1.534	1.505	1.482	1.465	1.45
70	252.5	19.48	8.566	5.679	4.422	3.730	3.294	2.994	2.776	2.610	2.147	1.932	1.724	1.621	1.558	1.516	1.486	1.463	1.445	1.430
80	252.7	19.48	8.561	5.673	4.415	3.722	3.286	2.986	2.768	2.601	2.137	1.922	1.712	1.608	1.544	1.502	1.471	1.448	1.429	1.415
90	252.9	19.48	8.557	5.668	4.409	3.716	3.280	2.980	2.761	2.594	2.130	1.913	1.703	1.597	1.534	1.491	1.459	1.436	1.417	1.402
100	253.0	19.49	8.554	5.664	4.405	3.712	3.275	2.975	2.756	2.588	2.123	1.907	1.695	1.589	1.525	1.481	1.450	1.426	1.407	1.392

Quantile der F-Verteilung $F_{m,n;\alpha}$ mit $\alpha = 0.975$

m \ n	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	60	70	80	90	100
1	647.8	38.51	17.44	12.22	10.01	8.813	8.073	7.571	7.209	6.937	6.200	5.871	5.568	5.424	5.340	5.286	5.247	5.218	5.196	5.179
2	799.5	39.00	16.04	10.65	8.434	7.260	6.542	6.059	5.715	5.456	4.765	4.461	4.182	4.051	3.975	3.925	3.890	3.864	3.844	3.828
3	864.2	39.16	15.44	9.979	7.764	6.599	5.890	5.416	5.078	4.826	4.153	3.859	3.589	3.463	3.390	3.343	3.309	3.284	3.265	3.250
4	899.6	39.25	15.10	9.605	7.388	6.227	5.523	5.053	4.718	4.468	3.804	3.515	3.250	3.126	3.054	3.008	2.975	2.950	2.932	2.917
5	921.8	39.30	14.88	9.364	7.146	5.988	5.285	4.817	4.484	4.236	3.576	3.289	3.026	2.904	2.833	2.786	2.754	2.730	2.711	2.696
6	937.1	39.33	14.74	9.197	6.978	5.820	5.119	4.652	4.320	4.072	3.415	3.128	2.867	2.744	2.674	2.627	2.595	2.571	2.552	2.537
7	948.2	39.36	14.62	9.074	6.853	5.695	4.995	4.529	4.197	3.950	3.293	3.007	2.746	2.624	2.553	2.507	2.474	2.450	2.432	2.417
8	956.7	39.37	14.54	8.980	6.757	5.600	4.899	4.433	4.102	3.855	3.199	2.913	2.651	2.529	2.458	2.412	2.379	2.355	2.336	2.321
9	963.3	39.39	14.47	8.905	6.681	5.523	4.823	4.357	4.026	3.779	3.123	2.837	2.575	2.452	2.381	2.334	2.302	2.277	2.259	2.244
10	968.6	39.40	14.42	8.844	6.619	5.461	4.761	4.295	3.964	3.717	3.060	2.774	2.511	2.388	2.317	2.270	2.237	2.213	2.194	2.179
15	984.9	39.43	14.25	8.657	6.428	5.269	4.568	4.101	3.769	3.522	2.862	2.573	2.307	2.182	2.109	2.061	2.028	2.003	1.983	1.968
20	993.1	39.45	14.17	8.560	6.329	5.168	4.467	3.999	3.667	3.419	2.756	2.464	2.195	2.068	1.993	1.944	1.910	1.884	1.864	1.849
30	1001.	39.46	14.08	8.461	6.227	5.065	4.362	3.894	3.560	3.311	2.644	2.349	2.074	1.943	1.866	1.815	1.779	1.752	1.731	1.715
40	1006.	39.47	14.04	8.411	6.175	5.012	4.309	3.840	3.505	3.255	2.585	2.287	2.009	1.875	1.796	1.744	1.707	1.679	1.657	1.640
50	1008.	39.48	14.01	8.381	6.144	4.980	4.276	3.807	3.472	3.221	2.549	2.249	1.968	1.832	1.752	1.699	1.660	1.632	1.61	1.592
60	1010.	39.48	13.99	8.360	6.123	4.959	4.254	3.784	3.449	3.198	2.524	2.223	1.940	1.803	1.721	1.667	1.628	1.599	1.576	1.558
70	1011.	39.48	13.98	8.346	6.107	4.943	4.239	3.768	3.433	3.182	2.506	2.205	1.920	1.781	1.698	1.643	1.604	1.574	1.551	1.532
80	1012.	39.48	13.97	8.335	6.096	4.932	4.227	3.756	3.421	3.169	2.493	2.190	1.904	1.764	1.681	1.625	1.585	1.555	1.531	1.512
90	1013.	39.49	13.96	8.326	6.087	4.923	4.218	3.747	3.411	3.160	2.482	2.179	1.892	1.751	1.667	1.611	1.570	1.540	1.516	1.496
100	1013.	39.49	13.96	8.319	6.080	4.915	4.210	3.739	3.403	3.152	2.474	2.170	1.882	1.741	1.656	1.599	1.558	1.527	1.503	1.483

Quantile der F-Verteilung $F_{m,n;\alpha}$ mit $\alpha = 0.99$

m \ n	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	60	70	80	90	100
1	4052.	98.50	34.12	21.20	16.26	13.74	12.25	11.26	10.56	10.04	8.683	8.096	7.562	7.314	7.171	7.077	7.011	6.963	6.925	6.895
2	5000.	99.00	30.82	18.00	13.27	10.92	9.547	8.649	8.022	7.559	6.359	5.849	5.390	5.179	5.057	4.977	4.922	4.881	4.849	4.824
3	5403.	99.12	29.46	16.69	12.06	9.780	8.451	7.591	6.992	6.552	5.417	4.938	4.510	4.313	4.199	4.126	4.074	4.036	4.007	3.984
4	5625.	99.25	28.71	15.98	11.39	9.148	7.847	7.006	6.422	5.994	4.893	4.431	4.018	3.828	3.720	3.649	3.600	3.563	3.535	3.513
5	5764.	99.30	28.24	15.52	10.97	8.746	7.460	6.632	6.057	5.636	4.556	4.103	3.699	3.514	3.408	3.339	3.291	3.255	3.228	3.206
6	5859.	99.33	27.91	15.21	10.67	8.466	7.191	6.371	5.802	5.386	4.318	3.871	3.473	3.291	3.186	3.119	3.071	3.036	3.009	2.988
7	5928.	99.36	27.67	14.98	10.46	8.260	6.993	6.178	5.613	5.200	4.142	3.699	3.304	3.124	3.020	2.953	2.906	2.871	2.845	2.823
8	5981.	99.37	27.49	14.80	10.29	8.102	6.840	6.029	5.467	5.057	4.004	3.564	3.173	2.993	2.890	2.823	2.777	2.742	2.715	2.694
9	6022.	99.39	27.34	14.66	10.16	7.976	6.719	5.911	5.351	4.942	3.895	3.457	3.067	2.888	2.785	2.718	2.672	2.637	2.611	2.590
10	6056.	99.40	27.23	14.55	10.05	7.874	6.620	5.814	5.257	4.849	3.805	3.368	2.979	2.801	2.698	2.632	2.585	2.551	2.524	2.503
15	6157.	99.43	26.87	14.20	9.722	7.559	6.314	5.515	4.962	4.558	3.522	3.088	2.700	2.522	2.419	2.352	2.306	2.271	2.244	2.223
20	6209.	99.45	26.69	14.02	9.553	7.396	6.155	5.359	4.808	4.405	3.372	2.938	2.549	2.369	2.265	2.198	2.150	2.115	2.088	2.067
30	6261.	99.47	26.50	13.84	9.379	7.229	5.992	5.198	4.649	4.247	3.214	2.778	2.386	2.203	2.098	2.028	1.980	1.944	1.916	1.893
40	6287.	99.47	26.41	13.74	9.291	7.143	5.908	5.116	4.567	4.165	3.132	2.695	2.299	2.114	2.007	1.936	1.886	1.849	1.820	1.797
50	6302.	99.48	26.35	13.69	9.238	7.091	5.858	5.065	4.517	4.115	3.081	2.643	2.245	2.058	1.949	1.877	1.826	1.788	1.759	1.735
60	6313.	99.48	26.32	13.65	9.202	7.057	5.824	5.032	4.483	4.082	3.047	2.608	2.208	2.019	1.909	1.836	1.785	1.746	1.716	1.692
70	6321.	99.48	26.29	13.62	9.176	7.032	5.799	5.007	4.459	4.058	3.022	2.582	2.181	1.991	1.880	1.806	1.754	1.714	1.684	1.659
80	6326.	99.49	26.27	13.60	9.157	7.013	5.781	4.989	4.441	4.039	3.004	2.563	2.160	1.969	1.857	1.783	1.730	1.690	1.659	1.634
90	6331.	99.49	26.25	13.59	9.142	6.998	5.766	4.975	4.426	4.025	2.989	2.548	2.144	1.952	1.839	1.764	1.711	1.671	1.639	1.614
100	6334.	99.49	26.24	13.58	9.130	6.987	5.755	4.963	4.415	4.014	2.977	2.535	2.131	1.938	1.825	1.749	1.695	1.655	1.623	1.598

4.11 Kolmogoroff-Smirnoff-Test

Kritische Werte für den Kolmogoroff-Smirnoff-Test

α = W.(Fehler 1. Art) $\rightsquigarrow$ kritischer Wert $c_{n;1-\alpha}$

Testentscheidung: Teststatistik $D > c_{n;1-\alpha}$ $\rightsquigarrow$ H_0 wird verworfen!

n	α 0.2	0.1	0.08	0.05	0.04	0.02	0.01
1	0.900	0.950	0.960	0.975	0.980	0.990	0.995
2	0.684	0.776	0.800	0.842	0.859	0.900	0.929
3	0.565	0.636	0.658	0.708	0.729	0.785	0.829
4	0.493	0.565	0.585	0.624	0.641	0.689	0.734
5	0.447	0.509	0.527	0.563	0.580	0.627	0.669
6	0.410	0.468	0.485	0.519	0.534	0.577	0.617
7	0.381	0.436	0.452	0.483	0.497	0.538	0.576
8	0.358	0.410	0.425	0.454	0.468	0.507	0.542
9	0.339	0.387	0.402	0.430	0.443	0.480	0.513
10	0.323	0.369	0.382	0.409	0.421	0.457	0.489
11	0.308	0.352	0.365	0.391	0.403	0.437	0.468
12	0.296	0.338	0.351	0.375	0.387	0.419	0.449
13	0.285	0.325	0.338	0.361	0.372	0.404	0.432
14	0.275	0.314	0.326	0.349	0.359	0.390	0.418
15	0.266	0.304	0.315	0.338	0.348	0.377	0.404
16	0.258	0.295	0.306	0.327	0.337	0.366	0.392
17	0.250	0.286	0.297	0.318	0.327	0.355	0.381
18	0.244	0.279	0.289	0.309	0.319	0.346	0.371
19	0.237	0.271	0.281	0.301	0.310	0.337	0.361
20	0.232	0.265	0.275	0.294	0.303	0.329	0.352
21	0.226	0.259	0.268	0.287	0.296	0.321	0.344
22	0.221	0.253	0.262	0.281	0.289	0.314	0.337
23	0.216	0.247	0.257	0.275	0.283	0.307	0.330
24	0.212	0.242	0.251	0.269	0.277	0.301	0.323
25	0.208	0.238	0.246	0.264	0.272	0.295	0.317
26	0.204	0.233	0.242	0.259	0.267	0.290	0.311
27	0.200	0.229	0.237	0.254	0.262	0.284	0.305
28	0.197	0.225	0.233	0.250	0.257	0.279	0.300
29	0.193	0.221	0.229	0.246	0.253	0.275	0.295
30	0.190	0.218	0.226	0.242	0.249	0.270	0.290
31	0.187	0.214	0.222	0.238	0.245	0.266	0.285
32	0.184	0.211	0.219	0.234	0.241	0.262	0.281
33	0.182	0.208	0.215	0.231	0.238	0.258	0.277
34	0.179	0.205	0.212	0.227	0.234	0.254	0.273
35	0.177	0.202	0.209	0.224	0.231	0.251	0.269
36	0.174	0.199	0.206	0.221	0.228	0.247	0.265
37	0.172	0.196	0.204	0.218	0.225	0.244	0.262
38	0.170	0.194	0.201	0.215	0.222	0.241	0.258
39	0.168	0.191	0.199	0.213	0.219	0.238	0.255
40	0.165	0.189	0.196	0.210	0.216	0.235	0.252
Approximation für > 40	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.27}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.40}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Die Approximation lässt sich verbessern, wenn man statt $\sqrt{n}$ den Ausdruck $\left(\sqrt{n} + 0.12 + 0.11/\sqrt{n}\right)$ verwendet, z. B. für $n = 40$ ergeben sich mit $\boxed{0.166\ \ 0.189\ \ 0.196\ \ 0.210\ \ 0.217\ \ 0.235\ \ 0.252}$ nur geringfügige Abweichungen zu den genauen Werten.

R	Excel	SPSS	STATA
`ks.test(x, ...)`	`<addon only>`	`[Analysieren->NP Tests->K-S]`	`ksmirnov`

Index

α-Niveau-Test, 29
Abhängigkeitstest, 31
Anpassungstest, 34
Approximationen, 25

Bayes, Satz von, 18
Bedingte Verteilungen, 21
Bestimmtheitsmaß, 15
Binomialkoeffizient, 19
Bivariates, 6, 21
Boxplot, 8

χ^2-Maß, 12
χ^2-Test
  Anpassungs-, 35
  Unabhängigkeits-, 32

Dichte, 20
  bedingte, 21

Einstichprobentests, 45
Ereignis, 16
Erwartungstreue, 27
Erwartungswert, 22

Fakultät, 19
Fehler 1., 2. Art, 28
Fisher-Test, 33

Ginikoeffizient, 11
Grenzwertsatz
  Moivre-Laplace, 25
  Poisson, 25
Grundgesamtheit, 4, 16
Gütefunktion, 29

Histogramm, 5
Homoskedastizität, 37
Häufigkeit, 4

Indexzahlen, 10
Indizes
  nach Laspeyres, 10
  nach Paasche, 10
  Umbasieren, 10
  Verketten, 10

Kolmogoroff-Smirnoff-Test, 34, 58
Kombinationen, 18
Konfidenzintervalle, 28, 44
Konsistenz, 27
Kontingenzkoeffizient
  Pearsonscher, 13
Kontingenztafel, 6
Konzentrationsindizes, 12
Konzentrationskurve, 12
Korrelation
  lineare, 24
Korrelationskoeffizient
  Bravais-Pearson-, 13
  Spearman-, 14
Korrelationstest, 31
Kovarianz, 23
  Empirische, 13

Lineare Regression
  Einfache
    Empirische, 14
  Multiple
    Empirische, 15
Lorenzkurve, 11

MAD, 9
Maximum-Likelihood-Methode, 27
Merkmal, 4
messbarer Raum, 17
Mittel
  Arithmetisches, 7
  Geometrisches, 7
  Getrimmtes, 7
  Harmonisches, 7
Modus, 8
Momenten-Schätzmethode, 27
Multinomialkoeffizient, 19

p-Wert, 29
Punktschätzverfahren, 27

QQ-Plot, 35
Quantil, 7
Quantile der
  χ^2-Vert., 51
  F-Vert., 53
  NV, 49
  t-Vert., 50
Quartil, 8
Quartilsabstand, 8

Randverteilung, 21
Rang, 13
Regressionsmodell, lineares
    Prognose, 39
    Tests im, 37

Schätzer
    Eigenschaften von, 27
Shapiro-Wilk-Test, 36
σ-Algebra, 16
Skalen, 4
Spannweite, 8
Standardabweichung, 23
    Empirische, 9
Statistik, 26
Stetigkeitskorrektur, 25
Stichprobenfunktionen, 26
    spezielle, 26

t-Test
    verbundener, 30
Tschebyschoff-Ungleichung, 24

Unabhängigkeit, stochastische, 18
Unabhängigkeit
    von Zufallsvariablen, 22

Varianz, 23
    Empirische, 9
    Stichproben-, 9
Variationen, 18
Variationskoeffizient, 9
Verschiebungssatz, 23
Verteilung
    Bernoulli-, 41
    Binomial-, 41
    Cauchy-, 42
    χ^2-, 43
    Exponential-, 42
    F-, Fisher-, 43
    Gamma-, 42
    Geometrische, 41
    Gleich-, 42
    Hypergeometrische, 41
    Laplace-, 42
    Normal-, 43
    Poission, 41
    t-, Student-, 43
Verteilungsf. der NV, 47
Verteilungsfunktion, 19
    2-dimensionale, 21
    Empirische, 5

Wahrscheinlichkeit
    bedingte, 17
    Laplace-, 17
    totale, 18
Wahrscheinlichkeitsdichte, 20
Wahrscheinlichkeitsfunktion, 20
Wahrscheinlichkeitsmaß, 17
Wahrscheinlichkeitsraum, 17

Zentraler Grenzwertsatz, 24
Ziehungen k aus n, 18
Zufallsstichproben, 26
Zufallsvariable, 19
    diskrete, 20
    stetige, 20
Zufallsvariablen
    Summen von, 25
Zweistichprobentests, 46