wiwi-online.net
Der Begleitfaden für Studium & Karriere

Dr. René M. Schröder
Michael Böttcher

Formelsammlung der Wirtschaftswissenschaften

SCHWERPUNKT STATISTIK

1. Auflage

Inhaltsverzeichnis

Die nächsten Ausgaben der Formelsammlung erscheinen:

Schwerpunkt BWL	**Oktober 2009**	(Semesteranfang)
Schwerpunkt VWL	**April 2010**	(Semesteranfang)

Hinweis: Eine für alle Hochschulen einheitliche Symbolisierung ist leider nicht realisierbar. Insofern bitten wir um Verständnis, falls die Symbole der Formelsammlung nicht mit den Ihrigen identisch sind. Sollten Sie Fehler finden oder Ergänzungsvorschläge haben, teilen Sie uns dieses bitte umgehend mit. Wir werden Ihre Hinweise schnellstmöglich mit einbinden. Eine aktuelle überarbeitete Fassung dieser Formelsammlung finden Sie ständig im Internet unter **www.wiwi-online.net**. Dort steht sie Ihnen zum kostenlosen Download bereit. Wir wünschen Ihnen weiterhin viel Erfolg bei Ihrem Studium.

1 Beschreibende Statistik

Symbole:

im diskreten Fall:

x_i^*: diskreter Wert $(i = 1, ..., k)$

n_i: absolute Häufigkeit von x_i^* $(i = 1, ..., k)$

h_i: relative Häufigkeit von x_i^* $h_i = n_i/n$ $(i = 1, ..., k)$

im stetigen Fall (klassifizierte Daten):

a_i: untere Klassengrenze von Klasse i $(i = 1, ..., k)$

b_i: obere Klassengrenze von Klasse i $(i = 1, ..., k)$

x_i^M: Klassenmitte von Klasse i $(i = 1, ..., k)$

n_i: absolute Häufigkeit der Werte in der Klasse $]a_i, b_i]$ $(i = 1, ..., k)$

allgemein:

x_i: Wert aus einer Datenliste $(i = 1, ..., n)$

n: Summe aller n_i bzw. Anzahl der Werte in einer Datenliste

x_p: $(p \cdot 100)$-Prozentpunkt/p-Quantil

x_{min}: minimaler Wert einer Datenliste

x_{max}: maximaler Wert einer Datenliste

$f(x)$: Häufigkeits- bzw. Dichtefunktion

$F(x)$: Verteilungsfunktion

$\overline{x}$: arithmetisches Mittel

s^2: Varianz

s: Standardabweichung

s_{xy}: Kovarianz

r_{xy}: Bestimmtheitsmaß

1.1 Merkmalstypen und Skalierungen

Diskretes Merkmal:

Ein Merkmal heißt diskret, wenn es endlich oder abzählbar unendlich viele Ausprägungen besitzt.

Stetiges Merkmal:

Ein Merkmal heißt stetig, wenn alle Werte eines Intervalls mögliche Ausprägungen sind. Das Merkmal besitzt also überabzählbar unendlich viele verschiedene Ausprägungen.

Merkmalstyp:	**Verhältnisse:**	**Skalierung:**	**Beispiele:**
qualitatives Merkmal	Verschiedenheit $x_i \neq x_j$	Nominalskala	Geschlecht, Telefonnumer
komparatives Merkmal	Rangfolge $x_i < x_j$	Ordinalskala	Schulnoten, IQ
quantitatives Merkmal	Abstände $x_i - x_j$ sinnvoll	Intervallskala	Geburtsjahr, Temp. in °C
	Verhältnisse $x_i : x_j$ sinnvoll	Verhältnisskala	Preis, Gewicht

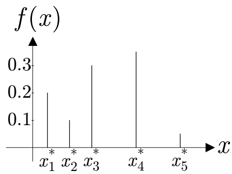
1.2 Empirische Häufigkeits- und Verteilungsfunktion

1.2.1 Diskrete Verteilung

Empirische Häufigkeitsfunktion:

(auch Stabdiagramm-Funktion)

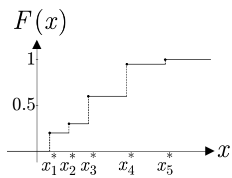
$$f(x) = \begin{cases} \frac{n_i}{n} & \text{für } x = x_i^* \quad (i = 1, ..., k) \\ 0 & \text{sonst} \end{cases}$$



Empirische Verteilungsfunktion (kumulierte rel. Häufigkeit):

mit $i = 1, ..., k-1$

$$F(x) = \begin{cases} 0 & \text{für } x < x_1 \\ \sum\limits_{j=1}^{i} \frac{n_j}{n} & \text{für } x_i^* \leq x < x_{i+1}^* \\ 1 & \text{für } x \geq x_k \end{cases}$$



p-Quantil: $x_p = \min\{x | F(x) \geq p\}$

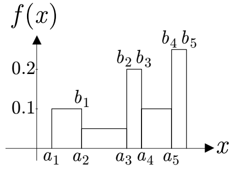
1.2.2 Stetige Verteilung (klassifizierte Daten)

Es gibt k Klassen $]a_i; b_i]$ mit $a_{i+1} = b_i$

Klassenbreite: $\Delta x_i = b_i - a_i \quad (i = 1, ..., k)$

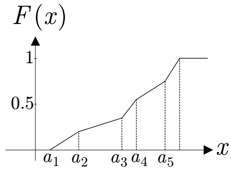
Empirische Dichtefunktion:

(auch Histogramm-Funktion)



$$f(x) = \begin{cases} \frac{n_i}{n \cdot \Delta x_i} & \text{für } a_i < x \leq b_i \quad (i = 1, ..., k) \\ 0 & \text{sonst} \end{cases}$$

Empirische Verteilungsfunktion: $(i = 1, ..., k)$



$$F(x) = \begin{cases} 0 & \text{für } x \leq a_1 \\ F(a_i) + \frac{x - a_i}{\Delta x_i} \cdot \frac{n_i}{n} & \text{für } a_i < x \leq b_i \\ 1 & \text{für } x > b_k \end{cases}$$

p-Quantil: $x_p = a_i + \dfrac{p - F(a_i)}{\frac{n_i}{n}} \cdot \Delta x_i \quad \text{mit } F(a_i) < p \leq F(b_i)$

1.3 Maßzahlen empirischer Verteilungen

1.3.1 Lagemaße

Arithmetisches Mittel:

bei Rohdaten: $\overline{x} = \dfrac{1}{n} \sum_{i=1}^{n} x_i$

bei diskreten Daten: $\overline{x} = \dfrac{1}{n} \sum_{i=1}^{k} x_i^* \cdot n_i = \sum_{i=1}^{k} \dfrac{n_i}{n} x_i^*$

bei klassifizierten Daten: $\overline{x} \approx \dfrac{1}{n} \sum_{i=1}^{k} x_i^M \cdot n_i \qquad \text{mit } x_i^M = (a_i + b_i)/2$

Harmonisches Mittel: $\overline{x}_{harm} = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

Geometrisches Mittel: $\overline{x}_{geom} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad x_i > 0$

Modus (Modalwert): $x_h \qquad$ wobei $f(x_h) \geq f(x)$ für alle $x \in \mathbb{R}$

Median (Zentralwert, 50%-Punkt, 0,5-Quantil):

bei geordneten Daten $x_1 \leq x_2 \leq ... \leq x_{n-1} \leq x_n$:

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{für n ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{für n gerade} \end{cases}$$

bei diskreten und klassifizierten Daten:

$\tilde{x} = x_{0,5} \qquad$ mit $F(x_{0,5}) = 0,5$

1.3.2 Streuungsmaße

Empirische Varianz:

bei Rohdaten: $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$

bei diskreten Daten: $s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i^* - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^{*2} - \overline{x}^2$

bei klassifizierten Daten: $s^2 \approx \frac{1}{n}\sum_{i=1}^{k} n_i(x_i^M - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^{M2} - \overline{x}^2$

mit $x_i^M = (a_i + b_i)/2$

Verschiebungssatz:

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 - (\overline{x} - c)^2 \qquad$ für jedes $c \in \mathbb{R}$

$s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i^* - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - c)^2 - (\overline{x} - c)^2 \qquad$ für jedes $c \in \mathbb{R}$

Streuungszerlegungssatz:

Bei Einteilung der Daten in k Gruppen mit jeweils n_i Beobachtungen

gilt: $s^2_{ges} = s^2_{intern} + s^2_{extern}$

$s_{ges} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_{ges})^2$ $\qquad s^2_{intern} = \frac{1}{n}\sum_{i=1}^{k} n_i s_i^2$

$s^2_{extern} = \frac{1}{n}\sum_{i=1}^{k}(\overline{x}_i - \overline{x}_{ges})n_i = \frac{1}{n}\sum_{i=1}^{k} n_i \overline{x}_i^2 - \overline{x}^2$

mit $s_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2$ (Varianz von Gruppe i)

$\overline{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ (Arithmetisches Mittel von Gruppe i)

Empirische Standardabweichung: $s = \sqrt{s^2}$

Empirischer Variationskoeffizient: $v = \frac{s}{\overline{x}}$

Mittlere absolute Abweichung (MAD):

bei Rohdaten: $\overline{d} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|$

bei diskreten Werten: $\overline{d} = \frac{1}{n}\sum_{i=1}^{k} n_i\,|x_i^* - \overline{x}|$

Empirische Spannweite: $Spannweite = x_{max} - x_{min}$

Emp. Interquartilsabstand: $Q = x_{0{,}75} - x_{0{,}25}$

1.4 Zweidimensionale Häufigkeitsverteilungen

Beobachtungswerte/Datenpaare: (x_t, y_t) $(t = 1, ..., n)$

mögliche Merkmalsausprägungen: (x_i, y_j) $(i = 1, ..., p)$ $(j = 1, ..., q)$

1.4.1 Häufigkeiten

absolute Häufigkeiten: n_{ij} $(i = 1, ..., p)$ $(j = 1, ..., q)$

relative Häufigkeiten: $h_{ij} = \frac{n_{ij}}{n}$ $(i = 1, ..., p)$ $(j = 1, ..., q)$

absolute Randhäufigkeiten:

des Merkmals X: $n_{i\bullet} = \sum_{j=1}^{q} n_{ij}$ $(i = 1, ..., p)$

des Merkmals Y: $n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$ $(j = 1, ..., q)$

relative Randhäufigkeiten:

des Merkmals X: $h_{i\bullet} = \sum_{j=1}^{q} h_{ij} = \frac{n_{i\bullet}}{n}$ $(i = 1, ..., p)$

des Merkmals Y: $h_{\bullet j} = \sum_{i=1}^{p} h_{ij} = \frac{n_{\bullet j}}{n}$ $(j = 1, ..., q)$

bedingte Häufigkeiten:

des Merkmals X $h_{i(j)} = \frac{n_{ij}}{n_{\bullet j}} = \frac{h_{ij}}{h_{\bullet j}}$ $(i = 1, ..., p)$

des Merkmals Y $h_{(i)j} = \frac{n_{ij}}{n_{i\bullet}} = \frac{h_{ij}}{h_{i\bullet}}$ $(j = 1, ..., q)$

Unabhängigkeit der Merkmale X und Y:

$h_{ij} = h_{i\bullet} \cdot h_{\bullet j} \quad \Leftrightarrow \quad n_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$ muss für alle i und j gelten

1.4.2 Empirische Korrelation und Regression

Empirische Kovarianz:

$$s_{xy} = s_{yx} = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})(y_t - \overline{y}) = \frac{1}{n}\sum_{t=1}^{n} x_t y_t - \overline{x}\cdot\overline{y}$$

Es gilt: $s_{xy} > 0$ $\rightarrow$ positive Korrelation
$s_{xy} < 0$ $\rightarrow$ negative Korrelation
$s_{xy} = 0$ $\rightarrow$ keine Korrelation

Emp. Korrelationskoeffizient nach Bravais/Pearson:

$$r_{xy} = r_{yx} = \frac{s_{xy}}{s_x \cdot s_y} \qquad \text{dabei gilt:} \quad -1 \leq r_{xy} \leq 1$$

Unabhängigkeit der Merkmale X und Y:

X und Y unabhängig verteilt $\Rightarrow s_{xy} = r_{xy} = 0$

Rangkorrelationskoeffizient nach Spearman:

Es existieren n Datenpaare: $(x_1, y_1), ..., (x_n, y_n)$

$R(x_t)$: Rang von x_t (kleinster Wert erhält Rang 1 usw.)

$R(y_t)$: Rang von y_t (kleinster Wert erhält Rang 1 usw.)

Sofern es keine Bindungen (gleiche Werte) gibt, gilt:

$$r_{Sp} = 1 - \frac{6 \cdot \sum_{t=1}^{n}(R(x_t) - R(y_t))^2}{n(n^2-1)} \qquad \text{Es gilt:} \quad -1 \leq r_{Sp} \leq 1$$

Empirische lineare Regression:

Es existieren n Beobachtungswerte/Datenpaare: $(x_1, y_1), ..., (x_n, y_n)$

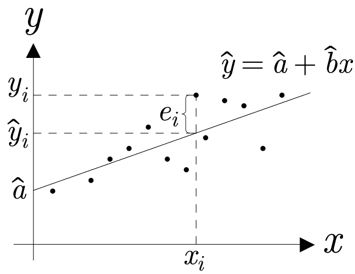
Regressionsmodell: $y_i = a + bx_i + e_i$

Regressionswerte: $\hat{y}_i = a + bx_i \qquad (i = 1, ..., n)$

Residuen: $e_i = y_i - \hat{y}_i \qquad (i = 1, ..., n)$

Methode der kleinsten Quadrate: $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \to \min!$

Die Parameter $\hat{a}$ und $\hat{b}$ der Regressionsgeraden $\hat{y} = \hat{a} + \hat{b}x$ werden bei der Methode der kleinsten Quadrate so bestimmt, dass die Summe der quadrierten Abstände e_i minimiert wird.



Steigung: $\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$

y-Achsenabschnitt: $\hat{a} = \overline{y} - \hat{b}\overline{x}$

Streuung der Regressionswerte: $s_{\hat{y}}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$

Rest- oder Residualstreuung: $s_e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$

Gesamtstreuung: $s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = s_{\hat{y}}^2 + s_e^2$

Bestimmtheitsmaß: $r_{xy}^2 = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = 1 - \frac{s_e^2}{s_y^2} = \frac{s_{\hat{y}^2}}{s_y^2}$

2 Wirtschaftsstatistik

2.1 Indexzahlen

Symbole:

p_{i0} bzw. p_{it}:	Preis von Gut i in Periode 0 bzw. t $\quad (i = 1, ..., n)$
q_{i0} bzw. q_{it}:	Menge von Gut i in Periode 0 bzw. t $\quad (i = 1, ..., n)$
Periode 0 bzw. t:	Basisperiode bzw. Berichtsperiode
I_{0t}:	Indexzahl (z.B. P_{0t}^L oder P_{0t}^P)

2.1.1 Preis-, Mengen- und Wertindizes

Preisindizes:

nach Laspeyres: $$P_{0t}^L = \frac{\sum\limits_{i=1}^n p_{it} q_{i0}}{\sum\limits_{i=1}^n p_{i0} q_{i0}} = \sum_{i=1}^n \frac{p_{it}}{p_{i0}} \cdot g_{i0} \qquad \text{mit } g_{i0} = \frac{p_{i0} q_{i0}}{\sum\limits_{i=1}^n p_{i0} q_{i0}}$$

nach Paasche: $$P_{0t}^P = \frac{\sum\limits_{i=1}^n p_{it} q_{it}}{\sum\limits_{i=1}^n p_{i0} q_{it}} = \frac{1}{\sum\limits_{i=1}^n \frac{p_{i0}}{p_{it}} \cdot g_{it}} \qquad \text{mit } g_{it} = \frac{p_{it} q_{it}}{\sum\limits_{i=1}^n p_{it} q_{it}}$$

nach Fisher: $$P_{0t}^F = \sqrt{P_{0t}^L \cdot P_{0t}^P}$$

Mengenindizes:

nach Laspeyres: $$Q_{0t}^L = \frac{\sum\limits_{i=1}^n p_{i0} q_{it}}{\sum\limits_{i=1}^n p_{i0} q_{i0}} = \sum_{i=1}^n \frac{q_{it}}{q_{i0}} \cdot g_{i0} \qquad \text{mit } g_{i0} = \frac{p_{i0} q_{i0}}{\sum\limits_{i=1}^n p_{i0} q_{i0}}$$

nach Paasche: $$Q_{0t}^P = \frac{\sum\limits_{i=1}^n p_{it} q_{it}}{\sum\limits_{i=1}^n p_{it} q_{i0}} = \frac{1}{\sum\limits_{i=1}^n \frac{q_{i0}}{q_{it}} \cdot g_{it}} \qquad \text{mit } g_{it} = \frac{p_{it} q_{it}}{\sum\limits_{i=1}^n p_{it} q_{it}}$$

nach Fisher: $Q_{0t}^F = \sqrt{Q_{0t}^L \cdot Q_{0t}^P}$

Wertindex (z.B. Umsatz- oder Ausgabenindex):

$$W_{0t} = \frac{\sum\limits_{i=1}^{n} p_{it}q_{it}}{\sum\limits_{i=1}^{n} p_{i0}q_{i0}} = \sum_{i=1}^{n} \frac{p_{it}q_{it}}{p_{i0}q_{i0}} \cdot g_{i0} = \frac{1}{\sum\limits_{i=1}^{n} \frac{p_{i0}q_{i0}}{p_{it}q_{it}} \cdot g_{it}} \qquad (g_{i0}, g_{it} \text{ siehe oben})$$

Zusammenhang der Indizes:

$$W_{0t} = P_{0t}^L \cdot Q_{0t}^P = P_{0t}^P \cdot Q_{0t}^L = P_{0t}^F \cdot Q_{0t}^F$$

2.1.2 Index-Anwendungen

Deflationierung (Preisbereinigung):

Deflationierung einer Wertgröße (z.B. Umsatz oder Ausgaben):

W_t^{real}: reale Wertgröße (zu konstanten Preisen der Basisperiode)
W_t^{nom}: nominale Wertgröße

$$W_t^{real} = \sum_{i=1}^{n} p_{i0}q_{it} = \frac{W_t^{nom}}{P_{0t}^P} \qquad \text{mit } W_t^{nom} = \sum_{i=1}^{n} p_{it}q_{it}$$

Deflationierung eines Wertindex:

Q_{0t}^L: Mengenindex nach Laspeyres als reale Größe
W_{0t}: Wertindex als nominale Größe

$$Q_{0t}^L = \frac{W_{0t}}{P_{0t}^P}$$

Umbasierung: $\hat{I}_{\tau t} = \frac{I_{0t}}{I_{0\tau}}$

Für die meisten Indexzahlen (insbesondere wenn I gleich P^L, P^P, Q^L oder Q^P ist) gilt: $\hat{I}_{\tau t} \neq I_{\tau t}$

Beispiel: $\hat{P}^L_{\tau t} = \frac{P^L_{0t}}{P^L_{0\tau}} = \frac{\sum\limits_{i=1}^{n} p_{it} q_{i0}}{\sum\limits_{i=1}^{n} p_{i\tau} q_{i0}}$ Es gilt: $\hat{P}^L_{\tau t} \neq P^L_{\tau t}$

Verkettung:

zweier Indexzahlen: $\hat{I}_{0t} = I_{0\tau} \cdot I_{\tau t}$ mit $t \geq \tau$

benachbarter Zeitperioden (Kettenindex):

$$\hat{I}_{0t} = I_{01} \cdot I_{12} \cdot ... \cdot I_{t-1,t} = \prod_{\tau=1}^{t} I_{\tau-1,\tau}$$

Beispiel: $\hat{P}^L_{0t} = P^L_{01} \cdot P^L_{12} \cdot ... \cdot P^L_{t-1,t}$ Es gilt: $\hat{P}^L_{0t} \neq P^L_{0t}$

Wenn bei der Umbasierung $\hat{I}_{\tau t} \neq I_{\tau t}$ gilt, dann gilt dies auch bei der Verkettung und vice versa.

mehrerer Wertindizes:

$$W_{0t} = W_{01} \cdot W_{12} \cdot ... \cdot W_{t-1,t} = \prod_{\tau=1}^{t} W_{\tau-1,\tau}$$

2.2 Konzentrations- und Disparitätsmessung

Bedingungen:
1. Merkmal X ist kardinalskaliert und alle Werte sind nichtnegativ.
2. Die Merkmalswerte sind aufsteigend geordnet: $x_1 \leq x_2 \leq ... \leq x_n$
3. Die Gesamtmerkmalssumme ist größer als null: $x_1 + ... + x_n > 0$

Symbole:

i: Nummer des sortierten Merkmalsträger bzw. Klassenangabe
x_i: Merkmalswert des i-ten Merkmalsträgers $(i = 1, ..., n)$
p_i: Anteil von i an der Gesamtmerkmalssume $(i = 1, ..., n)$
z_i: kumulierte Anteile der Merkmalsträger (Abszisse)
y_i: kumulierte relative Merkmalssummen (Ordinate)

2.2.1 absolute Konzentration

Konzentrationsraten:

$$CR_h = \sum_{i=1}^{h} p_{n-i+1} \quad \text{mit } p_i = x_i / \sum_{j=1}^{n} x_j \qquad (h = 1, ..., n)$$

Herfindahl(Hirschmann)-Index:

$$H = \sum_{i=1}^{n} p_i^2 \quad \text{mit } p_i = x_i / \sum_{j=1}^{n} x_j \qquad \text{Es gilt: } 1/n \leq H \leq 1$$

$x_1, ..., x_n$ sei der Markanteil von Unternehmen. Dann gilt:

$H = 1/n$	$\Rightarrow$	gleicher Marktanteil für alle Unternehmen
$H = 1$	$\Rightarrow$	die ganze Marktmacht konzentriert sich auf einen Monopolisten

2.2.2 Disparität (relative Konzentration)

Lorenzkurve:

1. bei geordneten Einzeldaten $x_1 \leq ... \leq x_n$:

$$z_i = i/n \qquad (i = 1, ..., n)$$

$$y_i = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{n} x_j} = \sum_{j=1}^{i} p_j \qquad (i = 1, ..., n)$$

2. bei Vorgabe von absoluten bzw. relativen Häufigkeiten:

Gegegeben sind k geordnete Merkmalswerte $a_1 < ... < a_k$ mit den absoluten bzw. relativen Häufigkeiten $n_1, ..., n_k$ bzw. $f_1, ..., f_k$.

$$z_i = \sum_{j=1}^{i} n_j/n = \sum_{j=1}^{i} f_j \quad \text{mit } n = \sum_{j=1}^{k} n_j \qquad (i = 1, ..., k)$$

$$y_i = \frac{\sum\limits_{j=1}^{i} n_j a_j}{\sum\limits_{j=1}^{k} n_j a_j} = \frac{\sum\limits_{j=1}^{i} f_j a_j}{\sum\limits_{j=1}^{k} f_j a_j} \qquad (i = 1, ...k)$$

3. bei gruppierten Daten (klassifizierte Daten):

Zu k aneinandergrenzenden Klassen $[c_0, c_1), [c_1, c_2), ..., [c_{k-1,c_k})$ sind die absoluten bzw. relativen Häufigkeiten $n_1, ..., n_k$ bzw. $f_1, ..., f_k$ gegeben.

Fall a) Die Merkmalswertsummen $n_i\overline{x}_i$ sind für jede Klasse bekannt.

$$z_i = \sum_{j=1}^{i} n_i/n = \sum_{j=1}^{i} f_i \quad \text{mit } n = \sum_{j=1}^{k} n_j \qquad (i = 1, ..., k)$$

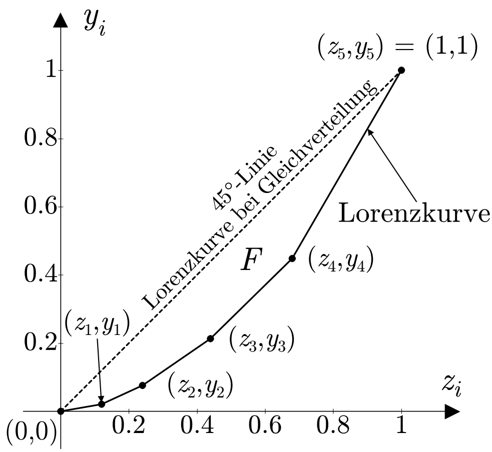
$$y_i = \frac{\sum\limits_{j=1}^{i} n_j \overline{x}_j}{\sum\limits_{j=1}^{k} n_j \overline{x}_j} \qquad (i = 1, ...k)$$

Fall b) Die Merkmalswertsummen $n_i\overline{x}_i$ sind unbekannt. Man verwendet die Klassenmitten $m_i = (c_i + c_{i-1})/2$.

$$z_i = \sum_{j=1}^{i} n_i/n = \sum_{j=1}^{i} f_i \quad \text{mit } n = \sum_{j=1}^{k} n_j \qquad (i = 1, ..., k)$$

$$y_i = \frac{\sum\limits_{j=1}^{i} n_j m_j}{\sum\limits_{j=1}^{k} n_j m_j} \qquad (i = 1, ...k)$$

Die Lorenzkurve ergibt sich als Streckenzug durch die Punkte $(0, 0), (z_1, y_1), ..., (z_k, y_k)$ bzw. (z_n, y_n), wobei (z_k, y_k) bzw. $(z_n, y_n) = (1, 1)$ ist.



$F =$ Fläche zwischen der 45°-Linie und der Lorenzkurve

Es gilt: Je größer F ist, desto größer ist die Disparität.

Gini-Koeffizient:

Der Gini-Koeffizient G ist folgendermaßen definiert:

$$G = \frac{\text{Fläche zwischen 45°-Linie und Lorenzkurve}}{\text{Fläche zwischen 45°-Linie und z-Achse}} = 2 \cdot F$$

<u>1. bei einer geordneten Urliste $x_1 \leq ... \leq x_n$:</u>

$$G = \frac{2 \cdot \sum\limits_{i=1}^{n} i \cdot x_i}{n \cdot \sum\limits_{i=1}^{n} x_i} - \frac{n+1}{n} = \frac{1}{n}\left(2 \cdot \sum_{i=1}^{n} i \cdot p_i - (n+1)\right) \quad \text{mit } p_i = \frac{x_i}{\sum\limits_{j=1}^{n} x_j}$$

<u>2. bei Vorgabe von absoluten bzw. relativen Häufigkeiten:</u>

Bezeichnungen siehe Seite 18 (Lorenzkurve–Punkt 2).

$$G = \frac{\sum\limits_{i=1}^{k} (z_{i-1} + z_i) n_i a_i}{\sum\limits_{i=1}^{k} n_i a_i} - 1$$

2.3 Zeitreihenanalyse

Eine emprische Zeitreihe besteht aus n zeitlich geordneten Beobachtungswerten y_t mit $t = 1, 2, ..., n$.

2.3.1 Komponentenmodelle

Komponenten einer Zeitreihe:

m_t: Trendkomponente
k_t: Konjunkturkomponente

g_t: glatte Komponente ($g_t = m_t + k_t$ bzw. $g_t = m_t \cdot k_t$)
s_t: Saisonkomponente
e_t: Restkomponente

Additives Modell: $y_t = g_t + s_t + e_t = m_t + k_t + s_t + e_t$

Multiplikatives Modell: $y_t = g_t \cdot s_t \cdot e_t = m_t \cdot k_t \cdot s_t \cdot e_t$
$\Leftrightarrow \ln y_t = \ln g_t + \ln s_t + \ln e_t$

Saisonbereinigung: $y_t^{saisonbereinigt} = y_t - s_t = g_t + e_t$

Beispiele für Trendfunktionen:

Linearer Trend:	$m_t = a + bt$
Quadratischer Trend:	$m_t = a + b_1 t + b_2 t^2$
Polynomialer Trend:	$m_t = a + b_1 t + ... + b_q t^q$
Exponentieller Trend:	$m_t = a \cdot \exp(bt) = a \cdot e^{bt}$
Logistischer Trend:	$m_t = \dfrac{a}{b_1 + \exp(b_2 t)} \quad (b_1 > 0, b_2 < 0)$

2.3.2 Schätzung der Trendkomponente

Methode der kleinsten Quadrate: (siehe auch Seite 14)

Dies ist eine globaler Ansatz. Es werden alle Beobachtungswerte für die Berechnung des Trends verwendet. Ein globaler Ansatz sollte bei Zeitreihen angewendet werden, bei denen die Parameter der Trendfunktion über den ganzen Beobachtungszeitraum hinweg konstant sind.

bei einer linearen Trendfunktion $g_t = a + bt$:

Es gilt: $\bar{t} = (n+1)/2$ und $s_t^2 = (n^2-1)/12$

$$\hat{b} = \frac{s_{ty}}{s_t^2} = \frac{\sum_{t=1}^{n}(y_t-\overline{y})(t-\bar{t})}{\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{\frac{1}{n}\sum_{t=1}^{n} y_t \cdot t - \overline{y} \cdot \bar{t}}{\frac{1}{n}\sum_{t=1}^{n} t^2 - \bar{t}^2} = \frac{\frac{1}{n}\sum_{t=1}^{n} y_t(t-\bar{t})}{\frac{1}{12}(n^2-1)} = \frac{12 \cdot \sum_{t=1}^{n} y_t(t-\bar{t})}{n(n^2-1)}$$

$$\hat{a} = \overline{y} - \hat{b} \cdot \bar{t} = \overline{y} - \hat{b} \cdot \frac{n+1}{2}$$


wiwi-journal
Das monatliche Online-Magazin
für den Management-Nachwuchs.
www.wiwi-journal.de

geschätzte Trendfunktion: $\hat{g}_t = \hat{a} - \hat{b}t$

Prognosegleichung (wenn kein Saisontrend vorliegt, also $y_t = g_t + e_t$):

$\hat{y}_{n+i} = \hat{a} + \hat{b}(n+i) \qquad (1 \leq i \leq 10)$

Einfacher und zentrierter gleitender Durchschnitt:

Ein gleitender Durchschnitt ist ein lokaler Ansatz, der im Falle sich verändernder Entwicklungsmuster angewendet werden sollte. Durch Verwendung weniger benachbarter Beobachtungswerte werden lokale Trends gleitend über den Beobachtungszeitraum hinweg berechnet.

<u>ungerader Ordnung $q = 2k+1$:</u>

$$\hat{g}_t = \frac{1}{2k+1} \sum_{i=-k}^{k} y_{t+i} \qquad (t = k+1, ..., n-k)$$

<u>gerader Ordnung $q = 2k$:</u>

$$\hat{g}_t = \frac{1}{2k} \left(\frac{1}{2} y_{t-k} + \sum_{i=-k+1}^{k-1} y_{t+i} + \frac{1}{2} y_{t+k} \right) \qquad (t = k+1, ..., n-k)$$

Gewichteter gleitender Durchschnitt:

λ_i: Gewichte $\qquad y_t$: Input $\qquad \hat{g}_t$: Output oder gefilterte Reihe

$$\hat{g}_t = \sum_{i=-k}^{s} \lambda_i \cdot y_{t+i} \qquad (t = k+1, ..., n-s) \qquad \text{mit} \sum_{i=-k}^{s} \lambda_i = 1$$

2.3.3 Schätzung der Saisonkomponente

Phasendurchschnittsverfahren:

<u>Definitionen:</u>

i: Saisonindex $\quad (i = 1, ..., p) \quad$ (z.B. ist bei Monatswerten $p = 12$)
j: Periodenindex $\quad (j = 1, ..., l)$
t: Beobachtungswertindex $\quad (t = 1, ..., n)$
y_{ij}: Beobachtungswert in Saison i und Periode j

Annahmen:

additives Modell: $y_t = g_t + s_t + e_t \qquad (t = 1, ..., n)$

konstante (stabile) Saisonfigur: $s_t = s_{t+p} \qquad (t = 1, ..., n - p)$

Ausgleich der Saisonausschläge: $s_t + s_{t+1} + ... + s_{t+p-1} = 0$
$(t = 1, ..., n - p + 1)$

Schritt 1: Die glatten Komponenten $\hat{g}_{ij}$ werden geschätzt (z.B. durch die Methode der Kleinsten Quadrate oder gleitende Duchschnitte).

Schritt 2: Berechnung einer trendbereinigten Zeitreihe durch Bildung der Differenzen zwischen den ursprünglichen und den geglätteten Zeitreihenwerten.

$$\hat{s}_{ij} = y_{ij} - \hat{g}_{ij} \qquad (t = 1, ..., n)$$

Schritt 3: Schätzung der rohen Saisonkomponenten

$$\hat{s}_i^{roh} = \frac{1}{l} \sum_{j=1}^{l} \hat{s}_{ij} \qquad (i = 1, ..., p)$$

Schritt 4: Normierung der rohen Saisonkomponenten

$$\hat{s}_i = \hat{s}_i^{roh} - \overline{\hat{s}^{roh}} \qquad (i = 1, ..., p) \qquad \text{mit} \quad \overline{\hat{s}^{roh}} = \frac{1}{p} \sum_{i=1}^{p} \hat{s}_i^{roh}$$

Schritt 5: Ermittlung der saisonbereinigten Zeitreihe, indem man die normierten Saisonabweichungen $\hat{s}_i$ von den Ursprungswerten y_{ij} subtrahiert.

$$\hat{y}_{ij}^{saisonbereinigt} = y_{ij} - \hat{s}_i$$

Prognosegleichung: $\hat{y}_{n+i} = \hat{a} + \hat{b}(n + i) + \hat{s}_{n+i} \qquad (i \geq 1)$

2.3.4 Exponentielles Glätten

Einfaches exponentielles Glätten:

Bemerkung: Anwendbar bei Zeitreihen ohne Trend und Saison

$$g_t^e = \beta g_{t-1}^e + (1 - \beta) y_t \qquad (t = 2, 3, ..., n)$$

Rekursionsformel:

$g_t^e = (1-\beta)\left[y_t + \beta_{t-1} + \beta^2 y_{t-2} + ... + \beta^{t-2} y_2\right] + \beta^{t-1} y_1$

$\beta \in]0;1[$ heißt Glättungsparameter (üblicher Wert: $0,6 \leq \beta \leq 0,9$)

Startwert: $g_1^e = y_1$

Exponentielles Glätten nach Holt-Winters:

Bemerkung: Anwendbar bei Zeitreihen mit Trend

$g_t^H = \beta(g_{t-1}^H + b_{t-1}) + (1-\beta) y_t$

mit $b_t = \alpha b_{t-1} + (1-\alpha)(g_t^H - g_{t-1}^H) \qquad (t = 2, ..., n)$

Es gilt: $0 < \alpha < 1$ und $0 < \beta < 1$
(übliche Werte: $0,5 \leq \alpha, \beta \leq 0,9$)

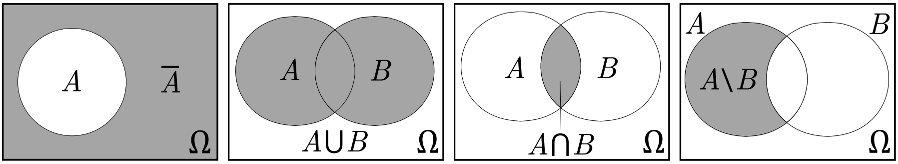
Startwerte: $g_1^H = y_1$ und $b_1 = 0$

Prognosegleichung: $\hat{y}_{n+i} = g_n^H + b_n i \qquad 1 \leq i \leq 10$

3 Wahrscheinlichkeitsrechnung

Symbole:

$\Omega = \{e_1, e_2, ..., e_m\}$:	Ereignisraum (Ergebnismenge)
$\{e_i\}$:	Elementarereignis $(i = 1, 2, ..., m)$
$\|\Omega\|$:	Anzahl der Elementarereignisse im Ereignisraum (m)
$\emptyset$:	unmögliches Ereignis
A:	zufälliges Ereignis A (Teilmenge von Ω)
$\overline{A}$:	Komplement (Gegenereignis) zu A
$A \cup B$:	A und/oder B (Vereinigung von A und B)
$A \cap B$:	A und B (Durchschnitt von A und B)
$A \backslash B$:	A aber nicht B (Differenz zwischen A und B)
$P(A)$:	Wahrscheinlichkeit dafür, dass Ereignis A eintritt
$P(A\|B)$:	Bedingte Wahrscheinlichkeit (A wenn B)



3.1 Elemente der Mengenlehre

Morgansche Formeln:

$\overline{A \cap B} = \overline{A} \cup \overline{B}$ und $\overline{A \cup B} = \overline{A} \cap \overline{B}$

Distributivgesetz:

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ und
$(A \cup B) \cap C = (A \cap B) \cup (A \cap C)$

Kommutativgesetz:

$A \cap B = B \cap A$ und $A \cup B = B \cup A$

Assoziativgesetz:

$(A \cap B) \cap C = A \cap (B \cap C)$ und
$(A \cup B) \cup C = A \cup (B \cup C)$

3.2 Wahrscheinlichkeiten

Axiome von Kolmogorov:

Axiom 1: $P(A) \geq 0$ für alle $A \subseteq \Omega$ (Nicht-Negativität)

Axiom 2: $P(\Omega) = 1$ (Normierung)

Axiom 3: $P(A \cup B) = P(A) + P(B)$
wenn $A \cap B = \emptyset$ (disjunkt) (Additivität)

Folgerungen:

$P(\overline{A}) = 1 - P(A)$

$P(\emptyset) = 0$

$P(A \backslash B) = P(A) - P(A \cap B)$

Additionssatz:

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

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

Bedingte Wahrscheinlichkeit:

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

Multiplikationssatz:

$P(A \cap B) = P(A|B) \cdot P(B) \qquad \text{mit } P(A) > 0$

$P(A \cap B) = P(B|A) \cdot P(A) \qquad \text{mit } P(B) > 0$

Für die Ereignisse $A_1, A_2, ..., A_n$ mit $P(A_1 \cap ... \cap A_{n-1}) > 0$ gilt:

$$P(A_1 \cap ... \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$
$$\cdot ... \cdot P(A_n|A_1 \cap ... \cap A_{n-1})$$

Unabhängigkeit von zwei Ereignissen:

A und B unabhängig $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

$\Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$

Totale Wahrscheinlichkeit:

Wenn $B_1 \cup B_2 \cup ... \cup B_n = \Omega$ (vollständiges System)
und $B_i \cap B_j = \emptyset$ für alle $i \neq j$ (disjunkt)

dann gilt: $P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$

Satz von Bayes:

Wenn $B_1 \cup B_2 \cup ... \cup B_n = \Omega$ (vollständiges System)
und $B_i \cap B_j = \emptyset$ für alle $i \neq j$ (disjunkt)

dann gilt: $$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum\limits_{i=1}^{n} P(A|B_i) \cdot P(B_i)} \qquad (k = 1, ..., n)$$

Laplace-Prozess:

$$P(e_i) = \frac{1}{|\Omega|} = \frac{1}{m} \quad (i = 1, ..., m) \quad \Leftrightarrow \quad P(e_1) = P(e_2) = ... = P(e_m)$$

Ein Laplace-Prozess ist ein Zufallsprozess, bei dem die Eintrittswahrscheinlichkeit für alle Elementarereignisse gleich ist.

Laplace-Formel:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n} \quad \text{mit } |A| = \text{Anzahl der der günstigen Fälle für A}$$

3.3 Kombinatorik

Fakultät: $N! = 1 \cdot 2 \cdot 3 \cdot ... \cdot N$

Binomialkoeffizient: $\binom{N}{n} = \frac{N!}{n! \cdot (N-n)!}$ („N über n“)

Anzahl verschiedener Reihenfolgen von N Elementen, die in k Gruppen mit jeweils $\mathbf{n_1, n_2, ..., n_k}$ gleichen Elementen aufgeteilt werden können:

$$\frac{N!}{n_1! \cdot n_2! \cdot ... \cdot n_k!} \quad \text{mit } N = \sum_{i=1}^{k} n_i$$

Anzahl der möglichen Stichproben vom Umfang n aus N verschiedenen Elementen:

	mit Berücksichtigung der Reihenfolge	ohne Berücksichtigung der Reihenfolge
mit Zurücklegen	N^n	$\binom{N+n-1}{n}$
ohne Zurücklegen	$\frac{N!}{(N-n)!}$	$\binom{N}{n}$

4 Zufallsvariable und theoretische Verteilungen

Symbole:

x_i:	möglicher Wert von X $(i = 1, ..., k)$	$V(X)$:	Varianz σ^2
x_p:	p-Quantil	$f(x, y)$:	gemeinsame Dichte- bzw. Wahrscheinlichkeitsfunktion
$f(x)$:	Wahrscheinlichkeits- bzw. Dichtefunktion	$f_X(x)$:	Randverteilung bzw. Randdichte
$F(X)$:	Verteilungsfunktion	$\text{Cov}(X, Y)$:	Kovarianz σ_{XY}
$E(X)$:	Erwartungswert μ	σ:	Standardabweichung

4.1 Zufallsvariable

4.1.1 Diskrete Zufallsvariable

Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} P(X = x_i) = p_i & \text{für } x = x_i \qquad (i = 1, ..., k) \\ 0 & \text{sonst} \end{cases}$$

mit Eigenschaften: 1.) $\sum_{i=1}^{k} p_i = 1$ 2.) $p_i \geq 0$ für alle i

Verteilungsfunktion: $F(x) = P(X \leq x) = \sum\limits_{x_i \leq x} f(x_i)$

p-Quantil: $x_p = \min(x|F(x) = P(X \leq x) \geq p)$

Unabhängigkeit von zwei diskreten Zufallsvariablen:

Zwei diskrete Zufallsvariablen X und Y sind unabhängig, wenn für alle möglichen Kombinationen von x und y gilt:

$P(X = x \text{ und } Y = y) = P(X = x) \cdot P(Y = y)$

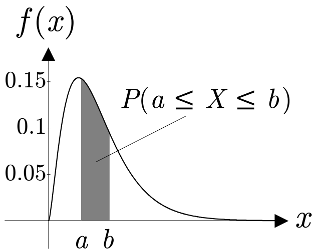
4.1.2 Stetige Zufallsvariable

Dichtefunktion:

$f(x) = \dfrac{\partial F(x)}{\partial x}$ mit Eigenschaften:

1.) $\int\limits_{-\infty}^{+\infty} f(x)\, dx = 1$

2.) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$



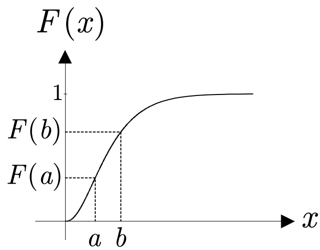
Verteilungsfunktion:

$F(x_0) = P(X \leq x_0) = \int\limits_{-\infty}^{x_0} f(x)dx$

mit Eigenschaften:

1.) $\lim\limits_{x \to -\infty} F(x) = 0 \qquad \lim\limits_{x \to +\infty} F(x) = 1$

2.) $F(b) - F(a) = P(a \leq X \leq b)$



Im stetigen Fall gilt: $P(X = x) = 0$ für jedes $x \in \mathbb{R}$. Daraus folgt:

$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$

p-Quantil: $x_p = F^{-1}(p)$ Es gilt: $P(X \leq x_p) = F(x_p) = p$

Unabhängigkeit von zwei stetigen Zufallsvariablen:

Zwei stetige Zufallsvariablen X und Y sind unabhängig, wenn für alle $x \in \mathbb{R}$ und $y \in \mathbb{R}$ gilt:

$$P(X \leq x \text{ und } Y \leq y) = P(X \leq x) \cdot P(Y \leq y) = F(x) \cdot F(y)$$

4.2 Maßzahlen theoretischer Verteilungen

4.2.1 Erwartungswert

diskrete Zufallsvariable:

$$E(X) = \mu = \sum_{i=1}^{k} x_i \cdot f(x) = \sum_{i=1}^{k} x_i \cdot P(X = x_i) = \sum_{i=1}^{k} x_i \cdot p_i$$

stetige Zufallsvariable:

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

allgemein transformierte Zufallsvariable X:

Transformation: $Y = g(X)$

wobei $g(X)$ eine beliebige reelwertige Funktion von X ist

$$E(g(X)) = E(Y) = \begin{cases} \sum_{i=1}^{k} g(x_i) \cdot f(x) & \text{X ist diskret} \\ \int_{-\infty}^{+\infty} g(x) \cdot f(x)dx & \text{X ist stetig} \end{cases}$$

linear transformierte Zufallsvariable X:

lineare Transformation: $Y = a + bX$ (a und b sind Konstanten)

$$E(a + bX) = E(Y) = a + b \cdot E(X)$$

Summe gewichteter Zufallsvariablen:

$$Y = \sum_{i=1}^{n} a_i X_i = a_1 X_1 + a_2 X_2 + ... + a_n X_n \quad \text{(alle } a_i \text{ sind Konstanten)}$$

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

4.2.2 Varianz

allgemein:

$$V(X) = \sigma^2 = E[(X - E(X))^2] = E(X^2) - E(X)^2 = E(X^2) - \mu^2$$

diskrete Zufallsvariable:

$$V(X) = \sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot f(x) = \sum_{i=1}^{k} x_i^2 \cdot f(x) - \mu^2$$

stetige Zufallsvariable:

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

einer Konstanten: $V(a) = 0$

linear transformierte Zufallsvariable X:

lineare Transformation: $Y = a + bX$ (a und b sind Konstanten)

$V(a + bX) = V(Y) = b^2 \cdot V(X)$

Summe gewichteter Zufallsvariablen:

$Y = \sum_{i=1}^{n} a_i X_i = a_1 X_1 + a_2 X_2 + ... + a_n X_n$ (alle a_i sind Konstanten)

allgemein gilt:

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j \cdot \mathrm{Cov}(X_i, X_j)$$

bei paarweise unabhängigen Zufallsvariablen gilt (Spezialfall):

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = V(Y) = \sum_{i=1}^{n} a_i^2 V(X_i)$$

Standardabweichung: $\sigma = \sqrt{\sigma^2} = \sqrt{V(X)}$

Variationskoeffizient: $v = \dfrac{\sqrt{V(X)}}{E(X)} = \dfrac{\sigma}{\mu}$

4.3 Zweidimensionale Zufallsvariablen

Wahrscheinlichkeitsfunktion bzw. Dichtefunktion:

Wahrscheinlichkeitsfunktion (X und Y sind diskret):

$$f(x,y) = \begin{cases} P(X = x \text{ und } Y = y) & \text{für alle möglichen } (x,y) \\ 0 & \text{sonst} \end{cases}$$

Dichtefunktion (X und Y sind stetig):

$f(x,y)$ wobei gilt: $P(a \leq X \leq b \text{ und } c \leq Y \leq d) = \int\limits_a^b \int\limits_c^d f(x,y)dxdy$

Verteilungsfunktion:

X und Y sind diskret:

$$F(x,y) = P(X \leq x \text{ und } Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j)$$

X und Y sind stetig:

$$F(x,y) = P(X \leq x \text{ und } Y \leq y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u,v)dudv$$

Randverteilung bzw. Randdichte:

X und Y sind diskret: $f_X(x) = \sum_j f(x, y_j) = P(X = x)$

$f_Y(y) = \sum_i f(x_i, y) = P(Y = y)$

X und Y sind stetig: $f_X(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy$

$$f_Y(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx$$

Randverteilungsfunktion:

X und Y sind diskret: $F_X(x) = P(X \leq x) = \sum\limits_{x_i \leq x} \sum\limits_{y_j} f(x_i, y_j)$

$$F_Y(y) = P(Y \leq y) = \sum\limits_{x_i} \sum\limits_{y_j \leq y} f(x_i, y_j)$$

X und Y sind stetig: $F_X(x) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{x} f(u,v)dudv$

$$F_Y(y) = \int\limits_{-\infty}^{y} \int\limits_{-\infty}^{\infty} f(u,v)dudv$$

Bedingte Verteilungen:

bedingte Dichte- bzw. Wahrscheinlichkeitsfunktion von X:

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

bedingte Dichte- bzw. Wahrscheinlichkeitsfunktion von Y:

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

Kovarianz:

Definition: $\sigma_{XY} = \mathrm{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$

$$= E(X \cdot Y) - E(X) \cdot E(Y)$$

X und Y sind diskret:

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = \sum_i \sum_j (x_i - E(X))(y_j - E(Y)) \cdot f(x_i, y_j)$$

$$= \sum_i \sum_j x_i y_j \cdot f(x_i, y_j) - E(X) \cdot E(Y)$$

X und Y sind stetig:

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E(X))(y-E(Y))\cdot f(x,y)dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f(x,y)dxdy - E(X)\cdot E(Y)$$

Korrelationskoeffizient:

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\cdot\sqrt{\mathrm{V}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y} \qquad \text{mit} \quad -1 \leq \rho \leq 1$$

Es gilt:
$\rho_{XY} > 0$ → X und Y sind positiv korreliert
$\rho_{XY} < 0$ → X und Y sind negativ korreliert
$\rho_{XY} = 0$ → X und Y sind unkorreliert

Erwartungswert der Summe/Differenz zweier Zufallsvariablen:

$$E(X+Y) = V(X) + V(Y) \qquad \text{bzw.} \qquad E(X-Y) = V(X) - V(Y)$$

Varianz der Summe/Differenz zweier Zufallsvariablen:

$$V(X \pm Y) = V(X) + V(Y) \pm 2\cdot\mathrm{Cov}(X,Y)$$

Unabhängigkeit von zwei Zufallsvariablen:

X und Y sind unabhängig voneinander, wenn für alle x und y gilt:

$$f(x,y) = f_X(x)\cdot f_Y(y)$$

Äquivalent dazu gilt: $F(x,y) = F_X(x)\cdot F_Y(y)$

bei Unabhängigkeit gilt:

$$\mathrm{Cov}(X,Y) = \rho_{XY} = 0$$
$$E(X\cdot Y) = E(X)\cdot E(Y)$$
$$V(X \pm Y) = V(X) + V(Y)$$

5 Spezielle Verteilungsmodelle

Symbole:

$X \sim$: Kurzschreibweise für die Verteilung der Zufallsvariablen X
W_X: Wertebereich der Zufallsvariablen X
$f(x)$: Wahrscheinlichkeits- bzw. Dichtefunktion von X
$F(X)$: Verteilungsfunktion der Zufallsvariablen X
$E(X)$: Erwartungswert der Zufallsvariablen X
$V(X)$: Varianz der Zufallsvariablen X
$\Phi(z)$: Verteilungsfunktion der Standardnormalverteilung

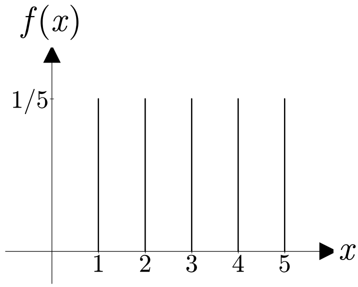
5.1 Diskrete Verteilungen

diskrete Gleichverteilung:

$X \sim G(x_1, ..., x_m) \quad x_i \in \mathbb{R} \quad (i = 1, ..., m)$

$W_X = \{x_1, ..., x_m\}$

$$f(x) = \begin{cases} 1/m & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$



Falls $x_i = i \quad (i = 1, ..., m)$: (s. Grafik)

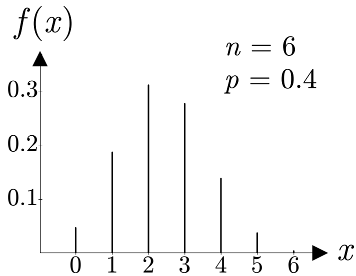
$$E(X) = \frac{m+1}{2} \qquad V(X) = \frac{m^2-1}{12}$$

Binomialverteilung:

$X \sim B(n; p) \qquad n \in \mathbb{N}, p \in \mathbb{R}\,]0; 1[$

$W_X = \{0, 1, ..., n\}$

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$



$$E(X) = n \cdot p \qquad V(X) = n \cdot p \cdot (1-p)$$

Additionseigenschaft: Wenn alle $X_t \sim B(n_t; p)$ und unabhängig sind,

$$\text{dann gilt:} \quad \sum_t X_t \sim B\left(\sum_t n_t; p\right)$$

Bernoulli-Verteilung: Spezialfall der Binomialverteilung $(n = 1)$

$X \sim B(1; p) \qquad p \in \mathbb{R}\,]0; 1[$

$W_X = \{0, 1\}$

$$f(x) = \begin{cases} 1 - p & \text{für } x = 0 \\ p & \text{für } x = 1 \end{cases}$$

$E(X) = p \qquad V(X) = p(1 - p)$

Hypergeometrische Verteilung:

$X \sim H(N; M; n) \qquad N, M, n \in \mathbb{N}; n \leq N; M \leq N$

$W_X = \{\max(0, n - (N - M)), ..., \min(n, M)\}$

$$f(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N - n}{N - 1}$$

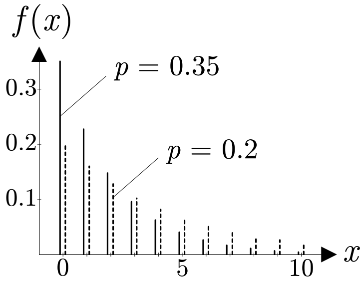
Geometrische Verteilung:

X beschreibt die Anzahl der Misserfolge (Wahrscheinlichkeit:1-p), bis der erste Erfolg mit der Wahrscheinlichkeit p auftritt.

$X \sim GV(p) \qquad p \in \mathbb{R}\,]0; 1[$

$W_X = \mathbb{N}_0 = \{0, 1, 2, ...\}$

$$f(x) = \begin{cases} (1-p)^x \cdot p & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$



$$F(x) = 1 - (1-p)^{x+1} \qquad \text{für } x \in W_X$$

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

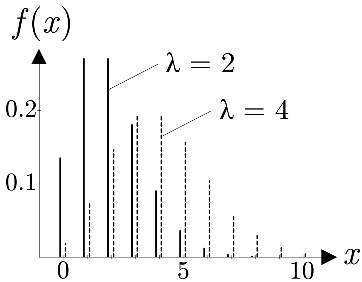
$$V(X) = \frac{1-p}{p^2}$$

Poisson-Verteilung:

$X \sim P(\lambda) \qquad \lambda \in \mathbb{R}^+$

$W_X = \mathbb{N}_0 = \{0, 1, 2, ...\}$

$$f(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x!} & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$



$E(X) = \lambda$

$V(X) = \lambda$

Additionseigenschaft: Wenn alle $X_t \sim P(\lambda_t)$ und unabhängig sind,

dann gilt: $\sum_t X_t \sim P(\sum_t \lambda_t)$

5.2 stetige Verteilungen

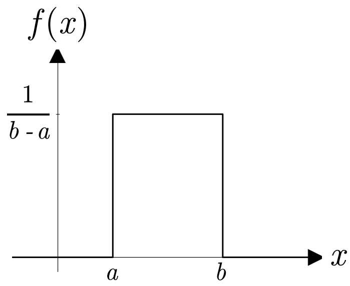
Rechteckverteilung (stetige Gleichverteilung):

$X \sim R(a; b) \qquad a, b \in \mathbb{R}; a < b$

$W_X = [a; b]$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{für } x < a \\ \dfrac{x-a}{b-a} & \text{für } x \in W_X \\ 1 & \text{für } x > b \end{cases}$$



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

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

Exponentialverteilung:

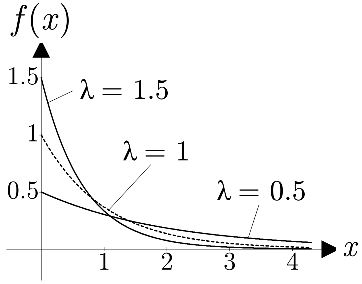
$X \sim E(\lambda) \qquad \lambda \in \mathbb{R}^+$

$W_X = \mathbb{R}^+{}_0 = [0; \infty[$

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{für } x \in W_X \\ 0 & \text{sonst} \end{cases}$$

$F(x) = 1 - e^{-\lambda x} \qquad \text{für } x \in W_X$

$$E(x) = \frac{1}{\lambda} \qquad V(X) = \frac{1}{\lambda^2}$$



Normalverteilung:

allgemein:

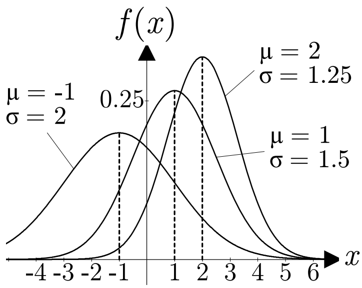
$X \sim N(\mu; \sigma^2) \qquad \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$

$W_X = \mathbb{R} =]-\infty; +\infty[$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$F(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

$$E(X) = \mu \qquad V(X) = \sigma^2$$



lineare Transformation: $Y = a + bX \quad (a \text{ und } b \text{ konstant})$

Ist $X \sim N(\mu, \sigma^2)$, dann gilt:
$\Rightarrow Y \sim N(a + b\mu; b^2\sigma^2)$

gewichtete Summe: $Y = \sum_{t=1}^{n} a_t X_t \quad (a_t \text{ konstant})$

Sind alle $X_t \sim N(\mu_t, \sigma_t^2) \quad (t = 1, ..., n)$
und unabhängig, dann gilt:

$$\Rightarrow Y \sim N\left(\sum_{t=1}^{n} a_t\mu_t; \sum_{t=1}^{n} a_t^2\sigma_t^2\right)$$

Additionseigenschaft: Sind alle $X_t \sim N(\mu_t, \sigma_t^2) \quad (t = 1, ..., n)$
und unabhängig, dann gilt:

$$\Rightarrow \sum_{t=1}^{n} X_t \sim N\left(\sum_{t=1}^{n} \mu_t; \sum_t \sigma_t^2\right)$$

Spezialfall der Normalverteilung: Standardnormalverteilung

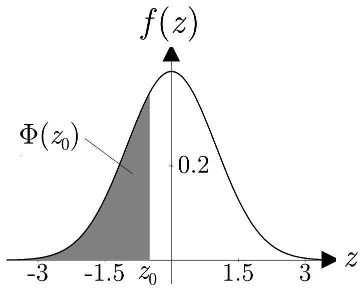
$Z \sim N(0;1)$

$W_Z = \mathbb{R} =]-\infty; +\infty[$

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

$$F(z_0) = \Phi(z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{1}{2}z^2} \, dz$$

$E(Z) = 0 \qquad V(X) = 1$



Symmetrie: $\Phi(-z) = 1 - \Phi(z) \quad \forall z \in \mathbb{R}$

Symmetrie der Quantile: $z_\alpha = -z_{1-\alpha}$ für alle $\alpha \in]0;1[$

symmetrisches Intervall: $P(-k \le Z \le k) = P(|Z| < k) = 2 \cdot \Phi(k) - 1$

mit $k \in \mathbb{R}$

Standardisierung einer normalverteilten Zufallsvariablen:

Standardisierung: $X \sim N(\mu;\sigma^2) \quad \Rightarrow \quad Z = \dfrac{X-\mu}{\sigma} \sim N(0;1)$

Durch die Standardisierung kann jede beliebig normalverteilte Zufallsvariable X in eine standardnormalverteilte Zufallsvariable Z mit den Parametern $\mu = 0$ und $\sigma = 1$ umgewandelt werden.

Verteilungsfunktion:

$$F(z) = F(x) = \Phi(z) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$
$$= P(Z \leq z) = P(X \leq x)$$

Wahrscheinlichkeiten:

$$P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$
$$P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$
$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Quantile: $x_\alpha = \mu + \sigma \cdot z_\alpha \qquad z_\alpha = \dfrac{x_\alpha - \mu}{\sigma}$

Chi-Quadrat-Verteilung:

Seien $Z_1, ..., Z_n$ alle $N(0;1)$ und unabhängig, dann gilt:

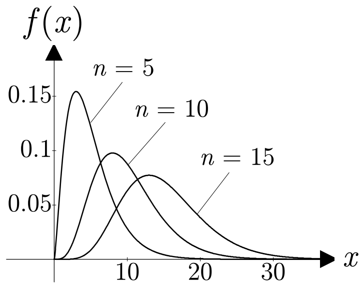
$X = \sum_{t=1}^{n} Z_t^2$ ist Chi-Quadrat-verteilt

mit n Freiheitsgraden

$X \sim \chi^2(n) \qquad n \in \mathbb{N}$

$W_X = \mathbb{R}^+{}_0 = [0;\infty[$

$E(X) = n \qquad V(X) = 2n$



Additionseigenschaft: Sind alle $X_t \sim \chi^2(n_t) \quad (t = 1, ..., n)$ und unabhängig, dann gilt:

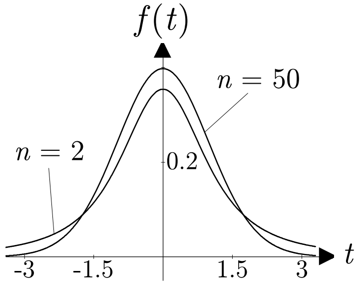
$$\Rightarrow \sum_{t=1}^{n} X_t \sim \chi^2\left(\sum_{t=1}^{n} n_t\right)$$

t-Verteilung (Studentverteilung):

Sei $Z \sim N(0;1)$, $X \sim \chi^2(n)$ und
Z und X unabhängig, dann gilt:

$T = \dfrac{Z}{\sqrt{X/n}}$ ist t-verteilt

mit n Freiheitsgraden



$T \sim t(n) \qquad n \in \mathbb{N}$

$W_T = \mathbb{R} = \,]-\infty; +\infty[$

$E(T) = 0 \qquad V(T) = \dfrac{n}{n-2} \quad (n > 2)$

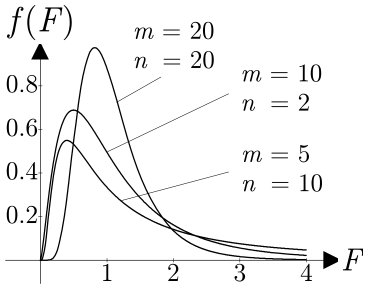
Symmetrie der Quantile: $t_\alpha = -t_{1-\alpha}$ für alle $\alpha \in \,]0;1[$

Fisher(F)-Verteilung:

Sei $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$
und unabhängig. Dann gilt:

$F = \dfrac{X/m}{Y/n}$ ist Fisher(F)-verteilt

mit den Freiheitsgraden m und n

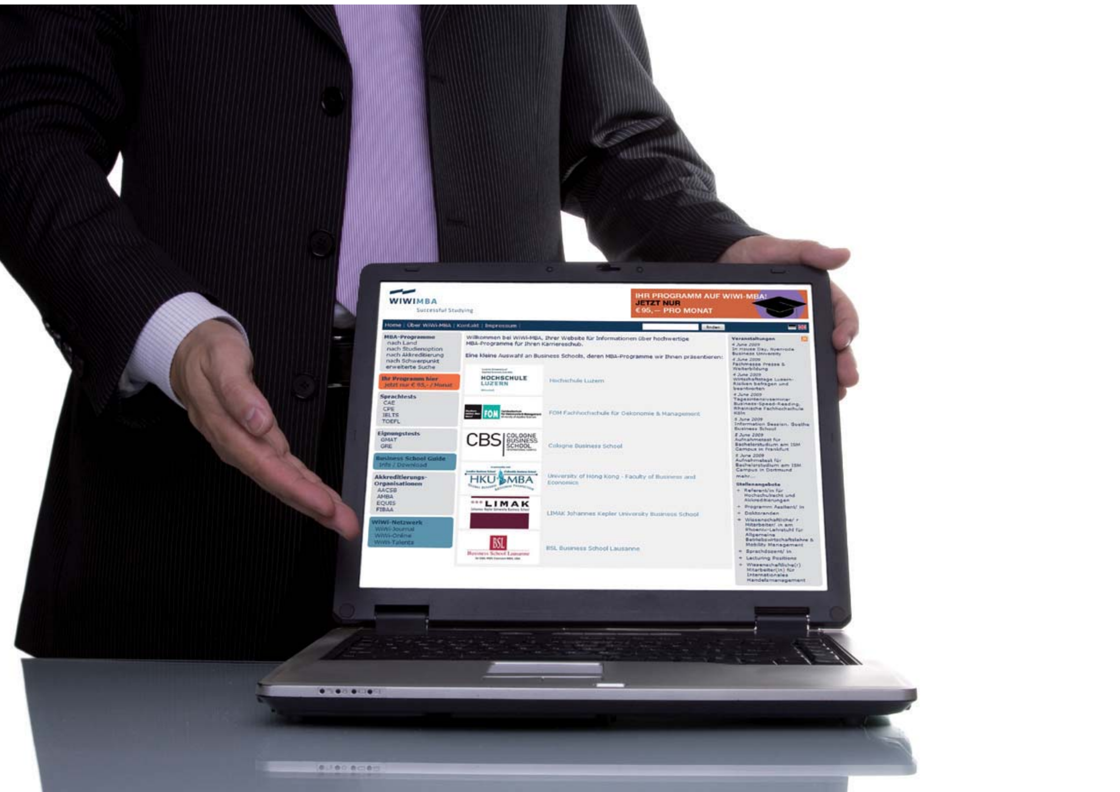


$F \sim F(m,n) \qquad m$ und $n \in \mathbb{N}$

$W_F =]0; \infty[$

$E(Z) = \dfrac{n}{n-2} \quad (n \geq 3) \qquad V(Z) = \dfrac{2n^2(n+m-2)}{m(n-4)(n-2)^2} \quad (n \geq 5)$

Für die Quantile gilt: $F_\alpha(m,n) = \dfrac{1}{F_{1-\alpha}(n,m)}$


WIWIMBA
Successful Studying
IHR PROGRAMM AUF WIWI-MBA!
JETZT NUR
€95,– PRO MONAT

5.3 Zentraler Grenzwertsatz

Bedingungen:

Die Zufallsvariablen $X_1, ..., X_n$ sind unabhängig und identisch verteilt mit Erwartungswert μ und Varianz $\sigma^2 > 0$.

$$Z = \frac{S_n - n \cdot \mu}{\sigma\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} X_i - n \cdot \mu}{\sigma\sqrt{n}} \quad \overset{n\to\infty}{\longrightarrow} \quad Z \sim N(0;1)$$

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \overset{n\to\infty}{\longrightarrow} \quad Z \sim N(0;1) \quad \text{mit } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Es gilt also jeweils: $\lim\limits_{n\to\infty} P(Z \leq z) = \Phi(z)$

Approximation durch die Normalverteilung:

$$S_n = \sum_{i=1}^{n} X_i \quad \overset{app.}{\sim} \quad N(n\mu; n\sigma^2) \quad \text{mit } n \geq 50$$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \overset{app.}{\sim} \quad N(\mu; \frac{\sigma^2}{n}) \quad \text{mit } n \geq 50$$

Beispiel mit einer Bernoulli-verteilten Zufallsvariablen:

$X_1, ..., X_n$ sind $B(1; p)$-verteilt und unabhängig. Bei $n \geq 50$ gilt:

$$Z = \frac{\sum\limits_{i=1}^{n} X_i - n \cdot p}{\sqrt{p(1-p)}\sqrt{n}} \quad \overset{app.}{\sim} \quad N(0;1)$$

$$Z = \frac{\overline{X} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} \quad \overset{app.}{\sim} \quad N(0;1) \quad \text{mit } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S_n = \sum_{i=1}^{n} X_i \quad \overset{app.}{\sim} \quad N(\mu = np; \sigma^2 = np(1-p))$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \overset{app.}{\sim} \quad N(\mu = p; \sigma^2 = \frac{p(1-p)}{n})$$

5.4 Approximation von Verteilungen

Stetigkeitskorrektur:

Diese wird bei der Approximation diskreter Verteilungen durch die Normalverteilung verwendet.

$P(X \leq x_i) = F(x_i) \approx F_{Normalverteilt}(x_i + 0,5)$

$P(X < x_i) = F(x_i - 1) \approx F_{Normalverteilt}(x_i - 0,5)$

Beispiel mit einer binomial-verteilten Zufallsvariablen:

$X \sim B(n;p) \quad$ mit $np(1-p) \geq 9$

$$P(X \leq x) = F(x) \approx \Phi\left(\frac{x + 0,5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X = x) \approx \Phi\left(\frac{x + 0,5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - 0,5 - np}{\sqrt{np(1-p)}}\right)$$

Approximation der Binomialverteilung:

durch die Poisson-Verteilung:

$B(n;p) \approx P(\lambda = np) \qquad$ für $n \geq 30$ und $p \leq 0,05$

durch die Normalverteilung:

$B(n;p) \approx N(\mu = np\ ;\ \sigma^2 = np(1-p)) \qquad$ für $np(1-p) \geq 9$

Approximation der Hypergeometrischen Verteilung:

durch die Binomialverteilung:

$H(N;M;n) \approx B(n\ ;\ p = \frac{M}{N}) \qquad$ für $\frac{n}{N} \leq 0,05$

durch die Poisson-Verteilung:

$H(N;M;n) \approx P(\lambda = n \cdot \frac{M}{N}) \qquad$ für $\frac{n}{N} \leq 0,05$

durch die Normalverteilung:

$$H(N;M;n) \approx N\left(\mu = n\frac{M}{N}\ ;\ \sigma^2 = n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}\right)$$

für $n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right) \geq 9$ und $\frac{n}{N} \leq 0{,}05$

Approximation der Poisson-Verteilung:

durch die Normalverteilung:

$P(\lambda) \approx N(\mu = \lambda\ ;\ \sigma^2 = \lambda)$ für $\lambda \geq 9$

Approximation der Chi-Quadrat-Verteilung:

durch die Normalverteilung:

$\chi^2(n) \approx N(\mu = n\ ;\ \sigma^2 = 2n)$ für $n \geq 30$

durch die Standardnormalverteilung:

Transformation: $z = \sqrt{2\chi^2(n)} - \sqrt{2n-1}$

$z \approx N(\mu = 0\ ;\ \sigma^2 = 1)$ für $n \geq 30$

Approximation der t-Verteilung:

durch die Standardnormalverteilung:

$t(n) \approx N(\mu = 0\ ;\ \sigma^2 = 1)$ für $n \geq 30$

6 Parameterschätzung

Symbole:

n:	Länge der Stichprobe
$\overline{X}$:	Schätzfunktion für den Erwartungswert μ
S^2:	Schätzfunktion für die Varianz σ^2 bei unbekanntem μ
$\tilde{S}^2$:	Schätzfunktion für die Varianz σ^2 bei bekanntem μ
P:	Schätzfunktion für den Anteilswert p
z_p:	p-Quantil der Standardnormalverteilung
$t_{n,p}$:	p-Quantil der t-Verteilung mit n Freiheitsgraden
$\chi^2_{n,p}$:	p-Quantil der Chi-Quadrat-Verteilung mit n Freiheitsgraden

6.1 Einfache Zufallsstichprobe

Eine einfache Zufallsstichprobe vom Umfang n aus X mit $f(x)$ liegt vor, wenn die Stichprobenvariablen $X_1, ..., X_n$

→ unabhängig und
→ identisch verteilt sind

Folglich beträgt die Wahrscheinlichkeit für die Realisation von n Stichprobenwerten $x_1, ..., x_n$: $f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$

6.2 Punktschätzung

6.2.1 Schätzfunktion

Defintion:

Eine Stichprobenfunktion, die zu jeder Realisierung $x_1, ..., x_n$ der Zufallsvariablen $X_1, ..., X_n$ einen Schätzwert $\hat{\theta}$ für den unbekannten Parameter θ liefert, nennt man Schätzfunktion. Man schreibt:

$\hat{\theta} = g(X_1, ..., X_n)$

Eine erwartungstreue (unverzerrte) Schätzfunktion liegt vor,
wenn gilt: $E(\hat{\theta}) = \theta$

Schätzfunktion für den Erwartungswert:

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

Es gilt: $E(\overline{X}) = \mu$ und $V(\overline{X}) = \frac{\sigma^2}{n}$

Schätzfunktion für den Anteilswert:

$$P = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{mit } X_i \sim B(1;p)$$

Es gilt: $E(P) = p$ und $V(P) = \frac{p(1-p)}{n}$

Schätzfunktion für die Varianz:

Erwartungswert unbekannt:

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

Es gilt: $E(D^2) = \frac{n-1}{n}\sigma^2$

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

Es gilt: $E(S^2) = \sigma^2$

Erwartungswert bekannt:

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

Es gilt: $E(\tilde{S}^2) = \sigma^2$

für das mittlere Abweichungsprodukt:

$$D_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} X_i \cdot Y_i - \overline{X}\cdot\overline{Y}$$

Schätzfunktion für die Kovarianz:

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{n}{n-1}D_{XY}$$

Schätzfunktion für den Korrelationskoeffizienten:

$$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2 \sum\limits_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

6.2.2 Maximum-Likelihood-Schätzung

Es sei γ ein unbekannter Parameter (Vektor) in der Funktion $f(x_1, ..., x_n; \gamma)$. Der (die) Paramater dieser Funktion ist (sind) mit Hilfe der Stichprobenrealisationen $x_1, ..., x_n$ der Zufallsvariablen X zu schätzen, wobei die Verteilungsfunktion bekannt ist.

Likelihood-Funktion:

$$L(\gamma) = \prod_{i=1}^{n} f(x_i; \gamma) = f(x_1; \gamma) \cdot ... \cdot f(x_n; \gamma)$$

Maximum-Likelihood-Prinzip γ:

Bestimme $\hat{\gamma}$ so, dass gilt: $\quad L(\hat{\gamma}) \geq L(\gamma)$ für alle zulässigen γ

Bei der Maximum-Likelihood-Schätzung wird (werden) der (die) Parameter $\hat{\gamma}$ so ermittelt, dass die Stichprobenrealisationen wahrscheinlicher sind als bei allen anderen möglichen Werten für γ.

Beispiele für Maximum-Likelihood-Schätzfunktionen:

$X \sim B(1, p) \quad \rightarrow \quad \hat{p} = \overline{X}$

$X \sim B(n, p) \quad \rightarrow \quad \hat{p} = \frac{1}{n}\overline{X}$

$X \sim P(\lambda) \quad \rightarrow \quad \hat{\lambda} = \overline{X}$

$X \sim E(\lambda) \quad \rightarrow \quad \hat{\lambda} = \frac{1}{\overline{X}}$

$X \sim N(\mu, \sigma^2) \quad \rightarrow \quad \hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = D^2$

6.3 Konfidenzintervalle

Ein Konfidenzintervall mit dem Konfidnzniveau $1-\alpha$ enthält mit Wahrscheinlichkeit $1-\alpha$ den unbekannten Parameter.

n_{min} ist der Mindesstichprobenumfang für ein Konfidenzintervall mit Niveau $1-\alpha$, das höchstens die Breite l hat.

6.3.1 Konfidenzintervall für den Erwartungswert

X ist normalverteilt, σ ist bekannt, n beliebig:

$$\overline{X} \pm z_{1-\alpha/2}\ \frac{\sigma}{\sqrt{n}} \qquad \text{mit Konfidenzniveau } 1-\alpha$$

$$n_{min} = \left(\frac{2 \cdot z_{1-\alpha/2} \cdot \sigma}{l}\right)^2$$

X ist normalverteilt, σ ist unbekannt, n beliebig:

$$\overline{X} \pm t_{n-1,1-\alpha/2}\ \frac{S}{\sqrt{n}} \qquad \text{mit Konfidenzniveau } 1-\alpha$$

$$n_{min} = \left(\frac{2 \cdot t_{1-\alpha/2} \cdot \hat{\sigma}}{l}\right)^2 \qquad \text{z.B. mit } \hat{\sigma} = S$$

X ist normalverteilt, σ ist unbekannt n $\geq$ 100:

$$\overline{X} \pm z_{1-\alpha/2}\ \frac{S}{\sqrt{n}} \qquad \text{mit Konfidenzniveau } \approx 1-\alpha$$

X ist nicht normalverteilt, σ ist bekannt, n $\geq$ 30:

$$\overline{X} \pm z_{1-\alpha/2}\ \frac{\sigma}{\sqrt{n}} \qquad \text{mit Konfidenzniveau } \approx 1-\alpha$$

$$n_{min} = \max\left\{30, \left(\frac{2 \cdot z_{1-\alpha/2} \cdot \sigma}{l}\right)^2\right\}$$

X ist nicht normalverteilt, $\boldsymbol{\sigma}$ ist unbekannt, n $\geq$ 30:

$$\overline{X} \pm t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}} \qquad \text{mit Konfidenzniveau} \approx 1-\alpha$$

$$n_{min} = \max\left\{30, \left(\frac{2 \cdot t_{1-\alpha/2} \cdot \hat{\sigma}}{l}\right)^2\right\} \qquad \text{z.B. mit } \hat{\sigma} = S$$

6.3.2 Konfidenzintervall für die Varianz

X ist normalverteilt, $\boldsymbol{\mu}$ ist unbekannt, n beliebig:

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \quad ; \quad \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right] \qquad \text{mit Konfidenzniveau } 1-\alpha$$

X ist normalverteilt, $\boldsymbol{\mu}$ ist bekannt, n beliebig:

$$\left[\frac{n\widetilde{S}^2}{\chi^2_{n,1-\alpha/2}} \quad ; \quad \frac{n\widetilde{S}^2}{\chi^2_{n,\alpha/2}}\right] \qquad \text{mit Konfidenzniveau } 1-\alpha$$

6.3.3 Konfidenzintervall für den Anteilswert

Voraussetzung: $n \cdot p^*(1-p^*) \geq 9$

Für p^* gilt: p^* ist der Schätzwert für p und unabhängig von der aktuellen Stichprobe

$$P \pm z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} \qquad \text{mit Konfidenzniveau} \approx 1-\alpha$$

mit $P = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ wobei $X \sim B(1;p)$

$$n_{min} = \max\left\{\frac{9}{p^*(1-p^*)}, \frac{4 \cdot z^2_{1-\alpha/2} \cdot p^*(1-p^*)}{l^2}\right\}$$

6.3.4 Konfidenzintervall für eine Anzahl

Die Vorgehensweise ist identisch wie in Kapitel 6.3.3. Das Konfidenzintervall für die Anzahl $N \cdot p$ wird bestimmt, indem zusätzlich die Grenzen mit N multipliziert werden.

Konfidenzintervall mit Konfidenzniveau $\approx 1 - \alpha$:

$$\left[N \cdot \left(P - z_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}} \right) \quad ; \quad N \cdot \left(P + z_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}} \right) \right]$$

$$n_{min} = \max \left\{ \frac{9}{p^*(1-p^*)}, \frac{4N^2 \cdot z^2_{1-\alpha/2} \cdot p^*(1-p^*)}{l^2} \right\}$$

7 Statistische Hypothesentests

Symbole:

H_0:	Nullhypothese
H_1:	Alternativhypothese (auch: Gegenhypothese)
T:	Testfunktion (auch: Teststatistik, Prüfgröße)
K_α:	kritischer Bereich (auch: Ablehnungsbereich)
α:	Signifikanzniveau
n:	Stichprobenlänge
n_X:	Stichprobenlänge für X bei unabhängigen Stichproben
$\overline{X}$:	Schätzfunktion für den Erwartungswert μ
P:	Schätzfunktion für den Anteilswert p
$B(n, p)$:	binomialverteilt mit den Paramtern n und p
z_p:	p-Quantil der Standardnormalverteilung
$t_{n,p}$:	p-Quantil der t-Verteilung mit n Freiheitsgraden
$\chi^2_{n,p}$:	p-Quantil der Chi-Quadrat-Verteilung mit n Freiheitsgraden
$F_{n_X, n_Y, p}$:	p-Quantil der F-Verteilung mit n_X und n_Y Freiheitsgraden

7.1 Einführung

Vorgehen beim Hypothesentest:

1. Formulierung der Nullhypothese H_0 und der logisch entgegengesetzten Alternativhypothese H_1
2. Festlegung des Signifikanzniveaus α
3. Auswahl der geeigneten Testfunktion $T(X_1, ..., X_n)$
4. Auftstellen einer Entscheidungsregel, indem der kritische Bereich K_α bestimmt wird. Dabei gilt: $P(T \in K_\alpha | H_0) = \alpha$
5. Berechnung der Testfunktion T mit den Daten der Stichprobe
6. Ablehnung von H_0, wenn der Wert von T_n im kritischen Bereich liegt, also $T \in K_\alpha$. Ansonsten wird H_1 angenommen.

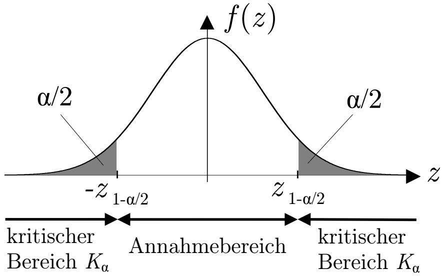
Entscheidungstabelle:

	H_0 ist wirklich wahr	H_1 ist wirklich wahr
H_0 wird angenommen	richtige Entscheidung	Fehler 2. Art Wahrscheinlichkeit: β
H_0 wird abgelehnt $\Rightarrow$ Annahme von H_1	Fehler 1. Art Wahrscheinlichkeit: α	richtige Entscheidung

P(H_0 wird abgelehnt|H_0 ist wahr)=α

P(H_0 wird angenommen|H_1 ist wahr)=β

Einseitige und zweiseitige parametrische Tests:



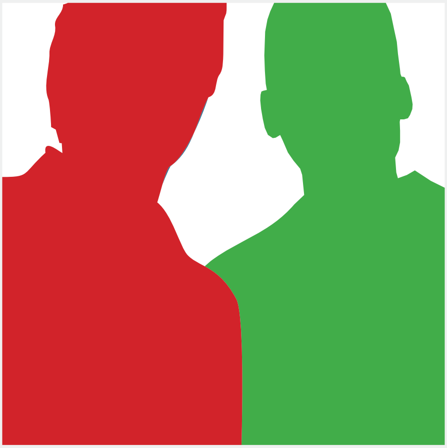
zweiseitiger Test:

Nullhypothese: $H_0 : \theta = \theta_0$

Alternativhypothese: $H_1 : \theta \neq \theta_0$

kritischer Bereich bei standardnormalverteilter Testfunktion:

$K_\alpha = \left\{ t \in \mathbb{R} \mid \quad |t| > z_{1-\alpha/2} \right\}$



Die WiWi-Media AG ist marktführender wirtschaftswissenschaftlicher Informationsdienstleister in Deutschland, Österreich und der Schweiz. Wir suchen als studienbegleitende Praxistätigkeit zum nächstmöglichen Zeitpunkt mehrere

Werkstudenten (m/w)

In enger Zusammenarbeit mit Praxis- und Forschungspartnern bearbeiten Sie herausfordernde Fragenstellungen im Themengebiet Mathematik, Statistik, Wirtschaftsstatistik und Ökonometrie. Wochenarbeitszeit nach Absprache (max. 20 Stunden pro Woche).

Wir bieten Ihnen:

- eigenverantwortliche Projektarbeit in der Weiterbildung, Forschung und Praxis
- ein praxisorientiertes und internationales Team
- eine offene und kooperative Atmosphäre
- Lösung von Problemstellungen auf höchstem wissenschaftlichem Niveau

Wir erwarten von Ihnen:

- Studierende/r der Wirtschaftswissenschaft oder Mathematik
- sehr gute Noten in Mathematik und Statistik (mind. 1,9)
- fundierte Kenntnisse mit Microsoft Formel-Editor oder in LaTeX
- Organisationstalent
- Ausgeprägte analytische und konzeptionelle Fähigkeiten

Senden Sie bitte Ihre aussagekräftigen Bewerbungsunterlagen (gerne per E-Mail) an folgende Adresse:

WiWi-Media AG
Neuer Wall 19
D-20354 Hamburg

schroeder@wiwi-media.ag
www.wiwi-media.ag

Wir bringen Sie nach vorn.
Tun Sie etwas für sich und Ihre Karriere.

WIWIMEDIA

Successful Branding

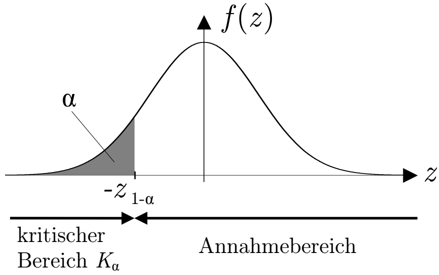
<ins>linksseitiger Test:</ins>

Nullhypothese: $H_0 : \theta \geq \theta_0$

Alternativhypothese: $H_1 : \theta < \theta_0$

kritischer Bereich bei standardnormalverteilter Testfunktion:

$K_\alpha = \{t \in \mathbb{R} \mid \quad t < -z_{1-\alpha}\}$



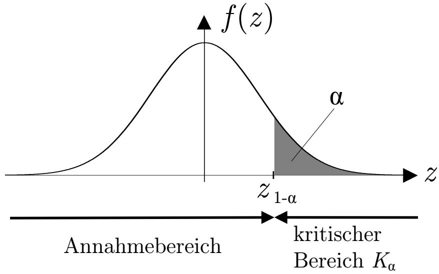
<ins>rechtsseitiger Test:</ins>

Nullhypothese: $H_0 : \theta \leq \theta_0$

Alternativhypothese: $H_1 : \theta > \theta_0$

kritischer Bereich bei standardnormalverteilter Testfunktion:

$K_\alpha = \{t \in \mathbb{R} \mid \quad t > z_{1-\alpha}\}$



7.2 Tests für den Ein-Stichprobenfall

7.2.1 Test auf den Erwartungswert

Hypothesen:

a) $H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

b) $H_0 : \mu \leq \mu_0$
$H_1 : \mu > \mu_0$

c) $H_0 : \mu \geq \mu_0$
$H_1 : \mu < \mu_0$

Annahmen:

i) X ist normalverteilt, σ ist bekannt, n ist beliebig
ii) X ist nicht normalverteilt, σ ist bekannt, $n \geq 30$
iii) X ist normalverteilt, σ ist unbekannt, n ist beliebig
iv) X ist nicht normalverteilt, σ ist unbekannt, $n \geq 30$

Testfunktion:

i) $T = \dfrac{\overline{X}_n - \mu_0}{\sigma} \cdot \sqrt{n} \quad \sim \quad N(0;1) \qquad$ (Gauß-Test)

ii) $T = \dfrac{\overline{X}_n - \mu_0}{\sigma} \cdot \sqrt{n} \quad \overset{app.}{\sim} \quad N(0;1)$ (Gauß-Test)

iii) $T = \dfrac{\overline{X}_n - \mu_0}{S} \cdot \sqrt{n} \quad \sim \quad t(n-1)$ (t-Test)

iv) $T = \dfrac{\overline{X}_n - \mu_0}{S} \cdot \sqrt{n} \quad \overset{app.}{\sim} \quad N(0;1)$

Entscheidung: Ablehnung von H_0, wenn

i,ii) a) $|T| > z_{1-\alpha/2}$ b) $T > z_{1-\alpha}$ c) $T < -z_{1-\alpha}$

iii,iv) a) $|T| > t_{n-1,1-\alpha/2}$ b) $T > t_{n-1,1-\alpha}$ c) $T < -t_{n-1,1-\alpha}$

7.2.2 Test auf den Anteilswert (Binomialtest)

Hypothesen:

a) $H_0 : p = p_0$ b) $H_0 : p \leq p_0$ c) $H_0 : p \geq p_0$
$H_1 : p \neq p_0$ $H_1 : p > p_0$ $H_1 : p < p_0$

Annahmen:

Für beide Fälle gilt: $X_i \sim B(1,p)$ $X = \sum_{i=1}^{n} X_i$ also $X \sim B(n;p)$

i) $n \geq 30$, $np_0 \geq 10$, $n(1-p_0) \geq 10$
ii) $n < 30$ oder $np_0 < 10$ oder $n(1-p_0) < 10$

Testfunktion:

i) $T = \dfrac{P - p_0}{\sqrt{p_0(1-p_0)}} \cdot \sqrt{n} \quad \overset{app.}{\sim} \quad N(0;1)$ (approx. Binomialtest)

ii) $T = X \quad \sim \quad B(n;p_0)$ (exakter Binomialtest)

Entscheidung: Ablehnung von H_0, wenn

i) a) $|T| > z_{1-\alpha/2}$ b) $T > z_{1-\alpha}$ c) $T < -z_{1-\alpha}$

ii) a) $T < c_{unten}$ oder $T > c_{oben}$ b) $T > c_{oben}$ c) $T < c_{unten}$

Bemerkung:

c_{unten} ist der erste Wert für x, bei der die Verteilungsfunktion $P(X \leq x)$ für $X \sim B(n;p_0)$ den Wert α (beim zweiseitigen Test: $\alpha/2$) überschreitet.

c_{oben} ist der erste Wert für x, bei der die Verteilungsfunktion den Wert $1 - \alpha$ (beim zweiseitigen Test: $1 - \alpha/2$) erreicht oder überschreitet.

7.2.3 Test auf die Varianz

Hypothesen:

a) $H_0 : \sigma^2 = \sigma_0^2$	b) $H_0 : \sigma^2 \leq \sigma_0^2$	c) $H_0 : \sigma^2 \geq \sigma_0^2$
$H_1 : \sigma^2 \neq \sigma_0^2$	$H_1 : \sigma^2 > \sigma_0^2$	$H_1 : \sigma^2 < \sigma_0^2$

Annahmen:

i) X ist normalverteilt, μ ist unbekannt, n beliebig
ii) X ist normalverteilt, μ ist bekannt, n beliebig

Testfunktion:

i) $$T = \frac{(n-1)S^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \sim \quad \chi^2(n-1)$$

ii) $$T = \frac{n\widetilde{S}^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2 \quad \sim \quad \chi^2(n)$$

Entscheidung: Ablehnung von H_0, wenn

i)	a) $T < \chi^2_{n-1,\alpha/2}$ oder $T > \chi^2_{n-1,1-\alpha/2}$	b) $T > \chi^2_{n-1,1-\alpha}$	c) $T < \chi^2_{n-1,\alpha}$
ii)	a) $T < \chi^2_{n,\alpha/2}$ oder $T > \chi^2_{n,1-\alpha/2}$	b) $T > \chi^2_{n,1-\alpha}$	c) $T < \chi^2_{n,\alpha}$

7.2.4 Test auf ein/en Prozentpunkt/Quantil

Definitionen:

x_p: $(p \cdot 100)$-Prozentpunkt/$p-$Quantil von X
A: Anzahl der Fälle, bei denen gilt: $x_i \leq \delta_0$

Hypothesen:

a) $H_0 : x_p = \delta_0$	b) $H_0 : x_p \geq \delta_0$	c) $H_0 : x_p \leq \delta_0$
$H_1 : x_p \neq \delta_0$	$H_1 : x_p < \delta_0$	$H_1 : x_p > \delta_0$

Annahmen:

Für Fall i) und ii) gilt jeweils: X besitzt stetige Verteilungsfunktion
i) $n \geq 30$ und $n \cdot p \geq 10$ und $n \cdot (1-p) \geq 10$
ii) $n < 30$ oder $n \cdot p < 10$ oder $n \cdot (1-p) < 10$

Testfunktion:

i) $T = \dfrac{\dfrac{A+0,5}{n} - p}{\sqrt{p(1-p)}} \cdot \sqrt{n} \quad \overset{app.}{\sim} \quad N(0;1)$

ii) $T = A \quad \sim \quad B(n;p)$

Entscheidung: Ablehnung von H_0, wenn

	a)	b)	c)
i)	a) $\|T\| > z_{1-\alpha/2}$	b) $T < -z_{1-\alpha}$	c) $T > z_{1-\alpha}$
ii)	a) $T < c_{unten}$ oder $T > c_{oben}$	b) $T < c_{unten}$	c) $T > c_{oben}$

Bemerkung:

Der kritischen Werte werden für Fall ii) wie beim Binomialtest beschrieben (Kapitel 7.2.2) bestimmt ($A \sim B(n;p)$).

7.2.5 Median- oder Vorzeichentest

Hypothesen:

a) $H_0 : x_{med} = \delta_0$
$H_1 : x_{med} \neq \delta_0$

b) $H_0 : x_{med} \geq \delta_0$
$H_1 : x_{med} < \delta_0$

c) $H_0 : x_{med} \leq \delta_0$
$H_1 : x_{med} > \delta_0$

Bemerkung:

Der Median- oder Vorzeichentest ist ein Spezialfall des Tests auf einen Prozentpunkt (siehe Kapitel 7.2.4). p ist in diesem Fall gleich $0,5$.

7.2.6 Chi-Quadrat-Anpassungstest

Annahmen:

i) X ist diskret verteilt und kann die Werte x_i $(i = 1, ..., k)$ annehmen.
ii) X ist stetig verteilt und der Wertebereich von X ist in k disjunkte Klassen K_i $(i = 1, ..., k)$ eingeteilt.

Definitionen:

$F_0(X)$: hypothetische Verteilung von X
p_i: hypothetische Wahrscheinlichkeit für $X = x_i$ bzw. $X \in K_i$
$\tilde{n}_i$: erwartete Anzahl der Beobachtungen für $X = x_i$ bzw. $X \in K_i$ in der Stichprobe der Länge n (es gilt: $np_i = \tilde{n}_i$)
n_i: Anzahl der Beobachtungen für $X = x_i$ bzw. $X \in K_i$ in der Stichprobe der Länge n
m: Anzahl der durch die Stichprobe geschätzten Parameter $\tilde{n}_i$

Hypothesen:

$H_0 : F(X) = F_0(X) \qquad H_1 : F(X) \neq F_0(X)$

äquivalente Formulierung für Fall i) und ii):
i) $H_0 : P(X = i) = p_i \qquad H_1 : P(X = i) \neq p_i$ für mindestens ein i
ii) $H_0 : P(X \in K_i) = p_i \qquad H_1 : P(X \in K_i) \neq p_i$ für mindestens ein i

Testfunktion:

i,ii) $$T = \sum_{i=1}^{k} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \overset{app.}{\sim} \chi^2(k - m - 1)$$

Entscheidung: Ablehnung von H_0, wenn

i,ii) $T > \chi^2_{1-\alpha,k-m-1}$

Bemerkung:

Alle erwarteten Häufigkeiten müssen größer oder gleich 5 sein:
$\tilde{n}_i \geq 5 \Leftrightarrow np_i \geq 5$.
Bei diesem Test reduziert sich die Anzahl der Freiheitsgrade um die Anzahl der Schätzungen m. Sind beispielsweise die hypothetischen Wahrscheinlichkeiten p_i vollständig spezifiziert und somit keine Schätzung erforderlich, beträgt die Anzahl der Freiheitsgrade $k - 1$.

7.3 Tests bei unabhängigen Stichproben

Definition:
Bei unabhängigen Stichprobe sind $X_1, ..., X_{n_X}, Y_1, ..., Y_{n_Y}$ unabhängig voneinander.

7.3.1 Vergleich von zwei Erwartungswerten

Hypothesen:

a) $H_0 : \mu_Y - \mu_X = \delta_0$ $H_1 : \mu_Y - \mu_X \neq \delta_0$

b) $H_0 : \mu_X - \mu_Y \geq \delta_0$ $H_1 : \mu_X - \mu_Y < \delta_0$

c) $H_0 : \mu_X - \mu_Y \leq \delta_0$ $H_1 : \mu_X - \mu_Y > \delta_0$

Annahmen:

für alle Fälle gilt: die Stichproben sind unabhängig
i) X,Y normalverteilt; σ_X^2,σ_Y^2 bekannt; n_X,n_Y beliebig
ii) X,Y nicht normalverteilt; σ_X^2,σ_Y^2 bekannt; n_X,$n_Y \geq 30$
iii) X,Y normalverteilt; $\sigma_X^2 = \sigma_Y^2$ unbekannt; n_X,n_Y beliebig
iv) X,Y nicht normalverteilt; $\sigma_X^2 = \sigma_Y^2$ unbekannt; n_X,$n_Y \geq 30$
v) X,Y nicht normalverteilt; σ_X^2 und σ_Y^2 unbekannt; n_X,$n_Y \geq 100$

Testfunktion:

i) $$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0;1)$$

ii) $$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{app.}{\sim} N(0;1)$$

iii) $$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \sim t(n_X + n_Y - 2)$$

iv,v) $$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \overset{app.}{\sim} N(0;1)$$

mit $\quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n_X}(X_i - \overline{X})^2 \quad$ und $\quad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n_Y}(Y_i - \overline{Y})^2$

Entscheidung: Ablehnung von H_0, wenn

Für die Fälle i),ii),iv) und v) gelten:

a) $|T| > z_{1-\alpha/2}$ b) $T < -z_{1-\alpha}$ c) $T > z_{1-\alpha}$

Für Fall iii) gilt:
a) $|T| > t_{n_X+n_Y-2,1-\alpha/2}$
b) $T < -t_{n_X+n_Y-2,1-\alpha}$
c) $T > t_{n_X+n_Y-2,1-\alpha}$

7.3.2 Einfache Varianzanalyse

Annahmen:

Es gibt k unabhängige normalverteilte Zufallsvariablen $X_1, ..., X_k$ mit unbekannten, aber identischen Varianzen.

Definitionen:

X_{ij}: j-ter Stichprobenwert von X_i $\quad (i = 1, ..., k) \quad (j = 1, ..., n_i)$
n_i: Stichprobenumfang von X_i $\quad (i = 1, ..., n_i)$
n: Gesamtlänge der Stichprobe (es gilt: $n = n_1 + ... + n_k$)

Hypothesen:

$H_0 : \mu_1 = \mu_2 = ... = \mu_k \qquad H_1 : \mu_i \neq \mu_j$ für mindestens ein Paar (i, j)

Testfunktion:

$$T = \frac{\frac{n}{k-1}\; s^2_{extern}}{\frac{n}{n-k}\; s^2_{intern}} = \frac{\sum_{i=1}^{k} n_i(\overline{X}_i - \overline{X})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2/(n-k)} \quad \sim \quad F(k-1, n-k)$$

mit: $\quad \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \quad$ und $\quad \overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i\overline{X}_i$

Entscheidung: Ablehnung von H_0, wenn $T > F_{k-1,n-k,1-\alpha}$

7.3.3 Vergleich von zwei Anteilswerten

Hypothesen:

a) $H_0 : p_X - p_Y = 0$
$H_1 : p_X - p_Y \neq 0$

b) $H_0 : p_X - p_Y \geq 0$
$H_1 : p_X - p_Y < 0$

c) $H_0 : p_X - p_Y \leq 0$
$H_1 : p_X - p_Y > 0$

Annahmen:

unabhänige Stichproben, $X \sim B(1; p_X)$ und $Y \sim B(1; p_Y)$,
$n_X \cdot P_X(1 - P_X) \geq 9$ und $n_Y \cdot P_Y(1 - P_Y) \geq 9$

Testfunktion:

$$T = \frac{P_X - P_Y}{\sqrt{\hat{p}(1-\hat{p})\frac{n_X + n_Y}{n_X \cdot n_Y}}} \quad \overset{app.}{\sim} \quad N(0;1)$$

mit $\quad \hat{p} = \dfrac{n_X \cdot p_X + n_Y \cdot p_Y}{n_X + n_Y} \qquad P_X = \dfrac{1}{n_X}\sum_{i=1}^{n_X} X_i \qquad P_Y = \dfrac{1}{n_Y}\sum_{i=1}^{n_Y} Y_i$

Entscheidung: Ablehnung von H_0, wenn

a) $|T| > z_{1-\alpha/2}$ b) $T < -z_{1-\alpha}$ c) $T > z_{1-\alpha}$

7.3.4 Vergleich von zwei Varianzen

Hypothesen:

a) $H_0 : \sigma_X^2 = \sigma_Y^2$
$H_1 : \sigma_X^2 \neq \sigma_Y^2$

b) $H_0 : \sigma_X^2 \geq \sigma_Y^2$
$H_1 : \sigma_X^2 < \sigma_Y^2$

c) $H_0 : \sigma_X^2 \leq \sigma_Y^2$
$H_1 : \sigma_X^2 > \sigma_Y^2$

Annahmen:

Für Fall i) und ii) gilt: unabhänige Stichproben
i) X und Y sind normalverteilt; μ_X, μ_Y unbekannt; n_X, n_Y beliebig
ii) X und Y sind normalverteilt; μ_X, μ_Y bekannt; n_X, n_Y beliebig

Testfunktion:

$$\text{i)} \quad T = \frac{S_X^2}{S_Y^2} = \frac{\sum_{i=1}^{n_X}(X_i - \overline{X})/(n_X - 1)}{\sum_{j=1}^{n_Y}(Y_i - \overline{Y})/(n_Y - 1)} \quad \sim \quad F(n_X - 1, n_Y - 1)$$

ii) $$T = \frac{\tilde{S}_X^2}{\tilde{S}_Y^2} = \frac{\sum\limits_{i=1}^{n_X} (X_i - \mu_X)/(n_X)}{\sum\limits_{j=1}^{n_Y} (Y_i - \mu_Y)/(n_Y)} \quad \sim \quad F(n_X, n_Y)$$

Entscheidung: Ablehnung von H_0, wenn

Fall i): a) $T < F_{n_X-1,n_Y-1,\alpha/2}$ oder $T > F_{n_X-1,n_Y-1,1-\alpha/2}$
b) $T < F_{n_X-1,n_Y-1,\alpha}$ c) $T > F_{n_X-1,n_Y-1,1-\alpha}$

Fall ii): a) $T < F_{n_X,n_Y,\alpha/2}$ oder $T > F_{n_X,n_Y,1-\alpha/2}$
b) $T < F_{n_X,n_Y,\alpha}$ c) $T > F_{n_X,n_Y,1-\alpha}$

Bemerkung:

Dieser Test kann auch durchgeführt werden, wenn nur eine von den zwei Varianzen bekannt ist. Ist z.B. nur σ_X^2 bekannt, dann muss der Test in Fall i) geändert werden, indem S_X^2 durch $\tilde{S}_X^2$ ersetzt wird. Die Anzahl der Freiheitsgrade erhöht sich von $n_X - 1$ auf n_X.

7.4 Tests bei verbundenen Stichproben

Es liegt ein verbundener (gepaarter, abhängiger) Zweistichprobentest vor, wenn die Beobachtungen der einen Stichprobe von denen der anderen Stichprobe abhängig sind. Zu jedem x aus der einen Stichprobe gehört ein y aus der anderen Stichprobe. Aus einer Stichprobe der Länge n resultieren n Stichprobenpaare: $(x_i, y_i), ..., (x_n, y_n)$.

7.4.1 Test auf gleiche Erwartungswerte

Hypothesen:

a) $H_0 : \mu_X = \mu_Y$
$H_1 : \mu_X \neq \mu_Y$

b) $H_0 : \mu_X \leq \mu_Y$
$H_1 : \mu_X > \mu_Y$

c) $H_1 : \mu_X \geq \mu_Y$
$H_1 : \mu_X < \mu_Y$

Annahmen:

i) X, Y normalverteilt; σ_X^2, σ_Y^2 unbekannt; n beliebig
ii) X, Y nicht normalverteilt; σ_X^2, σ_Y^2 unbekannt; $n \geq 100$

Definition: $D_i = X_i - Y_i \qquad (i = 1, ..., n)$

Testfunktion:

i) $T = \dfrac{\overline{D} - 0}{S_D}\sqrt{n} \quad \sim \quad t(n-1)$

ii) $T = \dfrac{\overline{D} - 0}{S_D}\sqrt{n} \quad \overset{app.}{\sim} \quad N(0;1)$

jeweils mit: $S_D^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$

und $\overline{D} = \dfrac{1}{n}\sum_{i=1}^{n} D_i = \overline{X} - \overline{Y}$

Entscheidung: Ablehnung von H_0, wenn

i) a) $|T| > t_{n-1,1-\alpha/2}$ b) $T > t_{n-1,1-\alpha}$ c) $T < -t_{n-1,1-\alpha}$

ii) a) $|T| > z_{1-\alpha/2}$ b) $T > z_{1-\alpha}$ c) $T < -z_{1-\alpha}$

7.5 Zusammenhangsanalyse

7.5.1 Chi-Quadrat-Unabhängigkeitstest

Definitionen:

n: Anzahl der Stichprobenpaare aus (X, Y)
x_i: mögliche Ausprägungen von $X \quad (i = 1, ..., p)$
y_j: mögliche Ausprägungen von $Y \quad (j = 1, ..., q)$
n_{ij}: beobachtete absolute Häufigkeiten für das Wertepaar (x_i, y_j)
$\tilde{n}_{ij}$: erwartete absolute Häufigkeiten für das Wertepaar (x_i, y_j) bei Unabhängigkeit
$n_{i\bullet}$ bzw. $n_{\bullet j}$: absolute Randhäufigkeit für x_i bzw. y_j
$h_{i\bullet}$ bzw. $h_{\bullet j}$: relative Randhäufigkeit für x_i bzw. y_j
$\hat{h}_{i\bullet}$ bzw. $\hat{h}_{\bullet j}$: geschätzte relative Randhäufigkeit für x_i bzw. y_j

Hypothesen:

$H_0 : X$ und Y sind unabhängig $\qquad H_1 : X$ und Y sind abhängig

Annahmen:

Für die Fälle i und ii gelten jeweils: Es liegt eine Kontingenztabelle vor und es wird kein Verteilungsmodell unterstellt.

i) $h_{i\bullet}$ und $h_{\bullet j}$ sind bekannt
ii) $h_{i\bullet}$ und $n_{\bullet j}$ sind unbekannt und werden aus den vorhandenen Daten der Kontingenztabelle geschätzt

Testfunktion:

i) $$T = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}} \quad \overset{app.}{\sim} \quad \chi^2(pq-1)$$

mit: $\tilde{n}_{ij} = n \cdot h_{i\bullet} h_{\bullet j}$

ii) $$T = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}} \quad \overset{app.}{\sim} \quad \chi^2((p-1)(q-1))$$

mit: $\tilde{n}_{ij} = n \cdot \hat{h}_{i\bullet} \hat{h}_{\bullet j}$ $\qquad \hat{h}_{i\bullet} = n_{i\bullet}/n \qquad \hat{h}_{\bullet j} = n_{\bullet j}/n$

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij} \qquad n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$$

Entscheidung: Ablehnung von H_0, wenn

i) $T > \chi^2_{pq-1,1-\alpha}$
ii) $T > \chi^2_{(p-1)(q-1),1-\alpha}$

Anmerkungen:

Voraussetzung zur Durchführbarkeit dieses Tests ist, dass alle erwarteten absoluten Häufigkeiten größer oder gleich 5 sind: $\tilde{n}_{ij} \geq 5$ für alle i und j.

7.5.2 Test auf den Korrelationskoeffizienten

Annahmen:

i) X und Y sind normalverteilt, n beliebig
ii) X und Y sind normalverteilt, $n \geq 30$

Hypothesen:

a) $H_0 : \rho_{XY} = \rho_0$
$H_1 : \rho_{XY} \neq \rho_0$

b) $H_0 : \rho_{XY} \geq \rho_0$
$H_1 : \rho_{XY} < \rho_0$

c) $H_0 : \rho_{XY} \leq \rho_0$
$H_1 : \rho_{XY} > \rho_0$

Testfunktion:

i,ii) $$T = \frac{1}{2}\left(\ln\left(\frac{1+r_{XY}}{1-r_{XY}}\right) - \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right)\sqrt{n-3}$$

$$\text{mit } r_{XY} = \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2\sum\limits_{i=1}^{n}(Y_i-\overline{Y})^2}}$$

Es gilt für: Fall i) $T \sim t(n-2)$ Fall ii) $T \overset{app.}{\sim} N(0;1)$

Entscheidung: Ablehnung von H_0, wenn

i) a) $|T| > t_{n-2,1-\alpha/2}$ b) $T < -t_{n-2,1-\alpha}$ c) $T > t_{n-2,1-\alpha}$
ii) a) $|T| > z_{1-\alpha/2}$ b) $T < -z_{1-\alpha}$ c) $T > z_{1-\alpha}$

Anmerkungen:

Wenn die Nullhypothese $\rho_{XY} = 0$ gewählt wird, dann ist dies ein Test auf lineare Unabhängigkeit zwischen X und Y.

7.6 Tests bei der linearen Einfachregression

7.6.1 Einführung

Modell der linearen Einfachregression (Stichprobenmodell):

$Y_i = \alpha + \beta x_i + U_i \qquad (i = 1, ..., n)$

Es gilt:
→ die Störvariablen $U_1, ..., U_n$ sind unabhängig und identisch verteilt
→ $E(U_i) = 0$ und $V(U_i) = \sigma_U^2$
→ α, β und σ_U^2 sind unbekannte Parameter und werden aus den Daten (x_i, y_i) geschätzt

Erwartungswert und Varianz der $\mathbf{Y_i}$:

$E(Y_i) = \alpha + \beta x_i \qquad V(Y_i) = \sigma_U^2$

Kleinste-Quadrate-Schätzer:

Steigungskoeffizient: $\hat{\beta} = r_{XY} \cdot \frac{S_Y}{S_X} = \frac{S_{XY}}{S_X^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=n}^{n}(x_i - \overline{x})^2}$

Konstante: $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x}$

Varianz der Residuen (Störgrößenvarianz):

$$\hat{\sigma}_U^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{U}_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \frac{n-1}{n-2}(S_Y^2 - \hat{\beta}^2 S_X^2)$$
$$= \frac{n-1}{n-2}\left(S_Y^2 - \frac{S_{XY}^2}{S_X^2}\right)$$

mit $\hat{U}_i = Y_i - \hat{Y}_i$ (Residuen) und

$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i$ (angepasste Werte)

Es gilt: $E(\hat{\beta}) = \beta$ $E(\hat{\alpha}) = \alpha$ $E(\hat{\sigma}_U^2) = \sigma_U^2$

erwartungstreuer Schätzer für die Varianz von $\hat{\beta}$:

$$\hat{V}(\hat{\beta}) = \hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma_U}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\hat{\sigma}_U^2}{\sum_{i=1}^{n}x_i^2 - n\overline{x}^2} = \frac{\hat{\sigma}_U^2}{(n-1)S_X^2} = \frac{\hat{\sigma}_U^2}{nD_X^2}$$

erwartungstreuer Schätzer für die Varianz von $\hat{\alpha}$:

$$\hat{V}(\hat{\alpha}) = \hat{\sigma}_{\hat{\alpha}}^2 = \hat{\sigma}_U^2\frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2} = \hat{\sigma}_U^2\frac{\sum_{i=1}^{n}x_i^2}{n(\sum_{i=1}^{n}x_i^2 - n\overline{x}^2)} = \hat{\sigma}_U^2\frac{\sum_{i=1}^{n}x_i^2}{n^2D_X^2}$$

Bestimmtheitsmaß: $B_{XY} = r_{XY}^2 = \frac{\left(\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})\right)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}$

7.6.2 Test auf die Konstante α

Annahmen:

$U_i \sim N(0;1)$ gilt für alle i; n beliebig

Hypothesen:

a) $H_0 : \alpha = \alpha_0$
$H_1 : \alpha \neq \alpha_0$

b) $H_0 : \alpha \geq \alpha_0$
$H_1 : \alpha < \alpha_0$

c) $H_0 : \alpha \leq \alpha_0$
$H_1 : \alpha > \alpha_0$

Testfunktion:

$$T = \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_{\hat{\alpha}}} \quad \sim \quad t(n-2)$$

Entscheidung: Ablehnung von H_0, wenn

a) $|T| > t_{n-2,1-\alpha/2}$ b) $T < -t_{n-2,1-\alpha}$ c) $T > t_{n-2,1-\alpha}$

Anmerkung:

Für $n \geq 30$ ist die Testfunktion approximativ standardnormalverteilt.

7.6.3 Test auf den Steigungskoeffizienten β

Annahmen:

$U_i \sim N(0;1)$ gilt für alle i; n beliebig

Hypothesen:

a) $H_0 : \beta = \beta_0$
$H_1 : \beta \neq \beta_0$

b) $H_0 : \beta \geq \beta_0$
$H_1 : \beta < \beta_0$

c) $H_0 : \beta \leq \beta_0$
$H_1 : \beta > \beta_0$

Testfunktion:

$$T = \frac{\hat{\beta} - \beta_0}{\hat{\sigma}_{\hat{\beta}}} \quad \sim \quad t(n-2)$$

Entscheidung: Ablehnung von H_0, wenn

a) $|T| > t_{n-2,1-\alpha/2}$ b) $T < -t_{n-2,1-\alpha}$ c) $T > t_{n-2,1-\alpha}$

Anmerkung:

Für $n \geq 30$ ist die Testfunktion approximativ standardnormalverteilt.

7.6.4 Test auf die Varianz

Annahmen:

$U_i \sim N(0; 1)$ gilt für alle i; n beliebig

Hypothesen:

a) $H_0 : \sigma_U^2 = \sigma_0^2$
$H_1 : \sigma_U^2 \neq \sigma_0^2$

b) $H_0 : \sigma_U^2 \geq \sigma_0^2$
$H_1 : \sigma_U^2 < \sigma_0^2$

c) $H_0 : \sigma_U^2 \leq \sigma_0^2$
$H_1 : \sigma_U^2 > \sigma_0^2$

Testfunktion:

$$T = (n-2)\frac{\hat{\sigma}_U^2}{\sigma_0^2} \quad \sim \quad \chi^2(n-2)$$

Entscheidung: Ablehnung von H_0, wenn

a) $T < \chi^2_{n-2,\alpha/2}$ oder $T > \chi^2_{n-2,1-\alpha/2}$

b) $T < \chi^2_{n-2,\alpha}$

c) $T > \chi^2_{n-2,1-\alpha}$

Binomialverteilung – Verteilungsfunktion $\mathbf{P_{n;p}(X \leq x)}$

			p									
n	x		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	0	0,	9500	9000	8500	8000	7500	7000	6500	6000	5500	5000
2	0	0,	9025	8100	7225	6400	5625	4900	4225	3600	3025	2500
	1		9975	9900	9775	9600	9375	9100	8775	8400	7975	7500
3	0	0,	8574	7290	6141	5120	4219	3430	2746	2160	1664	1250
	1		9928	9720	9393	8960	8438	7840	7183	6480	5748	5000
	2		9999	9990	9966	9920	9844	9730	9571	9360	9089	8750
4	0	0,	8145	6561	5220	4096	3164	2401	1785	1296	0915	0625
	1		9860	9477	8905	8192	7383	6517	5630	4752	3910	3125
	2		9995	9963	9880	9728	9492	9163	8735	8208	7585	6875
	3			9999	9995	9984	9961	9919	9850	9744	9590	9375
5	0	0,	7738	5905	4437	3277	2373	1681	1160	0778	0503	0313
	1		9774	9185	8352	7373	6328	5282	4284	3370	2562	1875
	2		9988	9914	9734	9421	8965	8369	7648	6826	5931	5000
	3			9995	9978	9933	9844	9692	9460	9130	8688	8125
	4				9999	9997	9990	9976	9947	9898	9815	9688
6	0	0,	7351	5314	3771	2621	1780	1176	0754	0467	0277	0156
	1		9672	8857	7765	6554	5339	4202	3191	2333	1636	1094
	2		9978	9842	9527	9011	8306	7443	6471	5443	4415	3438
	3		9999	9987	9941	9830	9624	9295	8826	8208	7447	6563
	4			9999	9996	9984	9954	9891	9777	9590	9308	8906
	5					9999	9998	9993	9982	9959	9917	9844
7	0	0,	6983	4783	3206	2097	1335	0824	0490	0280	0152	0078
	1		9556	8503	7166	5767	4449	3294	2338	1586	1024	0625
	2		9962	9743	9262	8520	7564	6471	5323	4199	3164	2266
	3		9998	9973	9879	9667	9294	8740	8002	7102	6083	5000
	4			9998	9988	9953	9871	9712	9444	9037	8471	7734
	5				9999	9996	9987	9962	9910	9812	9643	9375
	6						9999	9998	9994	9984	9963	9922

nicht aufgeführte Werte sind gleich 1,0000

$P_{n;p}(X = x) = P_{n;p}(X \leq x) - P_{n;p}(X \leq x-1)$

Für $p \geq 0,5$ gilt:
$P_{n;p}(X \leq x) = 1 - P_{n;1-p}(X \leq n-x-1)$

Binomialverteilung – Verteilungsfunktion $\mathbf{P_{n;p}(X \leq x)}$

							p					
n	x		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
8	0	0,	6634	4305	2725	1678	1001	0576	0319	0168	0084	0039
	1		9428	8131	6572	5033	3671	2553	1691	1064	0632	0352
	2		9942	9619	8948	7969	6785	5518	4278	3154	2201	1445
	3		9996	9950	9786	9437	8862	8059	7064	5941	4770	3633
	4			9996	9971	9896	9727	9420	8939	8263	7396	6367
	5				9998	9988	9958	9887	9747	9502	9115	8555
	6					9999	9996	9987	9964	9915	9819	9648
	7							9999	9998	9993	9983	9961
9	0	0,	6302	3874	2316	1342	0751	0404	0207	0101	0046	0020
	1		9288	7748	5995	4362	3003	1960	1211	0705	0385	0195
	2		9916	9470	8591	7382	6007	4628	3373	2318	1495	0898
	3		9994	9917	9661	9144	8343	7297	6089	4826	3614	2539
	4			9991	9944	9804	9511	9012	8283	7334	6214	5000
	5			9999	9994	9969	9900	9747	9464	9006	8342	7461
	6					9997	9987	9957	9888	9750	9502	9102
	7						9999	9996	9986	9962	9909	9805
	8								9999	9997	9992	9980
10	0	0,	5987	3487	1969	1074	0563	0282	0135	0060	0025	0010
	1		9139	7361	5443	3758	2440	1493	0860	0464	0233	0107
	2		9885	9298	8202	6778	5256	3828	2616	1673	0996	0547
	3		9990	9872	9500	8791	7759	6496	5138	3823	2660	1719
	4		9999	9984	9901	9672	9219	8497	7515	6331	5044	3770
	5			9999	9986	9936	9803	9527	9051	8338	7384	6230
	6				9999	9991	9965	9894	9740	9452	8980	8281
	7					9999	9996	9984	9952	9877	9726	9453
	8							9999	9995	9983	9955	9893
	9									9999	9997	9990

nicht aufgeführte Werte sind gleich 1,0000

$$P_{n;p}(X = x) = P_{n;p}(X \leq x) - P_{n;p}(X \leq x - 1)$$

Für $p \geq 0,5$ gilt:

$$P_{n;p}(X \leq x) = 1 - P_{n;1-p}(X \leq n - x - 1)$$

Standardnormalverteilung – Verteilungsfunktion Φ(z)

z		.,.0	.,.1	.,.2	.,.3	.,.4	.,.5	.,.6	.,.7	.,.8	.,.9
0,0	0,	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1		5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2		5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3		6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4		6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5		6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6		7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7		7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8		7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9		8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	0,	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1		8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2		8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3		9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4		9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5		9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6		9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7		9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8		9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9		9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	0,	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1		9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2		9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3		9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4		9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5		9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6		9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7		9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8		9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9		9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	0,	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1		9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2		9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3		9995	9995	9995	9996	9996	9996	9996	9996	9996	9997
3,4		9997	9997	9997	9997	9997	9997	9997	9997	9997	9998

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

Für $z \geq 3,90$ gilt: $\Phi(z) = 1,0000$ (bei der Angabe von vier Dezimalstellen)

Standardnormalverteilung – Quantile z_p

p	z_p	p	z_p	p	z_p
0,0001	-3,7190	0,2750	-0,5978	0,7500	0,6745
0,0005	-3,2905	0,3000	-0,5244	0,7750	0,7554
0,0010	-3,0902	0,3250	-0,4538	0,8000	0,8416
0,0050	-2,5758	0,3500	-0,3853	0,8250	0,9346
0,0100	-2,3263	0,3750	-0,3186	0,8500	1,0364
0,0200	-2,0537	0,4000	-0,2533	0,8750	1,1503
0,0250	-1,9600	0,4250	-0,1891	0,9000	1,2816
0,0300	-1,8808	0,4500	-0,1257	0,9250	1,4395
0,0400	-1,7507	0,4750	-0,0627	0,9400	1,5548
0,0500	-1,6449	0,5000	0,0000	0,9500	1,6449
0,0600	-1,5548	0,5250	0,0627	0,9600	1,7507
0,0750	-1,4395	0,5500	0,1257	0,9700	1,8808
0,1000	-1,2816	0,5750	0,1891	0,9750	1,9600
0,1250	-1,1503	0,6000	0,2533	0,9800	2,0537
0,1500	-1,0364	0,6250	0,3186	0,9900	2,3263
0,1750	-0,9346	0,6500	0,3853	0,9950	2,5758
0,2000	-0,8416	0,6750	0,4538	0,9990	3,0902
0,2250	-0,7554	0,7000	0,5244	0,9995	3,2905
0,2500	-0,6745	0,7250	0,5978	0,9999	3,7190

t-Verteilung – Quantile t_p

	p								
n	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
1	1,376	1,963	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	1,061	1,386	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,978	1,250	1,638	2,353	3,182	4,541	5,841	10,21	12,92
4	0,941	1,190	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,920	1,156	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,889	1,108	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,849	1,047	1,299	1,676	2,009	2,403	2,678	3,261	3,496
100	0,845	1,042	1,290	1,660	1,984	2,364	2,626	3,174	3,390
∞	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,090	3,291

$t_{n,p} = -t_{n,1-p}$

Für $n \geq 30$ können die Quantile durch die entsprechenden Quantile der Standardnormalverteilung angenähert werden.

Chi-Quadrat-Verteilung – Quantile χ^2_p

	p									
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
40	20,71	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,69	66,77
50	27,99	29,71	32,36	34,76	37,69	63,17	67,50	71,42	76,15	79,49
100	67,33	70,06	74,22	77,93	82,36	118,5	124,3	129,6	135,8	140,2

Für $n \geq 30$ ist folgende Approximation möglich:

$\chi^2_{n,p} \approx \frac{1}{2}(z_p + \sqrt{2n-1})^2$ mit z_p: p-Quantil der Standardnormalverteilung

Fisher(F)-Verteilung – 95%-Quantile

	m																			
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12

Kopfzeile: Anzahl der Freiheitsgrade m der Zählervariablen einer $F(m,n)$-verteilten Zufallsvariablen

Erste Spalte: Anzahl der Freiheitsgrade n der Nennervariablen einer $F(m,n)$-verteilten Zufallsvariablen

Fisher(F)-Verteilung – 99%-Quantile

	m																			
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0	26,9	26,9	26,8	26,8	26,8	26,7	26,7
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,5	14,4	14,3	14,2	14,2	14,2	14,1	14,1	14,0	14,0
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94

Kopfzeile: Anzahl der Freiheitsgrade m der Zählervariablen einer $F(m, n)$-verteilten Zufallsvariablen

Erste Spalte: Anzahl der Freiheitsgrade n der Nennervariablen einer $F(m, n)$-verteilten Zufallsvariablen

Stichwortverzeichnis