

Formelsammlung

Vorlesung Statistik für die Sozialwissenschaften

version: v3.3

Letzte Korrektur: Schätzung β_0 in linearer Regression subsection 10.2

Team Professur für Methoden der Politikwissenschaft
T_EX: B. Philipp Kleer

23. Juni 2022

Hinweis: Da das *typesetting* der Formelsammlung mit englischem Sprachformat erfolgt, ist das Dezimalzeichen ein Punkt (.) und nicht ein Komma (wie im Deutschen üblich).

Inhaltsverzeichnis

1	Notation in Urliste/Häufigkeitstabelle	1
1.1	Laufindex für Fälle eines Merkmals X	1
1.2	Laufindex für einzelne Merkmalsausprägungen	1
1.3	Absolute Häufigkeit einer Merkmalsausprägung	1
1.4	Relative Häufigkeit	1
2	Lagemaße	2
2.1	Median	2
2.2	Arithmetischer Mittelwert	2
2.3	Quantile (im Besonderen Quartile)	2
3	Streuemaße	3
3.1	Spannweite/Variationsweite	3
3.2	Interquartilsabstand	3
3.3	Varianz	3
3.4	Standardabweichung	3
3.5	Variationskoeffizient	3
3.6	Kovarianz	4
4	Kreuztabellen	4
4.1	Erwartete Häufigkeit	4
5	Zusammenhangsmaße	4
5.1	Zusammenhangsmaße basierend auf χ^2	5
5.1.1	χ^2	5
5.1.2	ϕ -Koeffizient	5
5.1.3	Cramers V	5
5.1.4	Kontingenzkoeffizient C	5
5.2	PRE-Maße	6
5.2.1	λ	6
5.2.2	η^2	6
5.3	Korrelationsmaße	7
5.3.1	Spearman's ρ	7

5.4	Pearsons r	7
6	Standardisierung	7
7	Population und Stichprobe	8
7.1	Population	8
7.2	Stichprobe	8
8	Signifikanztest	9
8.1	Testverfahren	9
8.1.1	z-Test	9
8.1.2	t-Test	9
8.1.3	χ^2 -Test	10
8.2	χ^2 -Test auf Unabhängigkeit	10
8.3	Berechnung Konfidenzintervall	10
8.3.1	anhand z-Verteilung (bekanntes σ)	10
8.3.2	anhand t-Verteilung (unbekanntes σ)	11
9	Mittelwertvergleich zwischen zwei Gruppen	11
9.1	Zweistichproben t-Test für unabhängige Stichproben mit homogenen Varianzen . . .	11
9.2	Zweistichproben t-Test für unabhängige Stichproben mit heterogenen Varianz . . .	12
9.3	Zweistichproben t-Test für abhängige Stichproben	12
10	Regression	13
10.1	Bivariate lineare Regression	13
10.2	Modellschätzung bivariate lineare Regression	13
10.3	Multivariate lineare Regression	14
10.4	Standardisierte Regressionskoeffizienten	15
11	z-Tabelle	16
12	t-Tabelle	18
13	χ^2-Tabelle	19

1 Notation in Urliste/Häufigkeitstabelle

1.1 Laufindex für Fälle eines Merkmals X

Indexierung für empirische (beobachtete) Werte x eines Merkmals X mit i für Fall und n für Gesamtanzahl:

x_i mit $i \in 1, 2, \dots, n$

z.B. 6 befragte Personen zur Augenfarbe: $X = (\text{blau}, \text{grün}, \text{braun}, \text{blau}, \text{grün}, \text{braun}) = (x_1, x_2, x_3, x_4, x_5, x_6)$, wobei $x_4 = \text{blau}$ ist und $n = 6$

1.2 Laufindex für einzelne Merkmalsausprägungen

Merkmalsausprägungen werden zur besseren Identifikation zwischen empirischen Werten und Merkmalsausprägungen eines Merkmals X mit ξ (ξ_i) oder a angegeben. Indexierung mit k für Fall und K für Gesamtanzahl:

ξ_k mit $k \in 1, 2, \dots, K$

a_k mit $a \in 1, 2, \dots, K$

z.B. 3 verschiedene Augenfarben
(*blau, grün, braun*), wobei $\xi_1 = a_1 = \text{blau}$ und $K = 3$

1.3 Absolute Häufigkeit einer Merkmalsausprägung

Oftmals erheben wir Merkmale, bei denen Merkmalsausprägungen mehrfach vorhanden sind (mehrere Beobachtungseinheiten weisen denselben Merkmalswert auf). Wir erwarten dies z.B. in unserer Erhebung der Augenfarbe: Mehrere Personen haben dieselbe Augenfarbe. Absolute Häufigkeiten erfassen, wie oft eine Merkmalsausprägung in dem Merkmal X vorliegt (also wie viele Merkmalswerte der jeweiligen Merkmalsausprägung entsprechen):

$$h_{\xi_k} = h_{a_k}$$

Im Beispiel der 6 Befragten von oben, haben 2 Personen eine grüne Augenfarbe:

$$h_{\xi_2} = h_{a_2} = 2$$

1.4 Relative Häufigkeit

Es können auch relative Häufigkeiten angegeben werden, die nicht die absolute Häufigkeit, sondern den Anteil einer Merkmalsausprägung angeben. Relative Häufigkeit eines Merkmals k in Variable X :

$$p_{\xi_k} = \frac{h_{\xi_k}}{n}$$
$$p_{a_k} = \frac{h_{a_k}}{n}$$

Im Beispiel:

$$p_{\xi_{\text{braun}}} = p_{\xi_3} = \frac{2}{6} \approx 0.333$$

$$p_{a_{braun}} = p_{a_3} = \frac{2}{6} \approx 0.333$$

2 Lagemaße

2.1 Median

Der Median ist der Wert, der in einer geordneten Liste genau in der Mitte liegt, d.h. dass sich genauso viele Werte oberhalb wie unterhalb des Wertes befinden (jeweils 50% der Fälle). Um den Median berechnen zu können, muss mindestens ordinales Skalenniveau vorliegen. Bei der Berechnung des Medians muss berücksichtigt werden, ob die Gesamtanzahl der Fälle n *gerade* oder *ungerade* ist.

Formel für ein *ungerades* n :

$$\tilde{x} = x_{\frac{n+1}{2}}$$

Formel für ein *gerades* n :

$$\tilde{x} = \frac{1}{2} * (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

2.2 Arithmetischer Mittelwert

Der Mittelwert, auch arithmetisches Mittel, gibt den durchschnittlichen Wert einer/s Variable/Merkmal an. Dazu müssen die Abstände zwischen den Werte interpretierbar sein, also muss mindestens Intervallskalenniveau vorliegen. Formel zur Berechnung des arithmetischen Mittelwertes:

$$\bar{x} = \frac{1}{n} * \sum_{i=1}^n x_i = \frac{\sum_{i=1}^n x_i}{n}$$

2.3 Quantile (im Besonderen Quartile)

Neben dem Median können auch Quantile berechnet werden, die einen gewählten Anteil (p) der Verteilung einer Variable angeben. Wie beim Median müssen die Werte erst geordnet werden, bevor Quantile berechnet werden können. Mit Quantilen wird der Anteil einer Verteilung berechnet, wobei p das gesuchte Quantil ist. Quantile sind daher Grenzen, die festlegen, wie viele Werte über oder unter einem gewissen Wert (p) liegen. p -Quantile können beliebig gewählt werden, gängig sind aber folgende spezielle Quantile (1.00-Quantile bzw. 0.00-Quantile decken jeweils die Gesamtheit ab und sind daher statistisch irrelevant!):

- Quartile ($\tilde{x}_{0.25}, \tilde{x}_{0.50}, \tilde{x}_{0.75}$)
- Quintile ($\tilde{x}_{0.20}, \tilde{x}_{0.40}, \tilde{x}_{0.60}, \tilde{x}_{0.80}$)
- Dezile ($\tilde{x}_{0.10}, \tilde{x}_{0.20}, \tilde{x}_{0.30}, \tilde{x}_{0.40}, \tilde{x}_{0.50}, \tilde{x}_{0.60}, \tilde{x}_{0.70}, \tilde{x}_{0.80}, \tilde{x}_{0.90}$)
- Perzentile ($\tilde{x}_{0.01}, \tilde{x}_{0.02}, \dots, \tilde{x}_{0.98}, \tilde{x}_{0.99}$)

Die Berechnung der Quantile hängt zum einen von der Wahl des p -Werts ab, zum anderen davon ob die Multiplikation der Gesamtzahl der Fälle n mit p *gerade* oder *ungerade* ist.

wenn $n * p$ ganzzahlig ist:

$$\tilde{x}_p = \frac{1}{2} (x_{n*p} + x_{n*p+1})$$

wenn $n * p$ nicht ganzzahlig ist:

$$\tilde{x}_p = x_{[n*p]}$$

Standardmäßig werden Quartile wie das 25%-Quantil ($p = 0.25$) und das 75%-Quantil ($p = 0.75$) berechnet. Das 50%-Quantil entspricht dem **Median**.

3 Streumaße

3.1 Spannweite/Variationsweite

Die Spannweite (Variationsweite) errechnet sich aus dem Abstand des kleinsten und größten Falls:

$$V = R = x_{\max} - x_{\min}$$

3.2 Interquartilsabstand

Der Interquartilsabstand ist der Abstand zwischen dem 25%-Quartil und dem 75%-Quartil. Der Interquartilsabstand gibt also die mittleren 50% einer Verteilung an:

$$IQR = \tilde{x}_{0.75} - \tilde{x}_{0.25}$$

3.3 Varianz

Die empirische Varianz einer Vollerhebung errechnet sich nach folgender Formel:

$$\sigma^2 = \frac{1}{n} * \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Für die Berechnung der Varianz einer Stichprobe wird die Formel um die Korrektur $\frac{1}{n-1}$ angepasst:

$$s^2 = \frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

3.4 Standardabweichung

Die empirische Standardabweichung einer Vollerhebung errechnet sich nach folgender Formel:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n} * \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Für die Berechnung der Standardabweichung einer Stichprobe wird die Formel um die Korrektur $\frac{1}{n-1}$ angepasst:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

3.5 Variationskoeffizient

Der Variationskoeffizient ist eine dimensionslose Größe und ein relatives Maß. Der Variationskoeffizient berechnet sich aus der **Standardabweichung** und dem **arithmetischen Mittelwert**:

$$v = \frac{s}{\bar{x}}$$

Es gilt $0 \leq v \leq \sqrt{n-1}$. Der Variationskoeffizient wird angewendet, wenn die Streuung verschiedener Variablen verglichen werden soll. Wichtig: Er ist nur ein sinnvolles Maß, sofern nur positive oder nur negative Werte im Merkmal vorliegen.

3.6 Kovarianz

Die Kovarianz ist ein Maß für den linearen Zusammenhang zweier Variablen.

Die Kovarianz wird in der Grundgesamtheit wie folgt berechnet:

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Die Kovarianz wird in der Stichprobe wie folgt berechnet:

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Für die Interpretation der Kovarianz gilt die folgende Hilfe: $\text{cov} < 0$: negativer linearer Zusammenhang $\text{cov} = 0$: kein linearer Zusammenhang (aber nicht lineare Zusammenhänge möglich) $\text{cov} > 0$: positiver linearer Zusammenhang

Wichtig: Die Berechnung prüft nicht, ob der Zusammenhang linear ist. Dies ist eine Annahme vor der Durchführung der Berechnung und muss zuvor überprüft werden.

4 Kreuztabellen

4.1 Erwartete Häufigkeit

Erwartete Häufigkeiten stellen die Werte dar, die sich bei Unwissenheit der einzelnen Zelle aus den Randsummen der Tabelle ergeben würden. Erwartete Häufigkeiten werden oft mit $\hat{h}_{i,j}$ angegeben. Erwartete Häufigkeit einer Zelle in einer Kreuztabelle errechnet sich nach folgender Formel:

$$\hat{h}_{i,j} = \frac{h_{i,.} * h_{.,j}}{n}$$

wobei $h_{i,.}$ = Zeilensumme und $h_{.,j}$ = Spaltensumme gilt

5 Zusammenhangsmaße

Zusammenhangsmaße können zum einen auf χ^2 beruhen oder zum anderen auf der Verbesserung des Vorhersagefehlers (PRE-Maße). Korrelationen sind ebenfalls Zusammenhangsmaße, die neben der Stärke auch die Richtung angeben können. Die Tabelle zeigt, welches Zusammenhangsmaß für welches Skalenniveau genutzt werden kann.

Tabelle 1: Zusammenhangsmaße & Skalenniveau

Skalenniveau	nominal	ordinal	metrisch
nominal	Cramers V / ϕ λ C	Cramers V / ϕ λ C	η^2
ordinal	Cramers V / ϕ λ C	Spearman's ρ	Spearman's ρ
metrisch	η^2	Spearman's ρ	Pearson's r

5.1 Zusammenhangsmaße basierend auf χ^2

5.1.1 χ^2

χ^2 ist ein Maß, um die Unabhängigkeit von zwei Variablen zu berechnen. Auf Grundlage von χ^2 werden die Zusammenhangsmaße ϕ und Cramers V berechnet.

Formel zur Berechnung von χ^2 :

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(h_{ij} - \hat{h}_{ij})^2}{\hat{h}_{ij}}$$

wobei I die Anzahl der Zeilen ist und J die Anzahl der Spalten.

5.1.2 ϕ -Koeffizient

Der ϕ -Koeffizient wird für 2x2-Kreuztabellen auf Grundlage von χ^2 berechnet:

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

wobei $\phi \in [0; 1]$

5.1.3 Cramers V

Cramers V ist ein Zusammenhangsmaß für alle Größen von Kreuztabellen. Zur Berechnung von Cramers V müssen die Variablen mindestens nominales Skalenniveau aufweisen. Es berechnet sich nach folgender Formel:

$$\text{Cramers V} = \sqrt{\frac{\chi^2}{\chi^2_{\max}}} = \sqrt{\frac{\chi^2}{n * (\min(I, J) - 1)}}$$

wobei $V \in [0; 1]$

wobei n : Anzahl Beobachtungen; I : Anzahl Spalten; J : Anzahl Zeilen

Für die Interpretation von Cramers V gelten folgende Daumenregeln:

$\text{Cramers V} < 0.10$: kein Zusammenhang

$0.10 \leq \text{Cramers V} < 0.30$: geringer Zusammenhang

$0.30 \leq \text{Cramers V} < 0.60$: mittlerer Zusammenhang und

$0.60 \leq \text{Cramers V}$: starker Zusammenhang

5.1.4 Kontingenzkoeffizient C

Der Kontingenzkoeffizient ist ein Zusammenhangsmaß für mindestens nominalskalierte Variablen. Er berechnet sich nach folgender Formel:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$C_{\max} = \sqrt{\frac{R - 1}{R}}$$

wobei gilt $R = \min(k, m)$, mit k : Anzahl Spalten und m : Anzahl Zeilen

$$C_{norm} = \frac{C}{C_{max}}$$

wobei gilt $C_{norm} \in [0, 1]$

5.2 PRE-Maße

Ziel dieser Maße ist es auszudrücken, wie gut durch die Kenntnis einer Variablen die Ausprägungen einer weiteren Variablen vorhergesagt werden können. Das errechnete Maß drückt dann diese Verbesserung aus. Bei diesen Maßen wird zwischen unabhängiger und abhängiger Variable unterschieden. Es wird die Wirkung der unabhängigen auf die abhängige Variable errechnet.

5.2.1 λ

λ wird zur Berechnung nominaler Variablen genutzt.

Die Formel lautet:

$$\lambda = \frac{E_1 - E_2}{E_1}$$

Der Fehler 1 (*Error 1* E_1) wird wie folgt berechnet:

$$E_1 = 1 - \frac{h_M}{n}$$

wobei gilt h_M : absolute Häufigkeit der Modalkategorie

Der Fehler 2 (*Error 2* E_2) wird wie folgt berechnet:

$$E_2 = \sum_{j=1}^J \frac{h_{\cdot j}}{n} \left(1 - \frac{h_{Mj}}{h_{\cdot j}}\right)$$

wobei gilt:

- h_M : absolute Häufigkeit der Modalkategorie
- $h_{\cdot j}$: absolute Häufigkeit für die jeweilige Kategorie der unabhängigen Variable
- h_{Mj} : absolute Häufigkeit der Modalkategorie in Abhängigkeit der unabhängigen Variable
- J : Anzahl Spalten

Generell gilt bei PRE-Maßen: $0 < PRE < 1$ und $0 < E_2 < E_1$.

Für die Interpretation von λ gelten folgende Daumenregeln (Cohen 1998):

$\lambda < 0.20$: kein Zusammenhang

$0.20 \leq \lambda < 0.50$: kleiner Zusammenhang

$0.50 \leq \lambda < 0.80$: mittlerer Zusammenhang und

$0.80 \leq \lambda$: großer Zusammenhang

5.2.2 η^2

η^2 wird genutzt, um den Zusammenhang zwischen einer metrischen abhängigen Variable und einer kategorialen unabhängigen Variable zu berechnen.

$$\eta^2 = \frac{QS_{zwischen}}{QS_{gesamt}} = \frac{QS_{zwischen}}{QS_{zwischen} + QS_{in}} = \frac{QS_{gesamt} - QS_{in}}{QS_{gesamt}}$$

wobei QS: Quadratsumme ist.

Für die Interpretation von η^2 gelten folgende Daumenregeln (Cohen 1998):

$\eta^2 < 0.01$: kein Zusammenhang

$0.01 \leq \eta^2 < 0.04$: geringer Zusammenhang

$0.04 \leq \eta^2 < 0.16$: mittlerer Zusammenhang und

$0.16 \leq \eta^2$: starker Zusammenhang

5.3 Korrelationsmaße

Gemeinsam haben diese Korrelationskoeffizienten, dass sie alle Werte zwischen $[-1, +1]$ einnehmen und gleich interpretiert werden. Positive Werte bedeuten, dass mit dem Steigen der Variable A auch Variable B steigt und negative Werte bedeuten, dass beim Steigen der Variable A die Variable B sinkt.

Für die Interpretation von Korrelationsmaßen (r) gilt die folgende Hilfe:

$r < |0.05|$: zu vernachlässigen

$|0.05| \leq r < |0.20|$: geringer Zusammenhang

$|0.20| \leq r < |0.50|$: mittlerer Zusammenhang

$|0.50| \leq r < |0.70|$: starker Zusammenhang

$r \geq |0.7|$: sehr starker Zusammenhang

5.3.1 Spearmans ρ

Spearmans ρ wird zur Berechnung von mindestens ordinalskalierten Variablen berechnet. Dabei werden den Ausprägungen auf den einzelnen Variablen Ränge zugeordnet und die Datenreihe nach den Rängen einer (!) Variablen geordnet.

Die vereinfachte Formel zur Berechnung von Spearmans ρ lautet (bei max. 20 % Ties):

$$\rho = 1 - \frac{6 * \sum_{i=1}^n d_i^2}{n * (n^2 - 1)}$$

wobei $d_i = R(x_i) - R(y_i)$ (Differenz der Ränge)

5.4 Pearsons r

Pearsons r kann für metrisch-skalierte Variablen genutzt werden, wenn der Zusammenhang monoton und linear ist. Pearsons r errechnet sich aus der Kovarianz. Die Formel zur Berechnung von Pearsons r lautet:

$$r = \frac{\text{cov}(x; y)}{s_x * s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 * \sum_{i=1}^n (y_i - \bar{y})^2}}$$

mit $x = (x_1, x_2, x_3, \dots, x_n)$ und $y = (y_1, y_2, y_3, \dots, y_n)$

6 Standardisierung

Mithilfe der z-Transformation werden Variablen auf die Maßeinheit Standardabweichung standardisiert. Dies berechnet sich nach folgender Formel:

$$z_i = \frac{x_i - \bar{x}}{s}$$

7 Population und Stichprobe

7.1 Population

Mittelwert: μ_x

Standardabweichung: σ_x

Varianz: σ_x^2

7.2 Stichprobe

Mittelwert: \bar{x}

Standardabweichung: s_x

Varianz: s_x^2

Standardfehler des Mittelwertes / Stichprobenfehler:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}$$

Schätzung der Populationsvarianz:

Um die Populationsvarianz zu schätzen, wird die folgende Korrektur angewandt $\frac{n}{n-1}$. Da in der Berechnung der Stichprobenvarianz (s^2) bereits der Korrekturfaktor $n - 1$ inkludiert ist, muss diese Korrektur hier nicht mehr angewandt werden:

$$\hat{\sigma}_x^2 = s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

In Lehrbüchern, die bei der Berechnung der Stichprobenvarianz (s^2) nicht die Korrektur ($n - 1$) anwenden, erhält man folgende Formel, die äquivalent zur obigen ist.

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} * \frac{n}{n - 1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Schätzung der Populationsstandardabweichung:

$$\hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2} = s_x$$

Auch hier berücksichtigen wir, dass wir bereits die Korrektur bei der Berechnung der Stichprobenvarianz angewendet haben.

geschätzter Standardfehler des Mittelwertes (Standardfehler mit geschätzter Populationsvarianz):

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}_x^2}{n}} = \frac{\hat{\sigma}_x}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n * (n - 1)}}$$

8 Signifikanztest

Nullhypothese: H_0

Alternativhypothese: H_A

z.B.: $H_A : \mu > 2300$ bzw. $H_0 : \mu \leq 2300$

Grenzen:

einseitiger Test: $p = 1 - \alpha$

zweiseitiger Test: $p = 1 - \frac{\alpha}{2}$

8.1 Testverfahren

8.1.1 z-Test

Der z-Test dient zum Vergleich eines Stichprobenmittelwertes mit einem Populationsmittelwert. Dazu muss μ und σ bekannt sein. Der empirische z-Wert wird entsprechend der Irrtumswahrscheinlichkeit mit der Prüfgröße (kritischer Wert) verglichen.

Vorannahme: μ und σ sind bekannt.

$$z_{emp} = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{\text{Stichprobenmittelwert} - \text{theoretisch erwarteter Mittelwert}}{\text{Standardfehler der Stichprobenmittelwerte}}$$

Der kritische Wert wird wie folgt festgelegt:

$$z_{krit} = z_{(\phi)}$$

wobei $\phi = 1 - \alpha$

Wenn beidseitig getestet wird, ist der Test signifikant, sofern $|z_{emp}| > z_{krit}$ gilt.

Ist der Test einseitig, wird die Richtungsgröße getestet. Also $z_{emp} < z_{krit}$ (negativ) bzw. $z_{emp} > z_{krit}$ (positiv).

8.1.2 t-Test

Der t-Test kann angewendet werden, wenn nur der Mittelwert der Grundgesamtheit (μ), nicht aber die Standardabweichung (σ) bekannt ist.

$$t_{emp} = \frac{\bar{x} - \mu_x}{\hat{\sigma}_x} * \sqrt{n}$$

Der kritische t-Wert wird in Abhängigkeit der Freiheitsgrade (df / v) und dem Signifikanzniveau ($1 - \alpha$) festgelegt.

$$t_{krit} = t_{1-\alpha, v}$$

Die Freiheitsgrade berechnen sich wie folgt:

$$v = n - 1$$

Ist der Test einseitig, wird die Richtungsgröße getestet. Also $t_{emp} < t_{krit}$ (negativ) bzw. $t_{emp} > t_{krit}$ (positiv).

8.1.3 χ^2 -Test

Dieser Test kann durchgeführt werden, wenn das Merkmal nominal ist.
 χ^2 berechnet sich wie zuvor:

$$\chi_{emp}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(h_{ij} - \hat{h}_{ij})^2}{\hat{h}_{ij}}$$

Der kritische χ^2 -Wert wird in Abhängigkeit des Signifikanzniveaus bzw. der Freiheitsgrade berechnet:

$$\chi_{krit}^2 = \chi_{1-\alpha, v}^2$$

Die Freiheitsgrade berechnen sich wie folgt:

$$v = n - 1$$

Ist der Test einseitig, wird die Richtungsgröße getestet. Also $\chi_{emp}^2 < \chi_{krit}^2$ (negativ) bzw. $\chi_{emp}^2 > \chi_{krit}^2$ (positiv).

8.2 χ^2 -Test auf Unabhängigkeit

Liegen zwei Variablen vor, kann ein χ^2 -Test auf Unabhängigkeit berechnet werden (siehe 5.1.1).
Dazu muss nur zusätzlich der kritische Wert bestimmt werden.

$$\chi_{krit}^2 = \chi_{1-\alpha, v}^2$$

Die Freiheitsgrade berechnen sich beim Unabhängigkeitstest wie folgt:

$$v = (I - 1) * (J - 1)$$

wobei I die Anzahl der Zeilen ist und J die Anzahl der Spalten.

8.3 Berechnung Konfidenzintervall

Das Konfidenzintervall gibt mit einer ausgewählten Irrtumswahrscheinlichkeit die Intervallgrenzen eines Wertes an. Um das Konfidenzintervall mit der z-Verteilung zu berechnen, muss die Standardabweichung der Grundgesamtheit bekannt sein.

8.3.1 anhand z-Verteilung (bekanntes σ)

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}$$

Obergrenze Intervall:

$$x_O = \bar{x} + z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}$$

Untergrenze Intervall:

$$x_U = \bar{x} - z_{1-\frac{\alpha}{2}} * \sigma_{\bar{x}}$$

Wenn σ_x unbekannt ist, kann der Standardfehler des Mittelwertes durch die Stichprobe geschätzt werden ($\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}_x}{\sqrt{n}}$). Es ergibt sich dabei dann aber eine doppelte Schätzung, da sowohl die Standardabweichung der Grundgesamtheit ($\hat{\sigma}_x$) als auch der Standardfehler des Mittelwertes ($\hat{\sigma}_{\bar{x}}$) über die Stichprobe geschätzt werden.

Für ein 99%-Konfidenzintervall beträgt $z = 2.58$, für ein 95%-Konfidenzintervall beträgt $z = 1.96$ und für ein 90%-Konfidenzintervall beträgt $z = 1.645$

8.3.2 anhand t-Verteilung (unbekanntes σ)

Oftmals wird die t-Verteilung anstatt der z-Verteilung genutzt, um Konfidenzintervalle zu berechnen. Um die t-Verteilung nutzen zu können, muss zuvor ein Verteilungsparameter festgelegt werden: Die Freiheitsgrade. Dieser Verteilungsparameter wird oft mit v oder df (*degree of freedom*) angegeben.

Zur Berechnung eines Konfidenzintervalls wird die Populationsvarianz dann auf Basis der t-Verteilung und der gewählten Irrtumswahrscheinlichkeit geschätzt.

$$\bar{x} \pm t_{(1-\frac{\alpha}{2};v)} * \frac{\sigma_x}{\sqrt{n}} = \bar{x} \pm t_{(1-\frac{\alpha}{2};v)} * \sigma_{\bar{x}}$$

Hierbei gilt für die Berechnung des Verteilungsparameters der t-Verteilung:

$$v = n - 1$$

Obergrenze Intervall:

$$x_O = \bar{x} + t_{(1-\frac{\alpha}{2};v)} * \sigma_{\bar{x}}$$

Untergrenze Intervall:

$$x_U = \bar{x} - t_{(1-\frac{\alpha}{2};v)} * \sigma_{\bar{x}}$$

Auch hier wird $\sigma_{\bar{x}}$ über die Stichprobe wie folgt geschätzt:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}_x^2}{n}} = \frac{\hat{\sigma}_x}{\sqrt{n}}$$

$$\text{wobei } \hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

9 Mittelwertvergleich zwischen zwei Gruppen

Beim Mittelwertvergleich zwischen zwei Gruppen werden verschiedene Testsituationen unterschieden. Je nach Testsituation wird der t-Test anders berechnet. Die für die Klausur relevanten Varianten des Mittelwertvergleichs für zwei Gruppen werden im Folgenden dargestellt. Der empirische t-Wert wird dabei mit t_{emp} dargestellt.

9.1 Zweistichproben t-Test für unabhängige Stichproben mit homogenen Varianzen

In diesem Beispiel wird der t-Test für unabhängige Stichproben mit homogenen Varianzen berechnet. Der empirische t-Wert berechnet sich wie folgt:

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} = \frac{\text{Mittelwertdifferenz der Stichproben}}{\text{Standardfehler der Mittelwertdifferenz}}$$

Der Standardfehler der Mittelwertdifferenz wird wie folgt geschätzt:

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Die Anzahl der Freiheitsgrade ergibt sich wie folgt:

$$v = n_1 + n_2 - 2$$

9.2 Zweistichproben t-Test für unabhängige Stichproben mit heterogenen Varianz

Dieser Test wird auch Welch-Test oder t-Test mit Welch-Korrektur genannt. Wichtig ist, dass die Anzahl der Freiheitsgrade differenziert berechnet werden.

Der empirische t-Wert berechnet sich wie folgt:

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Die Anzahl der Freiheitsgrade berechnet sich bei ungleichen Varianzen wie folgt:

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

9.3 Zweistichproben t-Test für abhängige Stichproben

Als letztes Testszenario wird der Zweistichproben t-Test für abhängige Stichproben vorgestellt. Der empirische t-Test berechnet sich wie folgt:

$$t_{emp} = \frac{\bar{x}_d - \mu_d}{\hat{\sigma}_{\bar{x}_d}} = \frac{\bar{x}_d - 0}{\hat{\sigma}_{\bar{x}_d}} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}$$

mit

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}}$$

und

$$\hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{x}_d)^2}{n-1}}$$

und

$$d_i = x_{1_i} - x_{2_i}$$

und

$$\bar{x}_d = \frac{1}{n} \sum_{i=1}^n x_{1_i} - x_{2_i}$$

und

n = Fallzahl der verbundenen Stichprobe

Die Freiheitsgrade werden wie folgt bestimmt:

$$v = n - 1$$

Alternativ ist auch folgende kurze Herbeiführung möglich:

$$t_{emp} = \frac{\bar{x}_d - \mu_d}{\frac{\hat{\sigma}_d}{\sqrt{n}}} = \frac{\bar{x}_d - 0}{\frac{\hat{\sigma}_d}{\sqrt{n}}} = \frac{\bar{x}_d}{\frac{\hat{\sigma}_d}{\sqrt{n}}}$$

10 Regression

Die lineare Regression beruht auf dem Modell der linearen Darstellung. Die Grundformel einer linearen Darstellung ist folgende:

$$y_i = \beta_0 + \beta_1 * x_i$$

10.1 Bivariate lineare Regression

Die Modellgleichung der linearen Regression der beobachteten abhängigen Variable (Y) durch die unabhängige Variable (X) lautet:

$$y_i = \beta_0 + \beta_1 * x_i + \epsilon_i$$

wobei ϵ_i die Störgröße darstellt.

10.2 Modellschätzung bivariate lineare Regression

Im Regressionsmodell werden die Regressionskoeffizienten geschätzt, woraus sich der geschätzte Wert der unabhängigen Variable ergibt. Dies wird Modellschätzung oder Schätzgleichung genannt. Sie lautet wie folgt:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i$$

Die Schätzgleichung für die beobachteten Werte lautet in Anpassung an die Modellgleichung wie folgt:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 * x_i + \hat{\epsilon}_i$$

Wichtig: $\hat{\epsilon}$ stellt nicht mehr Störgrößen dar, sondern das empirische Residuum. Das Residuum ($\hat{\epsilon}$) berechnet sich als Nebenprodukt der Schätzung aus der Differenz der geschätzten und beobachteten Werte bzw. der geschätzten Regressionskoeffizienten:

$$\epsilon_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 * x_i$$

Die Schätzung der Regressionskoeffizienten erfolgt aus den beobachteten Werten der abhängigen und unabhängigen Variable:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 * \bar{x} = \bar{y} - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} * \bar{x}$$

Aus der Schätzung der Regressionskoeffizienten ergeben sich geschätzte Werte der unabhängigen Variable. Um nun die beste Schätzung zu erreichen, wird durch OLS-Schätzung die Gerade gesucht, deren quadrierte Abstände der geschätzten Werte minimal zu den beobachteten Werten ist:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 * x_i)^2 \rightarrow \min.!$$

Varianzzerlegung der linearen Regression:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SS_{total} = SS_{model} + SS_{residual}$$

$$SS_Y = SS_X + SS_e$$

\hat{y}_i : geschätzte Werte,
 y_i : beobachtete Werte,
 \bar{y} : Mittelwert der beobachteten Werte

Bestimmtheitsmaß:

$$R^2 = \frac{SS_{model}}{SS_{total}}$$

wobei $R^2 \in [0; 1]$

korrigiertes R^2 : $R_{korr}^2 = 1 - \frac{(1-R^2)*(n-1)}{(n-p-1)}$

10.3 Multivariate lineare Regression

Für die multivariate lineare Regression werden nun k -viele weitere unabhängige Variablen hinzugefügt.

Die Modellgleichung lautet daher:

$$y_i = \beta_0 + \beta_1 * x_{1_i} + \beta_2 * x_{2_i} + \dots + \beta_k * x_{k_i} + \varepsilon_i$$

wobei k Laufindex für Variablen ist und i der Laufindex für die Beobachtung.

Die Schätzgleichung lautet wie folgt:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 * x_{1_i} + \hat{\beta}_2 * x_{2_i} + \dots + \hat{\beta}_{k_i} * x_{k_i}$$

Die Schätzgleichung für die beobachteten Werte lautet in Anpassung an die Modellgleichung wie folgt:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 * x_{1_i} + \hat{\beta}_2 * x_{2_i} + \dots + \hat{\beta}_{k_i} * x_{k_i} + \hat{\epsilon}_i$$

10.4 Standardisierte Regressionskoeffizienten

Um die Stärke von einzelnen Regressionskoeffizienten im Modell gegenüber anderen unabhängigen Variablen im Modell zu berechnen, werden standardisierte Regressionskoeffizienten berechnet.

$$\beta_k = b_k * \frac{s_{x_k}}{s_y}$$

11 z-Tabelle

Die Tabelle enthält z-Werte, die auf zwei Stellen hinter dem Komma gerundet sind: z.B. -2.03, 1.07 oder 1.96.

Leseübung: Die Tabelle ist in zwei Teile aufgeteilt und somit ist auch jeder z-Wert in zwei Teile aufgeteilt: *Teil 1* mit der ersten Nachkommastelle (Spalte 1) und *Teil 2* mit der zweiten Nachkommastelle (alle folgenden Spalten). Jetzt suchen wir die Wahrscheinlichkeit (als Wert der Funktion $\phi_{0,1}(z)$), dass maximal ein z-Wert von -1.44 auftritt: In *Teil 1* geht man zur Zeile -1.4 und in dieser Zeile dann in die Spalte mit der zweiten Nachkommastelle 0.04 (*Teil 2*). Die gesuchte Wahrscheinlichkeit beträgt $\phi_{0,1}(z) = 0.0749$. Das heißt, die Wahrscheinlichkeit das ein z-Wert kleiner gleich -1.44 ist beträgt 7.49 % (bzw. grafisch: die Fläche bis zum z-Wert von -1.44 beträgt 7.49 % der gesamten Fläche).

z-Wert	-.-0	-.-1	-.-2	-.-3	-.-4	-.-5	-.-6	-.-7	-.-8	-.-9
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224

Fortsetzung z-Tabelle

z-Wert	-.0	-.1	-.2	-.3	-.4	-.5	-.6	-.7	-.8	-.9
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

12 t-Tabelle

Die folgende Tabelle zeigt ausgewählte Werte der inversen Verteilungsfunktion der t -Verteilung: $t(1 - \alpha | \nu)$. Für ausgewählte Freiheitsgrade (ν) und Wahrscheinlichkeiten ($1 - \alpha$) werden die entsprechenden t -Werte dargestellt.

Leseübung: Gesucht sei der t -Wert, unter dem bei $\nu = 12$ 90% aller möglichen Werte einer t -verteilten Zufallsvariable liegen.

In der Zeile $\nu = 12$ geht man zur Spalte ($1 - \alpha = 0.9$) und findet den gesuchten Werte $t = 1.356$. 90% aller Werte einer mit 12 Freiheitsgraden t -verteilten Zufallsvariable sind kleiner als 1.356.

df/ν	Fläche ($1 - \alpha$)									
	0.65	0.7	0.75	0,8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12,706	31,821	63,656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
30	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
50	0.388	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678
60	0.387	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660
70	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648
80	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639
90	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632
100	0.386	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626
150	0.386	0.526	0.676	0.844	1.040	1.287	1.655	1.976	2.351	2.609
200	0.386	0.525	0.676	0.843	1.039	1.286	1.653	1.972	2.345	2.601
500	0.386	0.525	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586
1000	0.385	0.525	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581

13 χ^2 -Tabelle

In der Tabelle sind die χ^2 -Werte angegeben, die linksseitig eine Fläche der Größe $1 - \alpha$ mit einer bestimmten Anzahl an Freiheitsgraden (v/df) abschneidet. In der Tabelle sind also inverse Werte der Verteilungsfunktion $\chi^2(1 - \alpha|v)$ eingetragen.

df/v	Fläche ($1 - \alpha$)								
	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	1.07	1.32	1.64	2.07	2.71	3.84	5.02	6.63	7.88
2	2.41	2.77	3.22	3.79	4.61	5.99	7.38	9.21	10.60
3	3.66	4.11	4.64	5.32	6.25	7.81	9.35	11.34	12.84
4	4.88	5.39	5.99	6.74	7.78	9.49	11.14	13.28	14.86
5	6.06	6.63	7.29	8.12	9.24	11.07	12.83	15.09	16.75
6	7.23	7.84	8.56	9.45	10.64	12.59	14.45	16.81	18.55
7	8.38	9.04	9.80	10.75	12.02	14.07	16.01	18.48	20.28
8	9.52	10.22	11.03	12.03	13.36	15.51	17.53	20.09	21.95
9	10.66	11.39	12.24	13.29	14.68	16.92	19.02	21.67	23.59
10	11.78	12.55	13.44	14.53	15.99	18.31	20.48	23.21	25.19
11	12.90	13.70	14.63	15.77	17.28	19.68	21.92	24.73	26.76
12	14.01	14.85	15.81	16.99	18.55	21.03	23.34	26.22	28.30
13	15.12	15.98	16.98	18.20	19.81	22.36	24.74	27.69	29.82
14	16.22	17.12	18.15	19.41	21.06	23.68	26.12	29.14	31.32
15	17.32	18.25	19.31	20.60	22.31	25.00	27.49	30.58	32.80
16	18.42	19.37	20.47	21.79	23.54	26.30	28.85	32.00	34.27
17	19.51	20.49	21.61	22.98	24.77	27.59	30.19	33.41	35.72
18	20.60	21.60	22.76	24.16	25.99	28.87	31.53	34.81	37.16
19	21.69	22.72	23.90	25.33	27.20	30.14	32.85	36.19	38.58
20	22.77	23.83	25.04	26.50	28.41	31.41	34.17	37.57	40.00
21	23.86	24.93	26.17	27.66	29.62	32.67	35.48	38.93	41.40
22	24.94	26.04	27.30	28.82	30.81	33.92	36.78	40.29	42.80
23	26.02	27.14	28.43	29.98	32.01	35.17	38.08	41.64	44.18
24	27.10	28.24	29.55	31.13	33.20	36.42	39.36	42.98	45.56
25	28.17	29.34	30.68	32.28	34.38	37.65	40.65	44.31	46.93
30	33.53	34.80	36.25	37.99	40.26	43.77	46.98	50.89	53.67
40	44.16	45.62	47.27	49.24	51.81	55.76	59.34	63.69	66.77
50	54.72	56.33	58.16	60.35	63.17	67.50	71.42	76.15	79.49
60	65.23	66.98	68.97	71.34	74.40	79.08	83.30	88.38	91.95
70	75.69	77.58	79.71	82.26	85.53	90.53	95.02	100.43	104.21
80	86.12	88.13	90.41	93.11	96.58	101.88	106.63	112.33	116.32
90	96.52	98.65	101.05	103.90	107.57	113.15	118.14	124.12	128.30
100	106.91	109.14	111.67	114.66	118.50	124.34	129.56	135.81	140.17
150	158.58	161.29	164.35	167.96	172.58	179.58	185.80	193.21	198.36
200	209.99	213.10	216.61	220.74	226.02	233.99	241.06	249.45	255.26
500	516.09	520.95	526.40	532.80	540.93	553.13	563.85	576.49	585.21