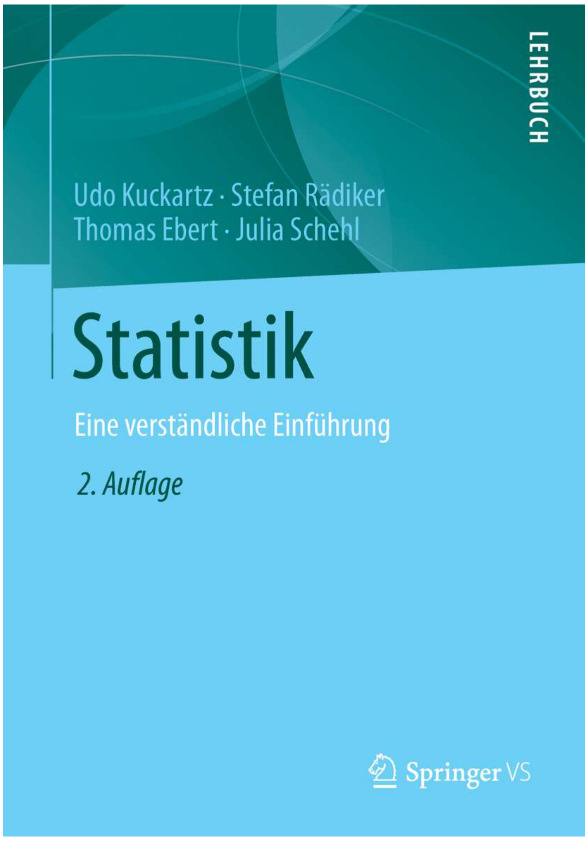
Udo Kuckartz, Stefan Rädiker, Thomas Ebert und Julia Schehl

Statistik
Eine verständliche Einführung

2. Auflage, Springer VS, 313 Seiten, 2013



Formelsammlung

Stand: 19.01.2019

magma Marburger Arbeitsgruppe für Methoden & Evaluation

info@magma-research.de

Kapitel 2 - Häufigkeitsverteilungen und ihre grafischen Darstellungen

Berechnung von Prozenten (S. 37)

$$\%k = \frac{f(k)}{n} \cdot 100\%$$

$f(k)$ = absolute Häufigkeit in der Kategorie k
n = Anzahl der Fälle

Berechnung von gültigen Prozenten (S. 38)

$$\%k = \frac{f(k)}{n_g} \cdot 100\%$$

$f(k)$ = absolute Häufigkeit in der Kategorie k
n_g = Anzahl der Fälle mit gültigen Werten

Berechnung von kumulierten Prozenten (S. 38)

$$\%k_{kum} = \frac{f_{kum}(k)}{n_g} \cdot 100\%$$

$f(k)$ = absolute Häufigkeit in der Kategorie k
n_g = Anzahl der Fälle mit gültigen Werten

Kapitel 3 - Mittelwerte und Streuungsmaße

Mittelwert (S. 64)

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

x_i = Messwert des i-ten Falls
n = Anzahl der Fälle

Mittelwert für eine Häufigkeitstabelle (S. 66)

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i \cdot k_i}{\sum_{i=1}^{m} f_i} = \frac{f_1 \cdot k_1 + f_2 \cdot k_2 + f_3 \cdot k_3 + \ldots + k_m}{f_1 + f_2 + f_3 + \ldots + f_m}$$

f_i = Die Häufigkeit des Vorkommens einer Kategorie
k_i = Der Wert einer Kategorie
m = Die Anzahl der Kategorien

Mittelwert für gruppierte Daten (S. 67)

$$\bar{x} = \frac{\sum_{k=1}^{m} f_k \cdot x_k}{\sum_{k=1}^{m} f_k} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + \ldots + f_m \cdot x_m}{\sum_{k=1}^{m} f_k}$$

f_k = Die Häufigkeit des Vorkommens einer Kategorie
x_k = Die Kategorienmitte
m = Die Anzahl der Kategorien

Spannweite (S. 69)

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

x_{max} = höchster Wert einer Verteilung
x_{min} = niedrigster Wert einer Verteilung

Interquartilsabstand (S. 71)

$$IQR = 3.\,Quartil - 1.\,Quartil = Q_{.75} - Q_{.25}$$

$Q_{.75}$ = Wert einer Verteilung, der die oberen 25% der Verteilung abschneidet
$Q_{.25}$ = Wert einer Verteilung, der die unteren 25% der Verteilung abschneidet

Varianz (S. 71)

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

x_i = Wert eines Falls
$\bar{x}$ = Mittelwert
n = Anzahl der Fälle

Varianz (Inferenzstatistik) (S. 71)

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

x_i = Wert eines Falls
$\bar{x}$ = Mittelwert
n = Anzahl der Fälle

Standardabweichung (S. 72)

$$s = \sqrt{s^2}$$

s^2 = Varianz

Variationskoeffizient (S. 76)

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

s = Standardabweichung
$\bar{x}$ = Mittelwert

z-Tranzformation (S. 77)

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

x_i = einzelner Messwert
$\bar{x}$ = Mittelwert der Verteilung
s = Standardabweichung der Verteilung

Kapitel 4 - Kreuztabelle, Chi-Quadrat und Zusammenhangsmaße

Berechnung der erwarteten Häufigkeit (S. 93)

$$f_{e(i,j)} = n_j \cdot p_i$$

n_j = Anzahl der Fälle in Kategorie j
p_i = relative Häufigkeit für die Ausprägung i

oder

$$f_{e(i,j)} = \frac{Zeilensumme\ i \cdot Spaltensumme\ j}{n}$$

n = Anzahl aller Fälle

Berechnung von Chi-Quadrat bei einer Kreuztabelle (S. 94)

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(f_{b(i,j)} - f_{e(i,j)})^2}{f_{e(i,j)}}$$

$f_{b(i,j)}$ = beobachtete Häufigkeit in der i-ten Zeile und j-ten Spalte
$f_{e(i,j)}$ = erwartete Häufigkeit in der i-ten Zeile und j-ten Spalte
k = Anzahl der Zeilen
l = Anzahl der Spalten

Freiheitsgrade $df = (k - 1) \cdot (l - 1)$

Phi-Koeffizient (S. 98)

$$\Phi = \frac{a \cdot d - b \cdot c}{\sqrt{(a + c) \cdot (b + d) \cdot (a + b) \cdot (c + d)}}$$

a = erste Zelle einer Vier-Felder-Tafel (links oben)
b = zweite Zelle einer Vier-Felder-Tafel (rechts oben)
c = dritte Zelle einer Vier-Felder-Tafel (links unten)
d = vierte Zelle einer Vier-Felder-Tafel (rechts unten)

Beziehung zwischen Phi und Chi-Quadrat (S. 99)

$$\Phi = \sqrt{\frac{\chi^2}{n}} \qquad bzw. \qquad \chi^2 = n \cdot \Phi^2$$

χ^2 = Chi-Quadrat
Φ = Phi-Koeffizient
n = Anzahl aller Fälle

Kontigenzkoeffizient C (S. 99)

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

χ^2 = Chi-Quadrat
n = Anzahl aller Fälle

Maximale Höhe von $C = C_{max} = \sqrt{\frac{R-1}{R}}$

R = Minimum der Kategorienanzahl in den Zeilen und Spalten

Cramers V (S. 99)

$$V = \sqrt{\frac{\chi^2}{n \cdot (R-1)}}$$

χ^2 = Chi-Quadrat
n = Anzahl aller Fälle
R = Minimum der Kategorienanzahl in den Zeilen und Spalten

Berechnung von Chi-Quadrat für univariate Verteilungen (S. 101)

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e\,(j)}}$$

$f_{b(j)}$ = beobachtete Anzahl in Kategorie j
$f_{e(j)}$ = erwartete Anzahl in Kategorie j
k = Anzahl der Kategorien

Kapitel 5 - Wahrscheinlichkeit und Wahrscheinlichkeitsverteilungen

Definition der theoretischen Wahrscheinlichkeit (S. 114)

$$\text{theoretische Wahrscheinlichkeit (Ereignis A)} = \text{P(A)} = \frac{\text{Anzahl der günstigen Ereignisse}}{\text{Anzahl der möglichen Ereignisse}}$$

Definition der empirischen Wahrscheinlichkeit (S. 117)

$$\text{empirische Wahrscheinlichkeit} = \lim_{n \to \infty} \frac{k}{n}$$

k = Anzahl der günstigen Ereignisse
n = Anzahl der durchgeführten Zufallsexperimente

Binomialgleichung (S. 124)

$$\text{P}(n; p; k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \text{wobei} \quad \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

n = Anzahl der durchgeführten Zufallsexperimente
$n!$ = $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ (gelesen: n Fakultät)
p = Wahrscheinlichkeit des interessierenden Ereignisses beim einmaligen Experiment
k = Anzahl, wie häufig das interessierende Ereignis auftreten soll

Binomialverteilung (S. 127)

Mittelwert der Verteilung $\mu = n \cdot p$

Standardabweichung der Verteilung $\sigma = n \cdot p \cdot q = n \cdot p \cdot (1-p)$

n = Anzahl der durchgeführten Zufallsexperimente
p = Wahrscheinlichkeit des interessierenden Ereignisses beim einmaligen Experiment

Kapitel 6 - Die Logik des statistischen Schließens

Standardfehler (= Standardabweichung von Stichprobenmittelwerten) (S. 141)

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

s^2 = Stichprobenvarianz als Schätzer für die Populationsvarianz
n = Anzahl der Fälle in der Stichprobe
s = Standardabweichung in der Stichprobe als Schätzer für die Population

Konfidenzintervall (S. 142)

Untere Grenze $= \mu - z \cdot \hat{\sigma}_{\bar{x}}$

Obere Grenze $= \mu + z \cdot \hat{\sigma}_{\bar{x}}$

μ = Mittelwert der Grundgesamtheit
z = z-Wert der Standardnormalverteilung
$\hat{\sigma}_{\bar{x}}$ = Standardfehler

Kapitel 7 - t-Test: zwei Mittelwerte vergleichen

Berechnung der Prüfgröße t für homogene Varianzen (S. 162)

$$\text{Prüfgröße t} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$$

$\bar{x}_1$ = Mittelwert der Stichprobe 1
$\bar{x}_2$ = Mittelwert der Stichprobe 2
$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}$ = geschätzter Standardfehler der Mittelwertsdifferenz in der Grundgesamtheit

Freiheitsgrade $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$

n_1 = Anzahl der Fälle in Stichprobe 1
n_2 = Anzahl der Fälle in Stichprobe 2

Berechnung der Prüfgröße t für heterogene Varianzen (S. 164)

$$\text{Prüfgröße t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = Mittelwert der Stichprobe 1
$\bar{x}_2$ = Mittelwert der Stichprobe 2
s_1^2 = Varianz der Stichprobe 1
s_2^2 = Varianz der Stichprobe 2
n_1 = Anzahl der Fälle in Stichprobe 1
n_2 = Anzahl der Fälle in Stichprobe 2

Standardfehler der Mittelwertdifferenz (S. 163)

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

n_1 = Anzahl der Fälle in Stichprobe 1
n_2 = Anzahl der Fälle in Stichprobe 2
s_1^2 = Varianz der Stichprobe 1
s_2^2 = Varianz der Stichprobe 2

Effektstärke einer Mittelwertdifferenz (S. 167-168)

$$\text{Effektstärke einer Mittelwertsdifferenz} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_p}$$

$\bar{x}_1$ = Mittelwert der Stichprobe 1 als Schätzung für die Grundgesamtheit
$\bar{x}_2$ = Mittelwert der Stichprobe 2 als Schätzung für die Grundgesamtheit
$\hat{\sigma}_\mathrm{p}$ = geschätzte Standardabweichung der Grundgesamtheit (Population)

$$\text{Cohens } d \text{ für unabhängige Stichproben} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

$\bar{x}_1$ = Mittelwert der Stichprobe 1 als Schätzung für die Grundgesamtheit
$\bar{x}_2$ = Mittelwert der Stichprobe 2 als Schätzung für die Grundgesamtheit
s_1 = Standardabweichung der Stichprobe 1
s_2 = Standardabweichung der Stichprobe 2

Berechnung der Prüfgröße t für abhängige Stichproben (S. 171)

$$\text{Prüfgröße } t = \frac{\bar{x}_D}{\hat{\sigma}_D / \sqrt{n}}$$

$\bar{x}_D$ = Mittelwert der Differenzen aller Wertepaare
(identisch mit der Differenz der beiden Stichprobenmittelwerte)
$\hat{\sigma}_D$ = geschätzte Standardabweichung der Mittelwertsdifferenz in der Grundgesamtheit
n = Anzahl der Wertepaare

$$\bar{x}_D = \frac{\sum_{i=1}^{n} x_{Di}}{n} = \frac{\sum_{i=1}^{n} x_{i1} - x_{i2}}{n}$$

x_{Di} = Differenz der Wertepaare x_{i1} und x_{i2}
x_{i1} = Wert 1 eines Wertepaares
x_{i2} = Wert 2 eines Wertepaares
n = Anzahl der Wertepaare

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

x_{Di} = Differenz der Wertepaare x_{i1} und x_{i2}
$\bar{x}_D$ = Mittelwert der Mittelwertsdifferenzen aller Wertepaare
n = Anzahl der Wertepaare

Freiheitsgrade $df = n - 1$

Effektstärke einer Mittelwertdifferenz bei abhängigen Stichproben (S. 172)

$$\text{Cohens } d_z \text{ für abhängige Stichproben} = \frac{\bar{x}_D}{\hat{\sigma}_D}$$

$\bar{x}_D$ = Mittelwert der Differenzen aller Wertepaare
(identisch mit der Differenz der beiden Stichprobenmittelwerte)
$\hat{\sigma}_D$ = geschätzte Standardabweichung der Differenzen in der Grundgesamtheit

Kapitel 8 - Varianzanalyse: mehr als zwei Mittelwerte vergleichen

Gesamtmittelwert (S. 189)

$$\bar{G} = \frac{G}{N}$$

G = Gesamtsumme aller beobachteten Werte
N = Anzahl aller Werte

Treatment-Quadratsumme (S. 191)

$$QS_{treat} = \sum_{j=1}^{k} m_j \cdot (\bar{A}_j - \bar{G})^2$$

m = Anzahl Personen unter der Faktor j
$\bar{A}_j$ = Gruppenmittelwert unter der Faktorstufe j
$\bar{G}$ = Gesamtmittelwert

Treatmentvarianz (S. 191)

$\hat{\sigma}^2_{treat} = \frac{QS_{treat}}{df}$; Freiheitsgrade $df = k - 1$

QS_{treat} = Treatment-Quadratsumme
k = Anzahl der Faktorstufen

Fehler-Quadratsumme (S. 192)

$$QS_{Fehler} = \sum_{i=1}^{m} \sum_{j=1}^{k} (x_{ij} - \bar{A}_j)^2$$

x_{ij} = Wert der Person i unter der Faktorstufe j
$\bar{A}_j$ = Gruppenmittelwert unter der Faktorstufe j
$\bar{G}$ = Gesamtmittelwert
m = Anzahl der Personen unter der Faktorstufe j
k = Anzahl der Faktorstufen

Fehlervarianz (S. 193)

$\hat{\sigma}^2_{Fehler} = \frac{QS_{Fehler}}{df}$; Freiheitsgrade $df = N - k$

QS_{Fehler} = Fehler-Quadratsumme
N = Anzahl aller Werte
k = Anzahl der Faktorstufen

Prüfgröße F (S. 194)

$$F = \frac{\hat{\sigma}^2_{treat}}{\hat{\sigma}^2_{Fehler}}$$

$\hat{\sigma}^2_{treat}$ = Treatmentvarianz
$\hat{\sigma}^2_{Fehler}$ = Fehlervarianz

Effektstärke eta-Quadrat (S. 195)

Effektstärke $\eta^2 = \frac{QS_{treat}}{QS_{tot}}$

QS_{treat} = Treatment-Quadratsumme
QS_{tot} = totale Quadratsumme

Kapitel 9 - Korrelation: Zusammenhänge identifizieren

Kovarianz (S. 211)

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

x_i und y_i = Wert der Untersuchungseinheit bei Variable x und bei Variable y
$\bar{x}$ und $\bar{y}$ = Mittelwerte der Variablen x und y
n = Anzahl der Untersuchungseinheiten, meistens Personen

Produkt-Moment-Korrelation (= Pearsons r) (S. 212)

$$r = \frac{cov(x, y)}{s_x \cdot s_y}$$

$cov(x,y)$ = Kovarianz der Variablen x und y
s_x und s_y = Standardabweichungen der Variablen x und y

Prüfgröße t für die Moment-Produkt-Korrelation (S. 215)

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

r = ermittelter Korrelationskoeffizient nach Pearson
n = Anzahl der Untersuchungseinheiten, meistens Personen

Spearmans rho (S. 217)

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

d_i^2 = quadrierte Rangplatzdifferenz der i-ten Untersuchungseinheit
n = Anzahl der Untersuchungseinheiten

Prüfgröße t für Spearmans rho (S. 219)

$$t = \frac{r_s}{\sqrt{(1 - r_s^2)/(n-2)}}$$

r_s = Spearmans rho
n = Anzahl der Untersuchungseinheiten

Kapitel 10 - Skalenbildung

Cronbachs-Alpha (S. 247)

$$\alpha = \frac{k \cdot \bar{r}}{1 + (k-1) \cdot \bar{r}}$$

k = Anzahl der Items in der Skala
$\bar{r}$ = durchschnittlicher Korrelationskoeffizient der Items (auch als Homogenität bezeichnet)

Kapitel 11 - Regression: komplexe Zusammenhänge analysieren und Vorhersagen treffen

Steigung der Regressionsgraden (S. 261)

$$b_1 = \frac{cov(x,y)}{s_x^2}$$

$cov(x,y)$ = Kovarianz der beiden Variablen x und y
s_x^2 = Varianz der Variablen x

Achsenabschnitt der Regressionsgeraden (S. 262)

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

$\bar{y}$ = Mittelwert der Variable y
b_1 = Steigung der Regressionsgrade
$\bar{x}$ = Mittelwert der Variable x

Standardisierter Regressionskoeffizient (S. 262)

$$\beta_1 = b_1 \cdot \frac{s_x}{s_y}$$

b_1 = unstandardisierter Regressionskoeffizient
$s_{x/y}$ = Standardabweichung der Variablen x/y

Bestimmtheitsmaß R-Quadrat (S. 263)

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

$s_{\hat{y}}^2$ = Varianz der vorhergesagten Werte
s_y^2 = Varianz der beobachteten Werte

Korrigiertes R-Quadrat (S. 270)

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} \cdot (1 - R^2)$$

n = Anzahl der Fälle
p = Anzahl der Prädiktoren
R^2 = unkorrigiertes Bestimmtheitsmaß R^2

Logistische Regressionsfunktion (S. 275)

Logistische Regressionsfunktion $\hat{p}_i = \frac{e^{a_i}}{1 + e^{a_i}}$

$\hat{p}_i$ = geschätzte Wahrscheinlichkeit, dass die vorhergesagte Variable den Wert 1 hat
e = Eulersche Zahl (2,71828)
a_i = Linearkombination der Prädiktoren

Odds-Ratio für ein Regressionsgewicht (S. 277)

Odds-Ratio für $b_i = \mathrm{Exp}(b_i) = e^{b_i}$

b_i = i-tes Regressionsgewicht
e = Eulersche Zahl (2,71828)