Udo Kuckartz, Stefan Rädiker, Thomas Ebert und Julia Schehl

Statistik
Eine verständliche Einführung

VS Verlag für Sozialwissenschaften, 274 Seiten, 2010

Formelsammlung

korrigierte Fassung vom 01.02.2011
Korrekturen gegenüber dem Buch sind rot markiert

magma Marburger Arbeitsgruppe für Methoden & Evaluation statistik-verstaendlich@methoden-evaluation.de

Kapitel 2 - Häufigkeitsverteilungen und ihre grafischen Darstellungen

Berechnung von Prozenten (S. 35)

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

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

Berechnung von gültigen Prozenten (S. 36)

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

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

Berechnung von kumulierten Prozenten (S. 36)

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

$f_{kum}(k)$ = die aufsummierten absoluten Häufigkeiten bis einschließlich zur Kategorie k
n = Anzahl der Fälle

Kapitel 3 - Mittelwerte und Streuungsmaße

Mittelwert (S. 60)

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

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

Mittelwert für eine Häufigkeitstabelle (S. 62)

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

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

Mittelwert für gruppierte Daten (S. 63)

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

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

Spannweite (S. 65)

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

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

Interquartilsabstand (S.67)

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

$Q_{.75}$ = obere 25% der Werte einer Verteilung
$Q_{.25}$ = untere 25% der Werte einer Verteilung

Varianz (S.67)

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

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

Varianz (Inferenzstatistik) (S. 67)

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

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

Standardabweichung (S. 68)

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

s^2 = Varianz

Variationskoeffizient (S. 72)

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

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

z-Tranzformation (S. 73)

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

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

Kapitel 4 - Kreuztabelle, Chi-Quadrat und Zusammenhangsmaße

Berechnung der erwarteten Häufigkeit (S. 87)

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

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. 90)

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

k = Anzahl der Zellen
f_b = beobachtete Häufigkeiten
f_e = erwartete Häufigkeiten

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

Phi-Koeffizient (S. 91)

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

a = erste Zelle einer Vier-Felder-Tafel
b = zweite Zelle einer Vier-Felder-Tafel
c = dritte Zelle einer Vier-Felder-Tafel
d = vierte Zelle einer Vier-Felder-Tafel

Beziehung zwischen Phi und Chi-Quadrat (S. 92)

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

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

Kontigenzkoeffizient C (S. 92)

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

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

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

R = kleinste Anzahl der Kategorien der Zeilen bzw. Spaltenvariable

Cramers V (S. 93)

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

χ^2 = Chi-Quadrat
n = Anzahl aller Fälle
R = kleinste Anzahl der Kategorien der Zeilen bzw. Spaltenvariable

Berechnung von Chi-Quadrat für univariate Verteilungen (S. 94)

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

$f_{b(j)}$ = 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. 107)

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

Definition der empirischen Wahrscheinlichkeit (S. 109)

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

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

Binomialgleichung (S. 116)

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

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. 118)

$$\textit{Mittelwert der Verteilung } (\mu) = n \cdot p$$

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

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. 133)

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

σ^2 = Varianz der Werte in der Stichprobe
n = Anzahl der Fälle

Konfidenzintervall (S. 134)

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

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

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

Kapitel 7 - t-Test: zwei Mittelwerte vergleichen

Berechnung der Prüfgröße t für homogene Varianzen (S. 150)

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

$\bar{x}_1$ = 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

$$\text{Freiheitsgrade } df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$$

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

Berechnung der Prüfgröße t für heterogene Varianzen (S. 152)

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

$\bar{x}_1$ = 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. 150)

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

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

Berechnung der Effektstärke einer Mittelwertdifferenz (S. 153-154)

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

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

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

$\bar{x}_1$ = 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. 156)

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

$\bar{x}_D$ = Mittelwert der Differenzen aller Wertepaare
$\hat{\sigma}_D$ = 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 $df = n - 1$

Kapitel 8 - Varianzanalyse: mehr als zwei Mittelwerte vergleichen

Gesamtmittelwert (S. 173)

$$\bar{G} = \frac{G}{m \cdot k}$$

m = Anzahl Personen in einer Gruppe
G = Gesamtsumme aller beobachteten Werte
k = Anzahl der Faktorstufen

Treatment-Quadratsumme (S. 173)

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

m = Anzahl Personen in einer Gruppe
$\bar{A}_j$ = Gruppenmittelwert unter der Faktorstufe i
$\bar{G}$ = Gesamtmittelwert

Treatmentvarianz (S. 173)

$$\hat{\sigma}^2_{treat} = \frac{QS_{treat}}{df}$$

QS_{treat} = Treatment-Quadratsumme
df = Freiheitsgrade $df = k - 1$

Fehler-Quadratsumme (S. 174)

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

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

Fehler-Varianz (S. 175)

$$\hat{\sigma}^2_{Fehler} = \frac{QS_{Fehler}}{df}; \text{ Freiheitsgrade } df = k \cdot (m-1)$$

QS_{Fehler} = Fehler-Quadratsumme

Prüfgröße F (S. 176)

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

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

Effektstärke eta-Quadrat (S. 177)

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

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

Kapitel 9 - Korrelation: Zusammenhänge identifizieren

Kovarianz (S. 193)

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

x_i und y_i = Wert einer Person 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. 194)

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

$cov(x,y)$ = 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. 196)

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

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

Spearmans rho (S. 199)

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

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

Prüfgröße t für Spearmans rho (S. 200)

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

r_s = Spearmans rho
n = Anzahl der Untersuchungseinheiten

Kapitel 10 - Skalenbildung

Cronbachs-Alpha (S. 223)

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

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. 235)

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

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

Achsenabschnitt der Regressionsgeraden (S. 236)

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

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

Standardisierter Regressionskoeffizient (S. 236)

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

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

Bestimmtheitsmaß R-Quadrat (S. 237)

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

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

Korrigiertes R-Quadrat (S. 242)

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

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