Formelsammlung

Statistische Formeln, die im Buch angesprochen wurden

1. Symbole

Symbol	Bedeutung
a	Ordinatenabschnitt (bivariate lineare Regression)
b	Steigung (bivariate lineare Regression)
b_{ij}	Beobachtete Häufigkeit in Spalte i und Zeile j einer Kreuztabelle
C	Kontingenzkoeffizient von Pearson
DK	Determinationskoeffizient
e_{ij}	Bei Unabhängigkeit erwartete Häufigkeit in Spalte i und Zeile j einer Kreuztabelle
f_i	Häufigkeit für Klasse i in einer Häufigkeitsverteilung
f(x)	Dichtefunktion (Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen)
i	Laufindex
j	Laufindex
k	Anzahl der Klassen (und Wert der Standardnormalvariablen K)
K	Standardnormalvariable
n	Anzahl der Beobachtungen (Stichprobenumfang)
P	Wahrscheinlichkeit
Φ	Vier-Felder-Koeffizient
r	Korrelationskoeffizient von Bravais/Pearson (und Zeilenzahl in Kreuztabellen)
ρ	Rangkorrelationskoeffizient von Spearman
R_x	Rangplatz eines Wertes der Variablen X
R_y	Rangplatz eines Wertes der Variablen Y
s	Standardabweichung (und Spaltenzahl in Kontingenztabellen)
u	Wert einer Variablen, die der Chi-Quadrat-Verteilung folgt
VAR	Varianz
x	Wert der stetigen Variablen X
$\bar{x}$	Arithmetisches Mittel
$\bar{x}_h$	Modus (häufigster Wert)
$\bar{x}_z$	Median (Zentralwert)
x_i	Merkmalswert Nr. i (Wert der diskreten Variablen X)
X	Variable
y_i	Merkmalswert Nr. i (Wert der diskreten Variablen Y)
Y	Variable
z	Zeilenzahl in Kontingenztabellen

2. Mittelwerte

ungewogenes arithmetisches Mittel	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
Voraussetzung	metrische Daten
angesprochen in Kapitel	12.1

gewogenes arithmetisches Mittel	$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{\sum_{i=1}^{k} f_i}$
Voraussetzung	metrische Daten
angesprochen in Kapitel	19.1

Median	Merkmalswert, der die der Größe nach sortierte Reihe der Merkmalswerte halbiert
Voraussetzung	ordinalskalierte Daten
angesprochen in Kapitel	12.1

Modus	Häufigster Wert
Voraussetzung	nominalskalierte Daten
angesprochen in Kapitel	12.1

3. Streuungsmaße

Varianz	$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
Voraussetzung	metrische Daten
angesprochen in Kapitel	12.1

Standardabweichung	s
Voraussetzung	metrische Daten
angesprochen in Kapitel	12.1

4. Bivariate lineare Regression

Steigung	$b = \dfrac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2}$
Ordinatenabschnitt	$a = \dfrac{\sum_{i=1}^{n} y_i}{n} - b \cdot \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{y} - b \cdot \bar{x}$
Voraussetzung	metrische Daten
angesprochen in Kapitel	16.4

5. Bivariate Korrelation

Korrelationskoeffizient von Bravais/Pearson	$r = \dfrac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[n \cdot \sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2 \right] \cdot \left[n \cdot \sum_{i=1}^{n} y_i^{\,2} - (\sum_{i=1}^{n} y_i)^2 \right]}}$ $-1 \leq r \leq 1$
Determinationskoeffizient	$DK = r^2$
Voraussetzung	metrische Daten
angesprochen in Kapitel	17.2

Rangkorrelationskoeffizient von Spearman	$\rho = 1 - \dfrac{6 \cdot \sum_{i=1}^{n} (R_x - R_y)^2}{n^3 - n}$ $\qquad -1 \leq \rho \leq 1$
Voraussetzung	ordinalskalierte Daten
angesprochen in Kapitel	17.3

Kontingenzkoeffizient von Pearson	$C=\sqrt{\frac{u}{u+n}}$ $\quad u=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(b_{ij}-e_{ij})^2}{e_{ij}}$ $\quad 0\le C<1$ $e_{ij}=\frac{b_{i*}b_{*j}}{n}$ $\quad b_{i*}=\sum_{j=1}^{s}b_{ij}$ $\quad b_{*j}=\sum_{i=1}^{r}b_{ij}$
Maximalwert von C	$C_{\max}=0{,}5\cdot\left(\sqrt{\frac{z-1}{z}}+\sqrt{\frac{s-1}{s}}\right)$
Voraussetzung	nominalskalierte Daten
angesprochen in Kapitel	17.4

Vierfelderkoeffizient	$\Phi=\sqrt{\frac{u}{n}}=\frac{b_{11}b_{22}-b_{12}b_{21}}{\sqrt{b_{1*}b_{2*}b_{*1}b_{*2}}}$ $\quad -1\le\Phi\le 1$
Voraussetzung	nominalskalierte Daten
angesprochen in Kapitel	17.4

6. Multiple Regression und Korrelation (Drei-Variablen-Fall)

Lineare Regressionsfläche	Abhängige Variable : Y Unabhängige Variablen : X_1 und X_2
Ordinatenabschnitt	$b_0=\bar{y}-b_1\bar{x}_1-b_2\bar{x}_2$
Steigung in Richtung der X_1-Achse	$b_1=\frac{r_{Y1}-r_{Y2}r_{12}}{1-r_{12}{}^2}\cdot\frac{\sqrt{\operatorname{var}(y)}}{\sqrt{\operatorname{var}(x_1)}}$
Steigung in Richtung der X_2-Achse	$b_2=\frac{r_{Y2}-r_{Y1}r_{12}}{1-r_{12}{}^2}\cdot\frac{\sqrt{\operatorname{var}(y)}}{\sqrt{\operatorname{var}(x_2)}}$
Voraussetzung	metrische Daten
angesprochen in Kapitel	16.5

Multipler Determinations-koeffizient	$r_{Y.12}{}^2=\frac{r_{Y1}{}^2+r_{Y2}{}^2-2r_{Y1}r_{Y2}r_{12}}{1-r_{12}{}^2}$
Voraussetzung	metrische Daten
angesprochen in Kapitel	17.6

Partieller Korrelations-koeffizient	$r_{Y1.2}=\frac{r_{Y1}-r_{Y2}r_{12}}{\sqrt{1-r_{Y2}{}^2}\sqrt{1-r_{12}{}^2}}$
Voraussetzung	metrische Daten
angesprochen in Kapitel	17.5

7. Wahrscheinlichkeitsverteilungen

Binomialverteilung	$P(X = x) = P(x) = \binom{n}{x}\pi^{x}(1-\pi)^{n-x}$
Voraussetzung	Bernoulli'sche Versuchsanordnung: n = voneinander unabhängige Versuche mit jeweils nur zwei Ergebnismöglichkeiten x = Anzahl der im Sinne der Fragestellung günstigen Ergebnisse π = Eintrittswahrscheinlichkeit des günstigen Ergebnisses bei einem Versuch
angesprochen in Kapitel	14.1

Gauß'sche Normalverteilung	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Voraussetzung	μ = Mittelwert der Verteilung σ = Standardabweichung der Verteilung π = Kreisparameter e = Euler'sche Zahl (Wachstumskonstante)
Standardnormalverteilung	$f(k) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}k^2}$ $\qquad k = \frac{x-\mu}{\sigma}$
angesprochen in Kapitel	15.1

Chi-Quadrat-Verteilung	$f(\chi^2) = \frac{0{,}5^{0{,}5\nu}}{\Gamma(0{,}5\nu)} \cdot (\chi^2)^{0{,}5\nu-1} \cdot e^{-0{,}5\chi^2}$ $\chi^2 = \sum_{i=1}^{\upsilon} K_i^{\,2}$ $\qquad \Gamma(\lambda) = \int_0^{\infty} e^{-x} \cdot x^{\lambda-1}dx$
Voraussetzung	χ^2 = Chi-Quadrat-Variable ν = Anzahl der Freiheitsgrade e = Euler'sche Zahl (Wachstumskonstante) Γ = Gamma-Funktion (siehe dritte Gleichung) λ = konstante Zahl größer null K_i = i-te Standardnormalvariable
angesprochen in Kapitel	17.4

t-Verteilung	$f(t) = \frac{\Gamma(0.5(\nu+1))}{\Gamma(0,5\nu)\sqrt{\nu\pi}}\left(1+\frac{t^2}{\nu}\right)^{-0.5(\nu+1)}$
Voraussetzung	t = t-verteilte Variable Andere Symbolik wie oben
angesprochen in Kapitel	Nicht angesprochen, aber von SPSS häufig verwendet (siehe z.B. 15.2 und 15.3)