Lösung zur Selbstkontrollarbeit 2

Vertiefung der Wirtschaftsmathematik und Statistik (Teil Statistik)

13. Januar 2020

Aufgaben

Aufgabe 1

$$
\begin{aligned}
H_0: &\quad \pi_{ij} = \pi_{i.} \cdot \pi_{.j} \text{ für alle } i = 1,2,\ j = 1,2,3 \\
H_1: &\quad \pi_{ij} \neq \pi_{i.} \cdot \pi_{.j} \text{ für mindestens ein } i, j
\end{aligned}
$$

Für den χ^2-Unabhängigkeitstest wird $\chi^2 = \sum_{i,j=1}^{I,J} \frac{(n_{ij} - N\hat{\pi}_{ij})^2}{N\hat{\pi}_{ij}}$ als Testgröße verwendet mit $N\hat{\pi}_{ij} = \frac{n_{i.} n_{.j}}{N}$ $(N = 100)$.

	n_{i1}	$N\hat{\pi}_{i1}$	n_{i2}	$N\hat{\pi}_{i2}$	n_{i3}	$N\hat{\pi}_{i3}$
n_{1j}	10	*10*	20	*15*	20	*25*
n_{2j}	10	*10*	10	*15*	30	*25*

$$
\chi^2 \quad = \quad \frac{25}{15} + \frac{25}{25} + \frac{25}{15} + \frac{25}{25} = 5\frac{1}{3}.
$$

Zum Signifikanzniveau $\alpha = 0.1$ ergibt sich der obere kritische Wert zu $c_o = \chi^2(1-\alpha, (I-1)(J-1)) = \chi^2(0.9, 2) = 4.605$. Mit $\chi^2 = 5.333 > 4.605$ kann die Nullhypothese abgelehnt und somit auf eine Abhängigkeit geschlossen werden.

Aufgabe 2

Für den χ^2-Unabhängigkeitstest wird $\chi^2 = \sum_{i,j=1}^{I,J} \frac{(n_{ij} - N\hat{\pi}_{ij})^2}{N\hat{\pi}_{ij}}$ als Testgröße verwendet mit $N\hat{\pi}_{ij} = \frac{n_{i.} n_{.j}}{N}$ $(N = 150)$. Die 3. und 4. Zeile werden zusammengefasst, da sonst die Bedingung $N\hat{\pi}_{ij} \geq 1, N\hat{\pi}_{ij} \geq 5$ für mindestens 80% der Klassen nicht erfüllt ist $(N\hat{\pi}_{41} = 0.333, N\hat{\pi}_{42} = 0.667)$. Es ergeben sich somit $(3-1)(2-1) = 2$ Freiheitsgrade.

$$
\begin{aligned}
H_0: &\quad \pi_{ij} = \pi_{i.} \cdot \pi_{.j} \text{ für alle } i = 1,2,3,\ j = 1,2 \\
H_1: &\quad \pi_{ij} \neq \pi_{i.} \cdot \pi_{.j} \text{ für mindestens ein } i, j
\end{aligned}
$$

	n_{i1}	$N\hat{\pi}_{i1}$	n_{i2}	$N\hat{\pi}_{i2}$
n_{1j}	15	*25*	60	*50*
n_{2j}	15	*15*	30	*30*
n_{3j}	20	*10*	10	*20*

$$\chi^2 = \frac{100}{25} + \frac{100}{50} + \frac{100}{10} + \frac{100}{20} = 4 + 2 + 10 + 5 = 21$$

Zum Signifikanzniveau $\alpha = 0.01$ ergibt sich der obere kritische Wert c_o zu $\chi^2(1-\alpha,(I-1)(J-1)) = \chi^2(0.99,2) = 9.21$ Wegen $\chi^2 = 21 > 9.21$ wird die Nullhypothese (Unabhängigkeit der Merkmale „Parteizugehörigkeit“ und „Posten“) abgelehnt.

Aufgabe 3

3.1

Für das obige Modell gilt:

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\ \hat{\beta} &= \frac{\sum_{n=1}^{N} X_n Y_n - N\bar{X}\bar{Y}}{\sum_{n=1}^{N} X_n^2 - N\bar{X}^2}\end{aligned}$$

$$\begin{aligned}\hat{\beta} &= \frac{20 - 10 \cdot 4 \cdot 0.8}{200 - 160} = -0.3 \\ \hat{\alpha} &= 0.8 + 0.3 \cdot 4 = 2\end{aligned}$$

3.2

Mittels der Streuungszerlegung ergibt sich

$$\widehat{\sigma^2} = \frac{1}{N-2}\sum_{n=1}^{N}\hat{\epsilon}_n^2 = \frac{1}{N-2}SQR$$

zu

$$\widehat{\sigma^2} = \frac{1}{N-2}(SQT - SQE)$$

mit

$$\begin{aligned}SQT &= \sum_{n=1}^{N}(Y_n - \bar{Y})^2 = \sum_{n=1}^{N} Y_n^2 - N\bar{Y}^2 \\ SQE &= \hat{\beta}^2\sum_{n=1}^{N}(X_n - \bar{X})^2 = \hat{\beta}^2(\sum_{n=1}^{N} X_n^2 - N\bar{X}^2)\end{aligned}$$

Obige Daten eingesetzt ergibt

$$\begin{aligned} SQT &= 26 - 6.4 = 19.6 \\ SQE &= 0.09 \cdot (200 - 160) = 3.6 \\ \widehat{\sigma^2} &= \frac{1}{8}(19.6 - 3.6) = 2 \end{aligned}$$

3.3

Gesucht ist das Bestimmtheitsmaß in der Form

$$R^2_{XY} = PRE = \frac{SQE}{SQT}$$

Es ergibt sich der Wert $r^2_{xy} = 0.1837$. Das Bestimmtheitsmaß gibt das Verhältnis von erklärter zu totaler Streuung an. Im vorliegenden Modell werden lediglich 18.37% der Streuung von Y durch X erklärt, d.h. die Residualstreuung ist hier recht hoch. Es liegen viele nicht-erklärbare Einflüsse vor.

3.4

Für die Hypothese $H_0 : \beta = 0$ wird die Testgröße

$$T_\beta = \frac{\hat{\beta} - 0}{\widehat{\sigma_\beta}}$$

verwendet, welche t-verteilt ist mit $(N - 2)$ Freiheitsgraden.

$$\widehat{\sigma_\beta} = \frac{\hat{\sigma}}{\sqrt{\sum_{n=1}^{N} X_n^2 - N\bar{X}^2}} = \frac{\hat{\sigma}}{\sqrt{200 - 10 \cdot 16}} = \sqrt{\frac{2}{40}} = 0.224$$

Es ergibt sich $t = -0.3/\sqrt{2/40} = -1.342$ und $t(1 - \alpha/2, 8) = t(0.975, 8) = 2.306$. Somit muss die Nullhypothese beibehalten werden.

3.5

Für die einseitigen Konfidenzintervalle wird das Quantil $t(1-\alpha, 8) = t(0.95, 8) = 1.860$ und für das zweiseitige das Quantil $t(1 - \alpha/2, 8) = t(0.975, 8) = 2.306$ verwendet. Die gesuchten Konfidenzintervalle lauten:

$$[\hat{\beta} - t(1 - \alpha/2, 8)\widehat{\sigma_\beta}, \hat{\beta} + t(1 - \alpha/2, 8)\widehat{\sigma_\beta}] = [-0.816, 0.216]$$

$$[\hat{\beta} - t(1-\alpha, 8)\widehat{\sigma_\beta}, \infty) = [-0.716, \infty)$$
$$(-\infty, \hat{\beta} + t(1-\alpha, 8)\widehat{\sigma_\beta}] = (-\infty, 0.116]$$

Auch anhand des zweiseitigen Konfidenzintervalls kann die Hypothese $H_0 : \beta = 0$ überprüft werden. Da das zweiseitige Konfidenzintervall den Wert $\beta = 0$ überdeckt, kann die Hypothese nicht abgelehnt werden.

3.6

Das Konfidenzintervall für die Regressionsgerade wird mit

$$\hat{E}[Y|X] \pm t(1-\alpha/2, N-2)\sqrt{\widehat{\mathrm{Var}}(\hat{E})}$$

angegeben, wobei

$$\sqrt{\widehat{\mathrm{Var}}(\hat{E})} = \hat{\sigma}\sqrt{\frac{1}{N} + \frac{(X-\bar{X})^2}{\sum_n X_n^2 - N\bar{X}^2}}$$

gilt.

$$\sqrt{\widehat{\mathrm{Var}}(\hat{E})} = \sqrt{\frac{1}{10} + \frac{(x-4)^2}{40}} = \sqrt{0.1 + 0.025(x-4)^2}$$

Mit $t(0.99, 8) = 2.896$ lautet das zweiseitige 98%-Konfidenzintervall

$$(2 - 0.3x) \pm 2.896\sqrt{0.1 + 0.025(x-4)^2}$$

An der Stelle $\bar{x} = 4$ befindet sich der minimale Wert

$$0.8 \pm 0.916$$

3.7

Zugrundegelegt wird die Testgröße

$$F = \frac{SQE}{SQR/(N-2)}$$

welche F-verteilt ist mit $(1, N-2)$ Freiheitsgraden.
Mit den oben berechneten Werten ergibt sich $F = 3.6/2 = 1.8$. Das Quantil lautet $f(0.99, 1, 8) = 11.3$. Da $F < f$ gilt, muss die Hypothese beibehalten werden.

Aufgabe 4

4.1

Allgemein gilt:

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\ \hat{\beta} &= \frac{\sum_{n=1}^{N} X_n Y_n - N\bar{X}\bar{Y}}{\sum_{n=1}^{N} X_n^2 - N\bar{X}^2}\end{aligned}$$

Für das hier angegebene Modell gilt:

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{X} \\ \hat{\beta} &= \frac{\sum_{n=1}^{N_1} Y_n - N\frac{N_1}{N}\bar{Y}}{N_1 - N\frac{N_1^2}{N^2}}\end{aligned}$$

Nach Umformungen ergeben sich die Schätzer

$$\begin{aligned}\hat{\alpha} &= \frac{1}{N - N_1} \sum_{n=N_1+1}^{N} Y_n \\ \hat{\beta} &= \frac{1}{N_1} \sum_{n=1}^{N_1} Y_n - \frac{1}{N - N_1} \sum_{n=N_1+1}^{N} Y_n\end{aligned}$$

In dem vorliegenden Modell gilt

$$E(Y_n) = \begin{cases} \alpha + \beta & \text{für } n = 1, ...N_1, \\ \alpha & \text{für } n = N_1 + 1, ..., N. \end{cases}$$

Daraus folgt

$$\begin{aligned}E(\hat{\alpha}) &= \frac{1}{N - N_1} \sum_{n=N_1+1}^{N} E(Y_n) = \frac{1}{N - N_1}(N - N_1)\alpha = \alpha \\ E(\hat{\beta}) &= \frac{1}{N_1} \sum_{n=1}^{N_1} E(Y_n) - \frac{1}{N - N_1} \sum_{n=N_1+1}^{N} E(Y_n) \\ &= \frac{1}{N_1} N_1(\alpha + \beta) - \frac{1}{N - N_1}(N - N_1)\alpha = \beta\end{aligned}$$

4.2

Allgemein gilt für KQ-Schätzer

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &= \sigma^2\left(\frac{1}{N}+\frac{\bar{X}^2}{\sum_{n=1}^{N}(X_n-\bar{X})^2}\right)=\sigma^2\left(\frac{\sum_{n=1}^{N}X_n^2}{N\sum_{n=1}^{N}(X_n-\bar{X})^2}\right)\\
\mathrm{Var}(\hat{\beta}) &= \frac{\sigma^2}{\sum_{n=1}^{N}(X_n-\bar{X})^2}
\end{aligned}$$

Somit gilt speziell hier

$$\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &= \sigma^2\left(\frac{\sum_{n=1}^{N}X_n^2}{N\sum_{n=1}^{N}X_n^2-N^2\bar{X}^2}\right)\\
&= \sigma^2\frac{N_1}{N\cdot N_1-N^2\frac{N_1^2}{N^2}}=\frac{\sigma^2}{N-N_1}\\
\mathrm{Var}(\hat{\beta}) &= \frac{\sigma^2}{\sum_{n=1}^{N}X_n^2-N\bar{X}^2}=\frac{\sigma^2}{N_1-N\frac{N_1^2}{N^2}}=\frac{\sigma^2}{N_1(1-\frac{N_1}{N})}
\end{aligned}$$

4.3

Für die Hypothese $H_0 : \alpha = 0$ wird die Testgröße

$$T_\alpha = \frac{\hat{\alpha}-0}{\widehat{\sigma_\alpha}}$$

verwendet, welche t-verteilt ist mit $(N-2)$ Freiheitsgraden. Speziell ist hier

$$T_\alpha = \frac{\frac{1}{N-N_1}\sum_{n=N_1+1}^{N}Y_n}{\frac{\hat{\sigma}}{\sqrt{N-N_1}}}=\frac{1}{\hat{\sigma}\sqrt{N-N_1}}\sum_{n=N_1+1}^{N}Y_n.$$

Aufgabe 5

5.1

Mittels der Streuungszerlegung $SQT = SQR + SQE$ reicht es aus, zwei Quadratsummen zu berechnen. Es gilt $\sum_{ij} y_{ij} = 6247$ und $\sum_{ij} y_{ij}^2 = 963329$.

$$\begin{aligned}
SQE &= \sum_{ij}(\bar{Y}_i-\bar{Y})^2=J\sum_{i}\bar{Y}_i^2-IJ\bar{Y}^2\\
SQT &= \sum_{ij}(Y_{ij}-\bar{Y})^2=\sum_{ij}Y_{ij}^2-IJ\bar{Y}^2
\end{aligned}$$

i	$\sum_j y_{ij}^2$	$\bar{y}_i$	$\bar{y}_i^2$
1	71155	107.5	11556.25
2	124823	142.83	20401.36
3	123155	142.5	20306.25
4	151351	158.17	25016.69
5	218130	189	35721
6	122904	142.67	20353.78
7	151811	158.5	25122.25
$\sum$	963329	1041.17	158477.58

$$
\begin{aligned}
SQE &= 6 \cdot 158477.58 - 42 \cdot 148.74^2 = 950865.498 - 929166.881 = 21698.6 \\
SQT &= 963329 - 42 \cdot 148.74^2 = 34162.12 \\
SQR &= 34162.12 - 21698.6 = 12463.5 \\
F &= 3616.43/356.1 = 10.156
\end{aligned}
$$

Somit lautet die Tabelle:

SQ	Wert	df	F-Statistik
SQE (zwischen)	21698.6	6	10.156
SQR (innerhalb)	12463.5	35	
SQT (total)	34162.12	41	

Das Quantil $F(0.99, 6, 35)$ nimmt den Wert 3.37 an, so dass die Nullhypothese, die Düngemittel wirken gleich, abgelehnt wird.

5.2

ONEWAY ANOVA

Hoehe

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	21698,619	6	3616,437	10,156	,000
Innerhalb der Gruppen	12463,500	35	356,100		
Gesamt	34162,119	41			

5.3

Das Modell lautet in der Effektdarstellung

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

mit der Restriktion $\sum_i \alpha_i = 0$. Die Schätzung der Effekte erfolgt über

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{Y}_i - \bar{Y},$$

wobei die obige Restriktion auch für die Schätzer gilt.

$$\hat{\alpha}_i = (-41.24, -5.91, -6.24, 9.43, 40.26, -6.07, 9.76)$$

Es gilt $\sum \hat{\alpha}_i = -41.24 - 5.91 - 6.24 + 9.43 + 40.26 - 6.07 + 9.76 = -0.01$. Somit ist die Restriktion bis auf Rundungsfehler erfüllt.

Aufgabe 6

Richtig ist nur die Aussage „Die Behauptung ist wahr. Es liegt eine Tautologie vor.“.

Aufgabe 7

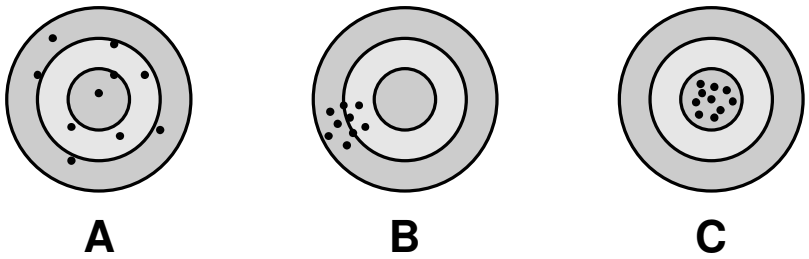
Es liegt eine Implikation der Form „Aus A folgt B“ vor, welche nur widerlegt ist, wenn B nicht folgt. Somit widerlegt folgende Aussage die Hypothese:
„Die Kinder spielen viel am Computer und die Aggressivität erhöht sich nicht (Formal: Das Verhalten X_1 wird verstärkt und die Auftretenswahrscheinlichkeit des Ereignisses Y_1 erhöht sich nicht).“

Aufgabe 8

8.1

Die Objektivität betrachtet die Unabhängigkeit einer Aussage bzw. Theorie von der Person des Beobachters und die Reliabilität untersucht die Zuverlässigkeit bzw. Reproduzierbarkeit.

8.2



A: Es liegt weder Reliabilität noch Validität vor.

B: Es liegt Reliabilität aber keine Validität vor.

C: Es liegt Reliabilität und Validität vor.

Aufgabe 9

9.1

$$\begin{aligned}\rho^2(X,T) &= \frac{\mathrm{Cov}^2(X,T)}{\mathrm{Var}(X)\cdot\mathrm{Var}(T)} = \frac{\mathrm{Cov}^2(T+\epsilon,T)}{\mathrm{Var}(X)\cdot\mathrm{Var}(T)} \\ &= \frac{\mathrm{Cov}^2(T,T)}{\mathrm{Var}(X)\cdot\mathrm{Var}(T)} = \frac{\mathrm{Var}^2(T)}{\mathrm{Var}(X)\cdot\mathrm{Var}(T)} = \frac{\mathrm{Var}(T)}{\mathrm{Var}(X)} = r\end{aligned}$$

9.2

$$\rho(X,X') = \frac{\mathrm{Cov}(X,X')}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(X')}} = \frac{\mathrm{Cov}(T,T)}{\mathrm{Var}(X)} = \frac{\mathrm{Var}(T)}{\mathrm{Var}(X)} = r$$

Aufgabe 10

10.1

$$\begin{aligned}G(-) &= \frac{12\cdot 10 + 24\cdot 20 + 35\cdot 40 + 24\cdot 25 + 5\cdot 5)}{10000} = 0.263 \\ G(+) &= \frac{7+10+25+15+3}{100} = 0.6 \\ \kappa &= \frac{0.6 - 0.2625}{0.7375} = 0.458\end{aligned}$$

10.2

Der Prozentsatz der Fehler entspricht $F(+)$ und der Prozentsatz der Fehler bei Unabhängigkeit der Merkmale entspricht $F(-)$.

$$\begin{aligned}F(+) = 1 - G(+) &= 1 - 0.6 = 0.4 \\ F(-) = 1 - G(-) &= 1 - 0.263 = 0.737\end{aligned}$$

Aufgabe 11

Es gilt $G(+) = 0.1 + 0.25 + 0.23 + 0.2 = 0.78$ und $G(-) = 0.03 + 0.09 + 0,15 + 0.125 = 0.395$. Somit ist

$$\kappa = \frac{G(+) - G(-)}{1 - G(-)} = 0.\overline{63}$$

Aufgabe 12

12.1

Es gilt $p < \alpha/k$, d.h. die maximale Überschreitungswahrscheinlichkeit liegt unter α/k.

12.2

Mit α/k als nach Bonferroni korrigiertes Signifikanzniveau ergibt sich das simultane Signifikanzneau

$$\alpha^* = 1 - (1 - \alpha/k)^k = 1 - (1 - 0.1/k)^k.$$

$$\begin{aligned} k = \ \ 2 \quad & \alpha^* = 0.0975 \\ k = \ \ 5 \quad & \alpha^* = 0.09608 \\ k = 10 \quad & \alpha^* = 0.09562 \end{aligned}$$

Aufgabe 13

13.1

$$S = \frac{1}{N-1}\big((x_1 - \bar{x})(x_1 - \bar{x})' + \cdots (x_N - \bar{x})(x_N - \bar{x})'\big)$$

mit

$$(x_n - \bar{x}) = \begin{pmatrix} x_{n1} - \bar{x}_1 \\ \vdots \\ x_{np} - \bar{x}_p \end{pmatrix}$$

für $n = 1, ..., N$. Es folgt $(x_1 - \bar{x})(x_1 - \bar{x})' =$

$$\begin{pmatrix} x_{11}-\bar{x}_1 \\ \vdots \\ x_{1p}-\bar{x}_p \end{pmatrix} \cdot \big((x_{11}-\bar{x}_1),\cdots,(x_{1p}-\bar{x}_p)\big)$$

$$= \begin{pmatrix} (x_{11}-\bar{x}_1)^2 & (x_{11}-\bar{x}_1)(x_{12}-\bar{x}_2) & \cdots & (x_{11}-\bar{x}_1)(x_{1p}-\bar{x}_p) \\ (x_{12}-\bar{x}_2)(x_{11}-\bar{x}_1) & (x_{12}-\bar{x}_2)^2 & \cdots & (x_{12}-\bar{x}_2)(x_{1p}-\bar{x}_p) \\ \vdots & & \vdots & \vdots \\ (x_{1p}-\bar{x}_p)(x_{11}-\bar{x}_1) & (x_{1p}-\bar{x}_p)(x_{12}-\bar{x}_2) & \cdots & (x_{1p}-\bar{x}_p)^2 \end{pmatrix}$$

Somit ist $S = \frac{1}{N-1}\big((x_1-\bar{x})(x_1-\bar{x})' + \cdots (x_N-\bar{x})(x_N-\bar{x})'\big) =$

$$\begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ \vdots & & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix}$$

Aufgabe 14

Gruppenstatistiken

	GESCHLECHT	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
LoyalU	weiblich	204	,0759	1,04077	,07287
	männlich	162	-,0986	,94284	,07408

Modellzusammenfassung

Abhängige Variable:LoyalU

Gleichung	Modellzusammenfassung				
	R-Quadrat	F	Freiheitsgrade 1	Freiheitsgrade 2	Sig.
Linear		2,759	1	364	,098

Die unabhängige Variable ist GESCHLECHT.

14.1

$$
\begin{aligned}
p &= 204/366 = 0.5574 \\
s_{*x} &= \sqrt{p(1-p)} = 0.4967 \\
\bar{y}_0 &= \hat{\alpha} = -0.0986 \\
\bar{y}_1 &= 0.0759 \\
\hat{\beta} &= \bar{y}_1 - \bar{y}_0 = 0.1745 \\
s_0 &= 0.94284 \\
s_1 &= 1.04077 \\
s_{*y} &\approx \sqrt{s_0^2(1-p) + s_1^2 p} = 0.9986 \\
r &= \hat{\beta} s_{*x}/s_{*y} = 0.0868 \\
r^2 &= 0.0075
\end{aligned}
$$

14.2

Betrachtet wird der t-Test im Zweistichprobenfall mit $N = N_0$ und $M = N_1$ und $X = X_w$ und $Y = Y_m$, d.h. die erste Stichprobe berücksichtigt die Werte GESCHLECHT=0 und die zweite die Werte GESCHLECHT=1. Da $N_0, N_1 > 30$ gilt, ist die Prüfgröße $T = \frac{\bar{X}-\bar{Y}-\delta_0}{S}$ approximativ standardnormalverteilt. Es ist $\delta_0 = 0$ und

$$
\begin{aligned}
S &= \sqrt{\left(\frac{1}{N_0} + \frac{1}{N_1}\right)\frac{(N_0-1)S_x + (N_1-1)S_y}{N_0+N_1-2}} \\
&= \sqrt{0.0110748(\frac{369.26232 + 151.79724}{364}} \\
&= \sqrt{0.0110748 \cdot 0.9974548} \\
&= 0.1051
\end{aligned}
$$

Die Hypothese $H_0 : \mu_0 = \mu_1$ wird zum Signifikanzniveau $\alpha = 0.05$ abgelehnt, wenn der Wert $\mu_0 - \mu_1$ nicht in dem Intervall

$$\left[\bar{X} - \bar{Y} - z \cdot S; \bar{X} - \bar{Y} + z \cdot S\right]$$

liegt mit $z = z(0.975) = 1.96$. Da $\mu_0 - \mu_1 = 0$ im Intervall $[0.1745 - 0.206; 0.1745 + 0.206] = [-0.032; 0.381]$ liegt, wird die Nullhypothese nicht abgelehnt.
Die Prüfgröße $T = \frac{\bar{X}-\bar{Y}}{S} = 0.1745/0.1051 = 1.66$ kann auch direkt aus dem F-Wert abgeleitet werden. Es gilt $F(1, 364) = 2.759 = t^2(364)$, so dass sich $t = \sqrt{2.759} = 1.661$ ergibt.

14.3

Im Falle einer dichotomen nominalskalierten Variablen X kann β als Mittelwertsunterschied interpretiert werden. Der Produkt-Moment-Korrelationskoeffizient zwischen stetigen und 0-1-Variablen r entspricht der standardisierten Mittelwertsdifferenz der Y-Variablen multipliziert mit der Standardabweichung der 0-1-Variablen.

Aufgabe 15

15.1

Symmetrische Zusammenhangsmaße deuten auf einen Zusammenhang hin, ohne jedoch einen Hinweis auf die Richtung der Beziehung zu geben. Es ist nicht zu erkennen, welche Variable die abhängige bzw. die unabhängige Variable ist, d.h. formal können beide Variablen ohne Ergebnisänderung in dem berechneten Maß vertauscht werden ($X \leftrightarrow Y$).

Asymmetrische Zusammenhangsmaße können eine gerichtete Kausalbeziehung der Form $X \rightarrow Y$ bzw. $Y \rightarrow X$ aufdecken. Hier muss formal zwischen abhängiger und unabhängiger Variable unterschieden werden. Eine Vertauschung beider Variablen ergibt für das berechnete Maß einen anderen Wert.

15.2

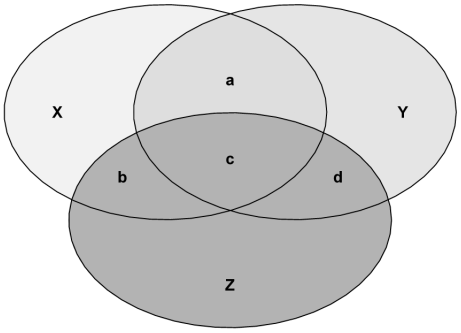
Von biserialer Korrelation wird gesprochen, wenn die Korrelation einer dichotomen bzw. dichotomisierten und einer intervallskalierten Variablen betrachtet wird.

15.3

Mittels der partiellen Korrelation wird der lineare Zusammenhang zweier Variablen betrachtet, wobei der Einfluss einer Drittvariablen eliminiert wird.

15.4

Gegeben ist das Venn-Diagramm, welches die Varianzen der drei Variablen X, Y und Z darstellt. Die Variable Z beeinflusst dabei den Zusammenhang zwischen X und Y. Mit den Bezeichnungen X, Y, Z ist hier die Gesamtfläche der entsprechenden Ellipse gemeint, während a, b, c, d Teilflächen bezeichnen.



Aufgrund der Streuungszerlegung gilt

$$S_y^2 = (1 - R_{xy}^2)S_y^2 + R_{xy}^2 S_y^2 = \text{residuale Varianz+erklärte Varianz.}$$

Die Fläche $a + c = R_{xy}^2 S_y^2$ entspricht der erklärten Varianz, d.h. dem Anteil der Varianz von Y, der durch X erklärt wird. Die Gesamtvarianz S_y^2 entspricht der gesamten Fläche Y, d.h. $S_y^2 = Y$. Für die bivariate Korrelation gilt somit $R_{xy} = \sqrt{(a+c)/Y}$.

Die partielle Korrelation entspricht dagegen der von Z bereinigten Korrelation, so dass $R_{xy.z} = \sqrt{a/(Y - c - d)}$ gilt.

Aufgabe 16

gemeinsame absolute Häufigkeiten: h_{ij}			
	keine Schulung	Schulung	$h_{i.}$
Unzufrieden	15	0	15
Zufrieden	5	30	35
$h_{.j}$	20	30	50

16.1

Zunächst wird ϕ berechnet, da dieser Koeffizient sehr leicht zu berechnen ist, und die anderen Koeffizienten daraus abgeleitet werden können.

$$\begin{aligned}
\phi &= \frac{ad - bc}{\sqrt{(c+d)(a+b)(b+d)(a+c)}} = \frac{15 \cdot 30 - 0}{\sqrt{35 \cdot 15 \cdot 30 \cdot 20}} = 0.802 \\
\chi^2 &= N\phi^2 = 32.143 \\
K &= \sqrt{\chi^2/(\chi^2 + N)} = \sqrt{32.143/(32.143 + 50)} = 0.626
\end{aligned}$$

16.2

$$
\begin{aligned}
\lambda(x \to y) &= \frac{45-30}{20} = 0.75 \\
\lambda(y \to x) &= \frac{45-35}{15} = 0.\overline{6} = 0.667 \\
\lambda_s &= \frac{45+45-30-35}{100-30-35} = \frac{25}{35} = 0.714
\end{aligned}
$$

$$
\begin{aligned}
\tau(x \to y) \quad : \quad & G(+) = \frac{225}{1000} + \frac{25}{1000} + \frac{900}{1500} = 0.85 \\
& G(-) = \frac{225}{2500} + \frac{1225}{2500} = 0.58 \\
\tau(x \to y) &= \frac{0.85-0.58}{1-0.58} = 0.643 \\
\tau(y \to x) \quad : \quad & G(+) = \frac{225}{750} + \frac{25}{1750} + \frac{900}{1750} = 0.8286 \\
& G(-) = \frac{400}{2500} + \frac{900}{2500} = 0.52 \\
\tau(y \to x) &= \frac{0.829-0.52}{1-0.52} = 0.643
\end{aligned}
$$

16.3

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	32,143[a]	1	,000		
Kontinuitätskorrektur[b]	28,671	1	,000		
Likelihood-Quotient	38,593	1	,000		
Exakter Test nach Fisher				,000	,000
Zusammenhang linear-mit-linear	31,500	1	,000		
Anzahl der gültigen Fälle	50				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6,00.

b. Wird nur für eine 2x2-Tabelle berechnet

Richtungsmaße

			Wert	Asymptotischer Standardfehler[a]
Nominal- bzgl. Nominalmaß	Lambda	Symmetrisch	,714	,126
		Zuf abhängig	,667	,172
		Schulung abhängig	,750	,097
	Goodman-und-Kruskal-Tau	Zuf abhängig	,643	,119
		Schulung abhängig	,643	,111

a. Die Null-Hyphothese wird nicht angenommen.

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,802	,000
	Cramer-V	,802	,000
	Kontingenzkoeffizient	,626	,000
Anzahl der gültigen Fälle		50	

Aufgabe 17

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,775[a]	,600	,590	7,988

a. Einflußvariablen : (Konstante), Größe

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	3644,374	1	3644,374	57,117	,000[a]
	Nicht standardisierte Residuen	2424,601	38	63,805		
	Gesamt	6068,975	39			

a. Einflußvariablen : (Konstante), Größe

b. Abhängige Variable: Gewicht

17.1

Mittels des globalen F-Tests wird die Hypothese „Es besteht kein linearer Zusammenhang“ überprüft.
Die Prüfgröße nimmt den Wert $\frac{SQE/1}{SQR/38} = \frac{MSE}{MSR} = \frac{3644.374}{63.805} = 57.117$ an. Die Hypothese wird hier aufgrund des geringen p-Wertes abgelehnt.

17.2

$R^2 = \frac{SSE}{SST} = \frac{3644.374}{6068.975} = 0.6$. Somit lautet $R = 0.775$.

17.3

Die Größe R^2 gibt an, wieviel Prozent der Variation von Y durch X erklärt wird. In diesem Fall wird 60% der Variation von Y durch X erklärt.

Aufgabe 18

18.1

ONEWAY ANOVA

reaktion

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	2880,000	2	1440,000	5,261	,012
Innerhalb der Gruppen	7390,000	27	273,704		
Gesamt	10270,000	29			

Es liegt ein signifikanter Gruppenunterschied vor, da $p < 0.05$. Allerdings kann nicht spezifiziert werden, welche Gruppen sich unterscheiden. Dies kann mittels Post-Hoc-Tests durchgeführt werden.

18.2

Das adjustierte Signifikanzniveau lautet $\alpha' = \alpha/k = 0.05/3 = 0.0167$. Die Post-Hoc-Tests zeigen, dass ein signifikanter Unterschied zwischen Gruppe 1 und 3 vorliegt. Alternativ können drei einzelne t-Tests durchgeführt werden, wobei zu beachten ist, dass für einen signifikanten Unterschied $p < \alpha' = 0.05/3$ gelten muss.

18.3

Hier liegt tatsächlich nur ein signifikanter Unterschied zwischen Gruppe 1 und 3 vor ($p < 0.0167$).

Aufgabe 19

19.1

Korrelationen

Korrelationen

		Umsatz	Mitarbeiter	LKW
Umsatz	Korrelation nach Pearson	1	,977**	,812**
	Signifikanz (2-seitig)		,000	,004
	Quadratsummen und Kreuzprodukte	66,400	1260,000	514,000
	Kovarianz	7,378	140,000	57,111
	N	10	10	10
Mitarbeiter	Korrelation nach Pearson	,977**	1	,772**
	Signifikanz (2-seitig)	,000		,009
	Quadratsummen und Kreuzprodukte	1260,000	25050,000	9500,000
	Kovarianz	140,000	2783,333	1055,556
	N	10	10	10
LKW	Korrelation nach Pearson	,812**	,772**	1
	Signifikanz (2-seitig)	,004	,009	
	Quadratsummen und Kreuzprodukte	514,000	9500,000	6040,000
	Kovarianz	57,111	1055,556	671,111
	N	10	10	10

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Partielle Korrelation

Korrelationen

Kontrollvariablen			Umsatz	LKW
Mitarbeiter	Umsatz	Korrelation	1,000	,421
		Signifikanz (zweiseitig)	.	,259
		Freiheitsgrade	0	7
	LKW	Korrelation	,421	1,000
		Signifikanz (zweiseitig)	,259	.
		Freiheitsgrade	7	0

Korrelationen

Kontrollvariablen			Umsatz	Mitarbeiter
LKW	Umsatz	Korrelation	1,000	,944
		Signifikanz (zweiseitig)	.	,000
		Freiheitsgrade	0	7
	Mitarbeiter	Korrelation	,944	1,000
		Signifikanz (zweiseitig)	,000	.
		Freiheitsgrade	7	0

Korrelationen

Kontrollvariablen			Mitarbeiter	LKW
Umsatz	Mitarbeiter	Korrelation	1,000	-,165
		Signifikanz (zweiseitig)	.	,671
		Freiheitsgrade	0	7
	LKW	Korrelation	-,165	1,000
		Signifikanz (zweiseitig)	,671	.
		Freiheitsgrade	7	0

Die Tabelle der bivariaten Korrelationen zeigt, dass bei jeder Paarbetrachtung ein positiver signifikanter Zusammenhang vorliegt. Werden dagegen die partiellen Korrelationen betrachtet, ergibt sich, dass lediglich ein positiver signifikanter Zusammenhang zwischen X (Umsatz) und Z (Mitarbeiter) besteht. Der positive Zusammenhang zwischen X und Y beruht somit auf der Korrelation zwischen X und Z. Es liegt eine Scheinkorrelation zwischen X und Y vor.

19.2

Es gilt $MQR(X,X) = 9\widehat{\mathrm{Cov}}(\tilde{X},\tilde{X})/8 = SQR(X,X)/8$ ($MQR(Y,Y)$ analog). Aus der Korrelationstabelle können alle Werte berechnet werden. Für MQR werden anstelle der Varianz und Kovarianz die Werte der Quadratsummen und Kreuzprodukte eingesetzt.

$$\begin{aligned}
\widehat{\mathrm{Cov}}(\tilde{X},\tilde{X}) &= 7.378 - 140^2/2783.333 = 0.336 \\
\widehat{\mathrm{Cov}}(\tilde{Y},\tilde{Y}) &= 671.111 - 1055.556^2/2783.333 = 270.8 \\
MQR(X,X) &= (66.4 - 1260^2/25050)/8 = 0.378 \\
MQR(Y,Y) &= (6040 - 9500^2/25050)/8 = 304.651
\end{aligned}$$

Die Berechnung von MQR kann auch direkt aus der ANOVA-Tabelle erfolgen.

$$\begin{aligned}
MQR(X,X) &= (66.4 - 63.377)/8 = 0.378 \\
MQR(Y,Y) &= (6040 - 3602.794)/8 = 304.651
\end{aligned}$$

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	63,377	1	63,377	167,734	,000[a]
	Nicht standardisierte Residuen	3,023	8	,378		
	Gesamt	66,400	9			

a. Einflußvariablen : (Konstante), Mitarbeiter

b. Abhängige Variable: Umsatz

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	3602,794	1	3602,794	11,826	,009[a]
	Nicht standardisierte Residuen	2437,206	8	304,651		
	Gesamt	6040,000	9			

a. Einflußvariablen : (Konstante), Mitarbeiter

b. Abhängige Variable: LKW

19.3

$$\begin{aligned}
\mathrm{Cov}(y,\hat{y}) &= \mathrm{Cov}(y,\alpha+\beta z) = \mathrm{Cov}(y,\beta z)\\
&= \beta\mathrm{Cov}(y,z)\\
&= \mathrm{Cov}(y,z)\mathrm{Cov}(z,z)^{-1}\mathrm{Cov}(y,z)\\
&= \mathrm{Cov}(y,z)\mathrm{Cov}(z,z)^{-1}\mathrm{Cov}(z,y)\\
&= \mathrm{Var}(\hat{y})
\end{aligned}$$

Aufgabe 20

Es gilt

$$\begin{aligned}
X &= \sum X_i && (1)\\
X_i &= T/k+\epsilon_i && (2)\\
\mathrm{Cov}(X_i,X_j) &:= \sigma_{ij} = \mathrm{Cov}(T,T)/(k^2) \geq 0 && (3)\\
\mathrm{Var}(X) &= \sum_i \sigma_i^2 + \sum_{i\neq j}\sigma_{ij} && (4)\\
\rho_{ij} &= \frac{\sigma_{ij}}{\sigma_i\sigma_j} && (5)\\
\bar{\rho} &= \frac{1}{k(k-1)}\sum_{i\neq j}\rho_{ij} && (6)\\
\sigma_i &= \sigma && (7)
\end{aligned}$$

Somit ergibt sich nach der allgemeinen Spearman-Brown-Formel (k Test-Teile) für die Reliabiliät:

$$\begin{aligned}
rel &= \frac{\mathrm{Var}(T)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(T,T)}{\mathrm{Var}\sum_i X_i}\\
&= \frac{k^2\sigma_{ij}}{\sum_i \mathrm{Var}(X_i) + \sum_{i\neq j}\mathrm{Cov}(X_i,X_j)}\\
&= \frac{k^2\sigma_{ij}}{\sum_i \sigma_i^2 + \sum_{i\neq j}\sigma_{ij}} = \frac{k^2\sigma_{ij}/\sigma^2}{\sum_i \sigma_i^2/\sigma^2 + \sum_{i\neq j}\sigma_{ij}/\sigma^2}\\
&= \frac{k^2\rho_{ij}}{k+\sum_{i\neq j}\rho_{ij}} = \frac{k\rho_{ij}}{1+\frac{1}{k}\sum_{i\neq j}\rho_{ij}}\\
&= \frac{k\rho_{ij}}{1+\frac{k-1}{k(k-1)}\sum_{i\neq j}\rho_{ij}} = \frac{k\rho_{ij}}{1+(k-1)\bar{\rho}}
\end{aligned}$$

Aufgabe 21

21.1

$$\begin{aligned} \bar{r} &= \frac{1}{k(k-1)} \sum_{i \neq j} r_{ij} = \frac{1}{12} 7.754 = 0.64617 \\ \hat{\alpha} &= \frac{k\bar{r}}{1 + (k-1)\bar{r}} = \frac{2.5847}{1 + 1.9385} = 0.88 \end{aligned}$$

21.2

$$p_i = (\{2.06, 2.19, 2.12, 2.26\} - 1)/(5 - 1)) = \{0.265, 0.298, 0.28, 0.315\}$$

Die Items sind eher leicht, d.h. die Personen sind eher positiv gegenüber der Filialgestaltung eingestellt.

21.3

Mittels der Trennschärfe wird hier analysiert, in welchem Maß die Antworten aller Items eines Konstruktes konstant bleiben. Die Trennschärfe gibt an, in wie weit eine Versuchsperson, die z.B. Item 1 hoch bewertert hat, auch alle anderen Items des Konstruktes hoch bewertet hat. Es wird untersucht, ob alle Items des Konstruktes dasselbe Merkmal erfassen.
Die Trennschärfe gibt hier die Korrelation eines einzelnen Items mit dem gesamten Konstrukt an, wenn das Item selbst unberücksichtigt bleibt. Anhand der Trennschärfen ist zu erkennen, dass das Item FRUHIG aus dem Rahmen fällt.

$$\begin{aligned} r_{X_1X} &= 0.816 \\ r_{X_2X} &= 0.823 \\ r_{X_3X} &= 0.839 \\ r_{X_4X} &= 0.498 \end{aligned}$$

21.4

Die itemspezifische Homogenität gibt an, inwieweit das Item zum Konstrukt passt. Items mit geringer itemspezifische Homogenität passen nicht zum Konstrukt. Das Item FRUHIG weist eine geringere itemspezifische Homogenität als die anderen Items auf.

$$r_{i+} = \{0.699, 0.703, 0.714, 0.468\}$$

Aufgabe 22

22.1

In der *Varianzanalyse* werden mit dem Begriff „Faktor“ die unabhängigen, qualitativen Variablen bezeichnet. Untersucht wird die Abhängigkeit einer quantitativen Variablen von unabhängigen nominalen Variablen, wobei die unabhängigen Variablen (Faktoren) beobachtbar sind.

In der *Faktorenanalyse* bezieht sich der Begriff „Faktor“ dagegen auf eine nicht beobachtbare (latente) Variable. Die latenten Variablen werden aus der Korrelationsmatrix der beobachtbaren Variablen „extrahiert“.

22.2

Die Faktorenanalyse ist ein Verfahren, mittels der die Dimensionalität der Datenstruktur betrachtet wird. Ziel ist es, die Dimensionalität zu reduzieren, in dem untersucht wird, ob sogenannte latente Faktoren existieren, welche die Korrelation ber beobachteten Variablen erklären, d.h. aus den beobachteten Variablen wird eine geringere Anzahl latenter Faktoren gesucht. Z.B. ist „Intelligenz“ ein latenter Faktor, der nicht direkt, sondern nur mittels mehrerer Variablen wie z.B. logisches Denkvermögen, Lesekompetenz usw., gemessen werden kann.
Um das Ziel der Reduzierung der Dimensionalität zu erreichen, wird u.a. die Hauptkomponentenanalyse verwendet. Die Hauptkomponentenanalyse ist ein Verfahren zur Extraktion der latenten Faktoren. Mittels der Hauptkomponentenanalyse werden unkorrelierte Linearkombinationen der beobachtbaren Variablen gebildet. Die erste Komponente besitzt den größten Varianzanteil. Nachfolgende Komponenten erklären stufenweise kleinere Anteile der Varianz.

22.3

Hauptkomponentenanalyse, Hauptachsentransformation, Maximum-Likelihood-Methode, Methode der kleinsten Quadrate.

22.4

Zur graphischen Bestimmung der Faktorenanzahl kann der sogenannte „scree plot“ herangezogen werden. Die Eigenwerte μ_i werden in absteigender Reihenfolge aufgetragen und miteinander verbunden. In der Regel ist an der Stelle mit der größten Differenz zwischen zwei Eigenwerten ein deutlicher Knick zu erkennen. Der letzte Punkt vor dem Knick bestimmt die Anzahl

der Faktoren. Es werden in der Analyse somit nur die Eigenwerte und die dazugehörigen Komponenten verwendet, die vor dem Knick liegen.

22.5

$$\begin{aligned} E(\xi|\mathbf{x}) &= \Lambda'\Sigma^{-1}\mathbf{x} = (P_1 M_1^{1/2})'(PMP')^{-1}\mathbf{x} \\ &= M_1^{1/2}P_1'P'^{-1}(PM)^{-1}\mathbf{x} = M_1^{1/2}P_1'P'^{-1}M^{-1}P^{-1}\mathbf{x} \\ &= M_1^{1/2}P_1'PM^{-1}P'\mathbf{x} = M_1^{1/2}[I_q,0]M^{-1}P'\mathbf{x} \\ &= M_1^{1/2}M^{-1}[I_q,0]P'\mathbf{x} = M_1^{1/2}M^{-1}[I_q,0][I_q,0]P'\mathbf{x} \\ &= M_1^{1/2}M_1^{-1}P_1\mathbf{x} \\ &= M_1^{-1/2}\mathbf{y}_1 \end{aligned}$$

Aufgabe 23

23.1

Die Kommunalitäten ergeben sich aus der Diagonalen der Matrix $\Lambda\Lambda'$, wobei Λ der Ladungsmatrix (Komponentenmatrix) entspricht. Mit

$$\hat{\Lambda} = \begin{pmatrix} 0.547 & 0.816 \\ 0.842 & -0.263 \\ 0.849 & -0.322 \\ 0.836 & 0.059 \end{pmatrix}$$

ergeben sich die geschätzten Kommunalitäten

$$\hat{h}_1^2 = 0.965, \quad \hat{h}_2^2 = 0.778, \quad \hat{h}_3^2 = 0.825, \quad \hat{h}_4^2 = 0.702.$$

23.2

Der erste Faktor erklärt 60.7% der Gesamt-Varianz.

23.3

Das erste Item (LHAUEFIG) passt nicht gut zu den anderen Items. Dies ist direkt an dem Komponentenplot zu erkennen. Besonders deutlich fällt der Unterschied bezüglich der zweiten Komponenten auf.