Lösungen zu den Übungsaufgaben in Kapitel 18

(1) Zeigen Sie, dass zwischen dem Partialregressionsgewicht und der Partialkorrelation die Beziehung F 18.13 gilt.

Nach Gleichung F 16.4 in Kapitel 16 lässt sich der Steigungskoeffizient einer einfachen Regressionsanalyse wie folgt bestimmen:

$$b_1 = r_{XY} \cdot \frac{s_Y}{s_X}$$

Das Regressionsgewicht in der multiplen Regression entspricht dem Regressionsgewicht in der einfachen Regression zwischen zwei Regressionsresiduen. Man erhält daher übertragen auf den Fall des Partialregressionsgewichts b_1:

$$b_1 = r_{E_{Y(X_2)}E_{X_1(X_2)}} \cdot \frac{s_{E_{Y(X_2)}}}{s_{E_{X_1(X_2)}}}.$$

Da die Korrelation zweier Residualvariablen der Partialkorrelation entspricht, gilt:

$$r_{E_{Y(X_2)}E_{X_1(X_2)}} = r_{YX_1 \bullet X_2}$$

Nach Gleichung F 16.14 gilt auch

$$s_{E_{Y(X_2)}} = \sqrt{s_Y^2 \cdot (1 - r_{X_2Y}^2)} \text{ und } s_{E_{X_1(X_2)}} = \sqrt{s_{X_1}^2 \cdot (1 - r_{X_2X_1}^2)}$$

Somit folgt aus Gleichung F 16.4 die Gleichung F 18.13 für den Fall eines Partialregressionsgewichts.

(2) Zeigen Sie, dass Gleichung F 18.19 aus Gleichung F 18.18 folgt.

Für die quadrierte Korrelation von Y und $\hat{Y}$ gilt:

$$r_{Y\hat{Y}}^2 = \frac{(s_{Y\hat{Y}})^2}{s_Y^2 \cdot s_{\hat{Y}}^2} = \frac{(s_{(\hat{Y}+E)\hat{Y}})^2}{s_Y^2 \cdot s_{\hat{Y}}^2} = \frac{(s_{\hat{Y}\hat{Y}} + s_{\hat{Y}E})^2}{s_Y^2 \cdot s_{\hat{Y}}^2} = \frac{(s_{\hat{Y}\hat{Y}})^2}{s_Y^2 \cdot s_{\hat{Y}}^2} = \frac{(s_{\hat{Y}}^2)^2}{s_Y^2 \cdot s_{\hat{Y}}^2} = \frac{s_{\hat{Y}}^2}{s_Y^2}$$

Die Beziehung $(s_{(\hat{Y}+E)\hat{Y}})^2 = (s_{\hat{Y}\hat{Y}} + s_{\hat{Y}E})^2$ gilt aufgrund der Rechenregeln für Kovarianzen (s. Rechenregel (4) in Abschnitt 15.4.1). Die Beziehung $(s_{(\hat{Y}+E)\hat{Y}})^2 = (s_{\hat{Y}\hat{Y}})^2$ gilt, da in der Regressionsanalyse die Variablen $\hat{Y}$ und E unkorreliert sind (s. Gleichung F 16.11). Da die Kovarianz einer Variablen mit sich selbst gleich der Varianz ist, gilt $(s_{\hat{Y}\hat{Y}})^2 = s_{\hat{Y}}^2$.

(3) In einer Untersuchung zur Lebenszufriedenheit (Y) wollen Sie die Hypothese überprüfen, dass sich die Arbeitszufriedenheit (X_1) umso stärker auf das Lebenszufriedenheitsurteil auswirkt, umso größer die Wichtigkeit (X_2) ist, die man der Arbeit zuschreibt. Hierzu erheben Sie die Lebenszufriedenheit, die Arbeitszufriedenheit und die Wichtigkeit mit intervallskalierten Skalen. Beschreiben Sie, wie Sie zur Überprüfung dieser Hypothese vorgehen. Formulieren Sie hierbei (mathematisch) das Regressionsmodell, das Sie Ihrer Hypothesenüberprüfung zugrunde legen, und formulieren Sie die statistische Nullhypothese, die Sie testen wollen. Beschreiben Sie auch, wie Sie zu einer statistischen Entscheidung gelangen.

Es handelt sich um die Frage, ob es sich bei der Wichtigkeit (X_2) um eine Moderatorvariable für den Zusammenhang zwischen der Lebenszufriedenheit (Y) und der Arbeitszufriedenheit (X_1) handelt. Daher formuliert man das folgende Regressionsmodell:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_1 \cdot X_2 + \varepsilon$$

Die Nullhypothese lautet: H_0: $\beta_3 \leq 0$, da bei der postulierten Moderatorbeziehung das Regressionsgewicht β_3 einen positiven Wert aufweisen muss. Nur dann gehen höhere Werte auf der Variablen X_2 mit einem höheren Regressionsgewicht für die Variable X_1 einher. Um diese gerichtete Nullhypothese zu überprüfen, testet man mit einem einseitigen t-Test nach Abschnitt 18.7.5, ob sich das Regressionsgewicht b_3, dass man anhand der Anwendung einer multiplen Regressionsanalyse erhält, bedeutsam von 0 in positiver Richtung unterscheidet. Ist der empirisch gefundene t-Wert größer als der kritische t-Wert, verwirft man die Nullhypothese. Hierzu muss das α-Niveau vorher festgelegt werden. Der üblichen Konvention folgend kann dies auf $\alpha = 0{,}05$ festgelegt werden. Um genug Teststärke zu haben, sollte vor der Durchführung der Studie eine A-priori-Poweranalyse zur Bestimmung der optimalen Stichprobengröße durchgeführt werden. Hierzu ist es notwendig, die Effektgröße und die Teststärke vorher festzulegen.

(4) In einer multiplen Regressionsanalyse wird der Einfluss der unabhängigen Variablen *A*, *B*, *C* und *D* auf die abhängige Variable *E* untersucht. Die einzelnen Variablen erfassen folgende Konstrukte:

***A*: Zufriedenheit mit dem Freundeskreis**
***B*: Zufriedenheit mit dem Partner**
***C*: Zufriedenheit mit der finanziellen Unterstützung durch die Eltern**
***D*: Zufriedenheit mit den Wohnverhältnissen**
***E*: Lebenszufriedenheit**

Die einzelnen unabhängigen Variablen wurden nacheinander in die Regressionsanalyse aufgenommen. In der folgenden Tabelle sind die Werte der zugehörigen Determinationskoeffizienten angegeben:

Prädiktoren in der Regressionsgleichung	R^2
A	0,16
A, *B*	0,27
A, *B*, *C*	0,49
A, *B*, *C*, *D*	0,65

Ergänzen Sie die folgenden Aussagen, so dass sie wahr sind.

Die Produkt-Momentkorrelation zwischen der Variablen *E* und der Variablen_______ beträgt *r* = 0.40.

Die _________________________-Korrelation zwischen der Variablen *E* und *D* beträgt *r* = 0,40.

Die multiple Korrelation zwischen der Variablen *E* und den Variablen *A*, *B* und *C* beträgt

___________________________.

Die Variablen *C* und *D* erklären 38 % der Variation in der Lebenszufriedenheit, unter der Bedingung, dass ____.

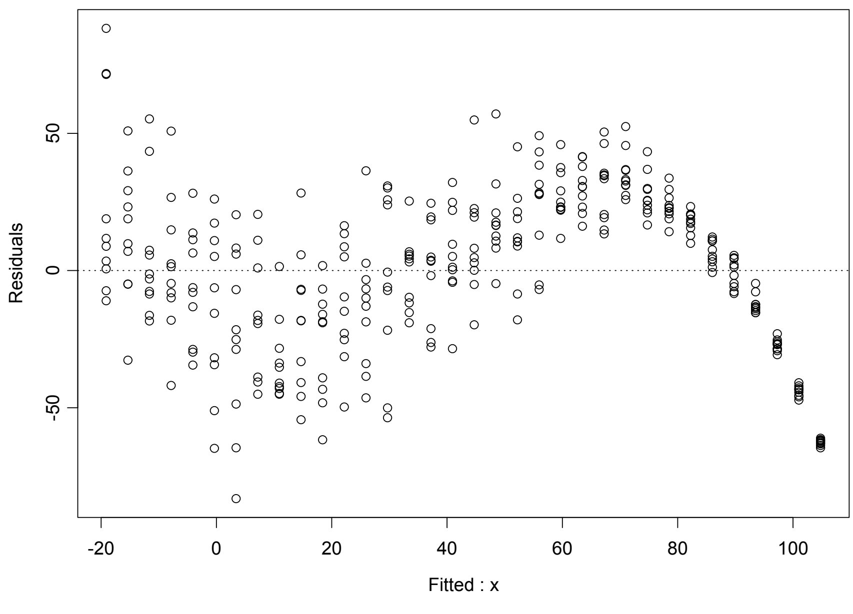
Die Produkt-Momentkorrelation zwischen der Variablen *E* und der Variablen___***A***___ beträgt $r = 0.40$.

Die ______________ **Semipartial** _____-Korrelation zwischen der Variablen *E* und *D* beträgt $r = 0{,}40$.

Die multiple Korrelation zwischen der Variablen E und den Variablen *A*, *B* und *C* beträgt ______________ **$R = 0{,}7$** ____.

Die Variablen *C* und *D* erklären 38 % der Variation in der Lebenszufriedenheit, unter der Bedingung, dass ___ **der Einfluss der Variablen *A* und *B* aus den Variablen *C* und *D* herauspartialisiert wurde.**______.

(5) Nachdem Sie eine einfache lineare Regression gerechnet haben, finden Sie folgenden Residuenplot. Die Achse »Fitted: x« kennzeichnet die vorhergesagten Werte. Welche Schlussfolgerungen ziehen Sie in Bezug auf die Gültigkeit der Annahmen der Regressionsanalyse in diesem Anwendungsfall? Welche Annahmen werden verletzt sein? Warum?



Erstens weist der Residuenplot darauf hin, dass die Annahme der Homoskedastizität verletzt ist. Die bedingte Varianz der Residuen ist im Bereich negativer gefitteter Werte größer als im Bereich positiver gefitteter Werte. Zweitens ist die Linearitätsannahme verletzt, da die Residuen nicht unsystematisch um den Wert 0 schwanken, sondern ein kurvilineares Muster aufweisen.

(6) In der folgenden Tabelle finden Sie zwei nominalskalierte Variablen. Sie möchten beide Variablen in eine Regressionsanalyse als unabhängige Variablen aufnehmen und neben ihren Haupteffekten auch den Effekt der Interaktion testen. Tragen Sie in die folgende Tabelle die Dummyvariablen ein, um die Information, die in den unabhängigen Variablen steckt, zu kodieren. Wählen Sie für jede Dummyvariable eine Spalte und geben Sie in den Zeilen für jede Dummyvariable die Werte an, die die Dummyvariable für diese Gruppe (Wertekombinationen) annimmt.

Variable 1 (Altersgruppe)	**Variable 2 (Bildungsniveau)**	**Dummyvariablen**
1	**1**	
2	**1**	
3	**1**	
1	**2**	
2	**2**	
3	**2**	

Variable 1 (Altersgruppe)	Variable 2 (Bildungsniveau)	Dummyvariablen				
		X_1	X_2	X_3	X_4 ($X_1 \cdot X_3$)	X_5 ($X_2 \cdot X_3$)
1	1	1	0	1	1	0
2	1	0	1	1	0	1
3	1	0	0	1	0	0
1	2	1	0	0	0	0
2	2	0	1	0	0	0
3	2	0	0	0	0	0