

Bachelorprüfung SS 2019 - MUSTERLÖSUNG

Fach: Praxis der empirischen Wirtschaftsforschung

Prüfer: Prof. Regina T. Riphahn, Ph.D.

Vorbemerkungen:

- Anzahl der Aufgaben:** Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen.
Es wird nur der Lösungsbogen eingesammelt. Angaben auf dem Aufgabenzettel werden nicht gewertet.
- Bewertung:** Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
- Erlaubte Hilfsmittel:**
- Formelsammlung (ist der Klausur beigelegt)
 - Tabellen der statistischen Verteilungen (sind der Klausur beigelegt)
 - Taschenrechner
 - Fremdwörterbuch
- Wichtige Hinweise:**
- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
 - Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1:**[15,5 Punkte]**

Sie interessieren sich für die Determinanten der Zahl gewonnener Medaillen bei den Olympischen Spielen 2016 in Brasilien. Dazu steht Ihnen ein Datensatz zu 81 Ländern mit folgenden Informationen zur Verfügung:

$medaillen_i$ Zahl der von Land i gewonnenen Medaillen
 $population_i$ Bevölkerung von Land i in Millionen Personen
 BIP_i Pro-Kopf-Bruttoinlandsprodukt von Land i in Tausend \$

Zunächst beschränken Sie die Analyse auf fünf Länder und unterstellen folgendes Regressionsmodell:

$$medaillen_i = \beta_0 + \beta_1 population_i + u_i$$

Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle.

a) Berechnen Sie anhand der folgenden Beobachtungen $\hat{\beta}_0$ und $\hat{\beta}_1$. (6 Punkte)

Land	$medaillen_i$	$population_i$
Jamaika	11	2,9
Bahamas	2	0,4
Neuseeland	18	4,7
Vietnam	2	95
Deutschland	42	82

Hinweis: $\sum_{i=1}^5 (population_i - \overline{population})^2 = 8934,66$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- $\bar{y} = \overline{medaillen} = (11 + 2 + 18 + 2 + 42)/5 = 15$ [1P]
- $\bar{x} = \overline{population} = (2,9 + 0,4 + 4,7 + 95 + 82)/5 = 37$ [1P]
- $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} = 2,9 \cdot 11 + 0,4 \cdot 2 + 4,7 \cdot 18 + 95 \cdot 2 + 82 \cdot 42 - 5 \cdot 37 \cdot 15 = 976,3$ [2P]
- $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{976,3}{8934,66} = 0,109$ [1P]
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 15 - 0,109 \cdot 37 = 10,967$ [1P]

Sie betrachten nun die ganze Stichprobe und schätzen folgendes Regressionsmodell mit SPSS:

$$medaillen_i = \beta_0 + \beta_1 population_i + \beta_2 population_i^2 + \beta_3 \ln(BIP_i) + u_i$$

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten			Signifikanz
	Regressionskoeffizient B	Standardfehler	T	
(Konstante)	-16,176	5,982	-2,700	0,008
<i>population</i>	0,186	0,034	5,530	0,000
<i>population</i> ²	-0,003	0,001	-3,000	0,000
<i>ln(BIP)</i>	6,528	1,770	3,690	0,000

a. Abhängige Variable: medaillen

b) Interpretieren Sie $\hat{\beta}_3$ inhaltlich und statistisch. (2 Punkte)

- Eine Erhöhung des Pro-Kopf-Bruttoinlandsproduktes um ein Prozent korreliert c.p.i.M. mit einer um $6,528/100 \approx 0,066$ höheren Anzahl an gewonnenen Medaillen. [1P]
- Der Koeffizient ist statistisch signifikant auf dem 1%-Niveau. [1P]

c) Bei welchem Wert für die Bevölkerung eines Landes ist die Anzahl an gewonnenen Medaillen am höchsten? Skizzieren Sie Ihren Rechenweg und interpretieren Sie Ihr Ergebnis. (2 Punkte)

- $\widehat{\frac{\Delta \text{medaillen}}{\Delta \text{population}}} = \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot \text{population}_i = 0,186 - 2 \cdot 0,003 \cdot \text{population}_i \stackrel{!}{=} 0$ [1P]
- $\Leftrightarrow \text{population}_{\max} = \frac{0,186}{0,006} = 31$ [0,5P]
- Bei einer Bevölkerung von 31 Millionen ist die Anzahl an gewonnenen Medaillen c.p. am höchsten. [0,5P]

d) Um wie viel unterscheidet sich die vorhergesagte Anzahl an gewonnenen Medaillen von Land A mit $BIP_A = 20$ von einem ansonsten identischen Land B mit $BIP_B = 30$? (1,5 Punkte)

- Der Unterschied in $\widehat{\text{medaillen}}$ beträgt $(\ln(30) - \ln(20)) \cdot \hat{\beta}_3 = 0,405 \cdot 6,528 = 2,644$ Medaillen. [1,5P]

e) Ein Kollege vermutet, dass die Annahme $E(u|X) = 0$ in Ihrem Modell verletzt sein könnte. Erklären Sie den Inhalt dieser Annahme. Beschreiben Sie die Auswirkung einer Verletzung dieser Annahme auf die geschätzten Koeffizienten. Beschreiben Sie anhand eines Beispiels, warum die Annahme hier verletzt sein könnte. (4 Punkte)

- Die Annahme $E(u|X) = 0$ fordert, dass der mittlere Störterm nicht von den erklärenden Variablen abhängt und für alle Werte von X gleich ist. [1P]
- Bei Verletzung dieser Annahme werden die Koeffizienten verzerrt geschätzt. [1P]
- Beispiel: In der obigen Schätzung wird z.B. die geographische Lage der einzelnen Länder nicht berücksichtigt. Die geographische Lage korreliert vermutlich mit der wirtschaftlichen Entwicklung und hat einen Einfluss auf die Anzahl an Medaillen in bestimmten Sportdisziplinen eines Landes. [2P]
- Andere Antworten möglich: Klima, sportliche Tradition, Kolonialgeschichte eines Landes.

Aufgabe 2:**[16,5 Punkte]**

Sie interessieren sich für das Abstimmungsverhalten von Mitgliedern des amerikanischen Senats zu einer Bildungsreform. Hierzu liegen Ihnen Daten zu 100 Senatorinnen und Senatoren, die alle Mitglied der republikanischen oder demokratischen Partei sind, mit folgenden Informationen vor:

- $reform_i$ = 1 wenn Person i der Reform zustimmt; =0 sonst
 $republican_i$ = 1 wenn Person i Mitglied der republikanischen Partei ist; =0 wenn Person i Mitglied der demokratischen Partei ist
 $democrat_i$ = 1 wenn Person i Mitglied der demokratischen Partei ist; =0 wenn Person i Mitglied der republikanischen Partei ist
 $female_i$ = 1 wenn Person i eine Frau ist; =0 wenn Person i ein Mann ist
 $male_i$ = 1 wenn Person i ein Mann ist; =0 wenn Person i eine Frau ist
 age_i Alter von Person i in Jahren

Sie unterstellen folgendes Regressionsmodell und schätzen dieses mit SPSS:

$$reform_i = \beta_0 + \beta_1 republican_i + \beta_2 female_i + \beta_3 age_i + \beta_4 (republican_i \cdot female_i) + u_i$$

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten			Signifikanz
	Regressionskoeffizient B	Standardfehler	T	
(Konstante)	0,289	0,103	2,757	0,006
republican	0,106	0,015	???	????
female	-0,009	0,007	-1,287	0,117
age	0,004	0,001	4,259	0,000
republican · female	-0,114	0,006	-19,150	0,000

a. Abhängige Variable: reform

Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle.

a) Interpretieren Sie $\hat{\beta}_3$ inhaltlich und statistisch. (2 Punkte)

- Eine Erhöhung des Alters um ein Jahr korreliert c.p.i.M. mit einer 0,4 Prozentpunkte höheren Wahrscheinlichkeit, für die Bildungsreform zu stimmen [1P]
- Der Koeffizient ist statistisch signifikant auf dem 1%-Niveau. [1P]

b) Sie möchten testen, ob der Koeffizient $\hat{\beta}_1$ signifikant ist. Führen Sie einen entsprechenden Test durch. Geben Sie Testverfahren, Null- und Alternativhypothese, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das 5%-Signifikanzniveau an. Für genau welche Gruppen testet dieser Test im Modellzusammenhang auf signifikante Unterschiede? (5 Punkte)

- Testverfahren: zweiseitiger t-Test [0,5P]
- Hypothesen: $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ [1P]

- Teststatistik: $t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0,106}{0,015} = 7,067$ [1P]
- Freiheitsgrade: $n - k - 1 = 100 - 4 - 1 = 95$ [0,5P]
- Kritischer Wert c 5%-Signifikanzniveau: $c = t_{\alpha; n-k-1} = t_{0,05; 95} = 1,987$ [0,5P]
(krit. Wert nicht tabelliert für df=95, nächstgelegener niedrigerer Wert: df=95) [Anmerkung: df=120 wird ebenfalls als richtig gewertet]
- Testentscheidung: Da $t_{empirisch} = 7,067 > 1,987 = c$ kann die Nullhypothese auf dem 5%-Niveau verworfen werden [0,5P]. Senatoren der republikanischen Partei haben eine auf dem 5%-Niveau signifikant andere Wahrscheinlichkeit, der Reform zuzustimmen, als Senatoren der Demokraten. [0,5P].
- Der Test testet, ob die Wahrscheinlichkeit, der Reform zuzustimmen, für männliche Senatoren je nach Parteizugehörigkeit signifikant unterschiedlich ist. [0,5P]

c) Berechnen Sie den marginalen Effekt des Geschlechts auf das Abstimmungsverhalten. Wie unterscheidet sich die Wahrscheinlichkeit, dem Gesetzentwurf zuzustimmen von republikanischen Senatorinnen im Vergleich zu demokratischen Senatorinnen? (2,5 Punkte)

- $\frac{\Delta reform}{\Delta female} = \hat{\beta}_2 + \hat{\beta}_4 republican_i = -0,009 - 0,114 \cdot republican_i$ [1P]
- Weiter ableiten nach *republican*: $\hat{\beta}_4 = -0,114$ [0,5P]
- Senatorinnen der republikanischen Partei haben c.p. im Mittel eine um 11,4 Prozentpunkte niedrigere Wahrscheinlichkeit, der Reform zuzustimmen, als Senatorinnen der Demokraten. [1P]

d) Frank (60 Jahre alt) und Claire Underwood (54 Jahre alt) sind beide Senatoren der demokratischen Partei. Wie hoch ist jeweils die vorhergesagte Wahrscheinlichkeit, dem Gesetzentwurf zuzustimmen? (3 Punkte)

- Frank: $\widehat{reform}_{Frank} = 0,289 + 60 \cdot 0,004 = 0,529$ [1P]
- Die vorhergesagte Wahrscheinlichkeit, dem Gesetzentwurf zuzustimmen, beträgt für Frank 52,9%. [0,5P]
- Claire: $\widehat{reform}_{Claire} = 0,289 + (-0,009) + 54 \cdot 0,004 = 0,496$ [1P]
- Die vorhergesagte Wahrscheinlichkeit, dem Gesetzentwurf zuzustimmen, beträgt für Claire 49,6%. [0,5P]

e) Ein Kommilitone ist der Meinung, dass das Modell fehlspezifiziert ist, weil die Variable *democrat_i* in Ihrer Schätzung nicht berücksichtigt wurde. Er schlägt folgendes Regressionsmodell vor:

$$reform_i = \gamma_0 + \gamma_1 republican_i + \gamma_2 female_i + \gamma_3 age_i + \gamma_4 (republican_i \cdot female_i) + \gamma_5 democrat_i + v_i$$

Welches Problem tritt hier auf? Wie muss das Modell spezifiziert werden, um die Dummy-Variablen *republican* und *democrat* gemeinsam als Regressoren zu verwenden? (2 Punkte)

- Es liegt perfekte Multikollinearität vor, MLR.3 ist verletzt. [1P]
- Man könnte das Modell mit den Dummyvariablen, aber ohne Konstante schätzen. [1P]

f) Nennen Sie zwei Schwächen des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

- Es ist möglich, dass vorhergesagte Werte außerhalb des Intervalls (0,1) liegen. [1P]
- Es liegt zwangsläufig Heteroskedastie vor. [1P]
- (Andere Lösungen denkbar.)

Aufgabe 3:

[18 Punkte]

Sie untersuchen die Determinanten des Interesses an der Fußball-Weltmeisterschaft der Frauen ("Frauen WM") 2019. Ihr Datensatz enthält für 500 Individuen folgende Informationen:

- Fernsehen_i* Anzahl der im Fernsehen angeschauten Spiele der Frauen WM 2019 von Person *i*
Alter_i Alter von Person *i* in Jahren
Frau =1, wenn Person *i* eine Frau ist; =0 sonst
FB_1_i =1, wenn Person *i* höchstens 100 Facebook-Freunde hat; =0 sonst
FB_2_i =1, wenn Person *i* mindestens 101 und höchstens 300 Facebook-Freunde hat; =0 sonst
FB_3_i =1, wenn Person *i* mindestens 301 Facebook-Freunde hat; =0 sonst

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Fernsehen_i = \beta_0 + \beta_1 Alter_i + \beta_2 Frau_i + \beta_3 FB_2_i + \beta_4 FB_3_i + u_i$$

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten		
	Regressionskoeffizient B	Standardfehler	T
(Konstante)	3,676	1,488	2,47
Alter	-0,054	0,006	-9,261
Frau	0,810	0,090	8,58
FB_2	0,388	0,041	9,463
FB_3	1,014	1,060	0,957

a. Abhängige Variable: *Fernsehen*

Hinweis: Das R^2 des Modells beträgt 0,455.

- a) Interpretieren Sie den geschätzten Koeffizienten von *Frau* inhaltlich. Ist der geschätzte Koeffizient signifikant? Begründen Sie. (2 Punkte)

- Frauen schauten c.p. im Mittel 0,81 Fußballspiele der Frauen WM 2019 mehr als Männer. [1P]
- Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden [0,5P] auf dem 1% Niveau (t-Wert: 8,58 > 2,576) [0,5P].

- b) Interpretieren Sie das R^2 des Modells. (1 Punkt)

- Das Modell erklärt 45,5% der Variation in der abhängigen Variable. [1P]

c) Berechnen Sie das korrigierte R^2 des Modells. (1,5 Punkte)

- $\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$
- $\bar{R}^2 = 1 - (1 - 0,455) \frac{500-1}{500-4-1}$ [1P]
 $= 1 - 0,549 = 0,451$ [0,5P]

d) Sie möchten testen, ob die Variablen FB_2 und FB_3 gemeinsam signifikant zum Erklärungsgehalt des Modells beitragen. Sie schätzen das Modell erneut ohne die beiden Variablen und erhalten ein R^2 von 0,375. Führen Sie einen entsprechenden Test am 5% Signifikanzniveau durch. Geben Sie Null- und Alternativhypothese, Freiheitsgrade, Teststatistik, kritischen Wert und Testentscheidung an. (5 Punkte)

- Nullhypothese: $H_0: \beta_3 = \beta_4 = 0$ [0,5P]
- Alternativhypothese: H_1 : mindestens ein $\beta_j \neq 0$ mit $j = 3, 4$ [0,5P]
- Freiheitsgrade (aus unrestringiertem Modell): $n - k - 1 = 500 - 4 - 1 = 495$ [0,5P] und $q = 2$ [0,5P]
- Teststatistik: $F_{\text{empirisch}} = \frac{(R_u^2 - R_R^2)/q}{(1 - R_u^2)/(n - k - 1)} = \frac{(0,455 - 0,375)/2}{(1 - 0,455)/(500 - 4 - 1)} = \frac{0,04}{0,001} = 40$ (alternativ ca. 36,4, wenn Nenner nicht gerundet) [1P]
- Kritischer Wert: $F_{\text{kritisch}} = F_{0,05;2;495} = 3,00$ [1P]
- Testentscheidung: Da $F_{\text{empirisch}} > F_{\text{kritisch}}$ wird die Nullhypothese auf dem 5% Niveau verworfen. [1P]
- Die Variablen tragen somit gemeinsam signifikant zum Erklärungsgehalt bei.

e) Bestimmen Sie das 99%-Konfidenzintervall für $\hat{\beta}_1$ und interpretieren Sie Ihr Ergebnis. Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (3 Punkte)

- $[-0,054 - t_{\frac{0,01}{2};500-4-1} \cdot 0,006; -0,054 + t_{\frac{0,01}{2};500-4-1} \cdot 0,006] = [-0,054 - 2,576 \cdot 0,006; -0,054 + 2,576 \cdot 0,006]$
 $= [-0,069; -0,039]$ [2P]
- Für wiederholte Stichproben liegt in 99% der Fälle der wahre Wert innerhalb der auf diese Weise berechneten Intervallgrenzen. [1P]

f) Welchen Wert würde $\hat{\beta}_1$ annehmen, wenn das Alter der Individuen nicht in Jahren, sondern in Monaten gemessen worden wäre? (1,5 Punkte)

- Umskalierung der unabhängigen Variable $\widetilde{\text{Alter}} = 12 \cdot \text{Alter}$:
- $\widehat{\text{Fernsehen}}_i = \hat{\beta}_0 + \hat{\beta}_1 \widetilde{\text{Alter}}_i + \hat{\beta}_2 \text{Frau}_i + \hat{\beta}_3 \text{FB_2}_i + \hat{\beta}_4 \text{FB_3}_i$
- $\frac{-0,054}{12} = -0,0045$ [1,5P]

- g) Für eine alternative Modellspezifikation berechnen Sie zunächst den natürlichen Logarithmus ($\ln$) der abhängigen Variable und schätzen das folgende neue lineare Regressionsmodell:

$$\ln(\text{Fernsehen}_i) = \beta_0 + \beta_1 \text{Alter}_i + \beta_2 \text{Frau}_i + \beta_3 \text{FB_2}_i + \beta_4 \text{FB_3}_i + u_i$$

- i. Das R^2 des Modells beträgt 0,450.
- ii. Das angepasste R^2 des Modells beträgt 0,446.
- iii. Der Korrelationskoeffizient zwischen den tatsächlichen und den vorhergesagten Werten für die Variable Fernsehen_i beträgt 0,691.

Welche der drei Maße sind ungeeignet, um die Schätzungsgüte des neuen Modells mit der des ursprünglichen Modells zu vergleichen? Begründen Sie kurz Ihre Antwort. Welches Kriterium würden Sie zum Vergleich der Modelle heranziehen? Welchem der beiden Modelle würden Sie den Vorrang geben? (4 Punkte)

- Ungeeignet sind das R^2 und das angepasste R^2 , [1P]
- Begründung: Die abhängige Variable ist nicht gleich. [1P]
- Der quadrierte Korrelationskoeffizient zwischen den tatsächlichen und den vorhergesagten Werten für die Variable Fernsehen_i kann herangezogen werden. [1P]
- Dem alternativen Modell (mit $\ln(\text{Fernsehen}_i)$) würde der Vorrang gegeben werden, da der quadrierte Korrelationskoeffizient größer als das R^2 des ursprünglichen Modells ist ($0,691^2 = 0,477 > 0,455$) [1P].

Aufgabe 4 - MC Fragen

[40 Punkte]

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. Zu jeder Frage gibt es genau eine richtige Antwort. Für jede korrekt beantwortete Frage erhalten Sie einen Punkt. Falsche Antworten führen nicht zu Punktabzug. Bei mehr oder weniger als einer markierten Antwort auf eine Frage gilt diese als nicht beantwortet. **Angaben auf dem Aufgabenblatt werden nicht gewertet.**

1.	Gegeben ist folgendes Modell: $y_i = \beta_0 + \beta_1 educ_i + \beta_2 age_i + \beta_3 exp_i + u_i$, wobei <i>educ</i> die Bildungsjahre angibt, <i>age</i> das Alter in Jahren und <i>exp</i> die potenzielle Berufserfahrung als Alter-Bildungsjahre-6. Welche Aussage ist korrekt?
a	Das Modell ist nicht linear in den Parametern.
b	X Es liegt perfekte Multikollinearität vor.
c	$\hat{y}_i$ hat keine sinnvolle Interpretation.
d	Es liegt Homoskedastie vor.

2.	Welche Aussage bezüglich der Annahmen MLR.1 - MLR.5 des multiplen linearen Regressionsmodells ist zutreffend?
a	X Gelten MLR.1 - MLR.5, so ist der Schätzer unverzerrt und effizient.
b	Wenn die Annahme MLR.3 zutrifft, ist der Erklärungsgehalt des Modells besonders hoch.
c	Eine Verletzung von MLR.5 führt zu verzerrt geschätzten Koeffizienten.
d	Alle Annahmen MLR.1 - MLR.5 müssen zutreffen, damit eine KQ-Schätzung durchgeführt werden kann.

3.	Wann bezeichnet man eine Variable in einer Schätzung als irrelevant?
a	Beim Vorliegen perfekter Multikollinearität.
b	Wenn die Variable nicht mit anderen erklärenden Variablen korreliert.
c	X Wenn die Variable nicht mit der abhängigen Variablen korreliert.
d	Wenn der Schätzer BLUE ist.

4.	Gegeben ist folgende Regression: $happiness_i = \beta_0 + \beta_1 age_i + \beta_2 female_i + \beta_3 female_i \cdot age_i + u_i$. Was beschreibt den Zusammenhang zwischen den Variablen <i>age</i> und <i>happiness</i> für Frauen?
a	X $\beta_1 + \beta_3$
b	$\beta_2 + \beta_3$
c	β_2
d	β_1

5.	Welche Aussage bezüglich Homo- und Heteroskedastie ist korrekt?
a	Der KQ-Schätzer ist auch im Falle von Homoskedastie ineffizient.
b	X Heteroskedastie bedeutet, dass die Varianz des Störterms nicht für alle Beobachtungen konstant ist.
c	Heteroskedastie führt zu verzerrt geschätzten Koeffizienten.
d	Homoskedastie bedeutet, dass $Var(u) = 1$.

6.	Sie schätzen das Modell $\log(lohn_i) = \beta_0 + \beta_1 bildung_i + \beta_2 berufserfahrung_i + u_i$. Sie vermuten, dass ein Problem ausgelassener Variablen vorliegt, da die Variable <i>Mann</i> nicht berücksichtigt wurde. Welche Aussage ist korrekt?
a	X Wenn $Cov(Mann, bildung) > 0$ und $Cov(Mann, lohn) > 0$, dann ist β_1 überschätzt.
b	Das Auslassen von <i>Mann</i> in der Schätzung führt zu Homoskedastie.
c	Wenn $Cov(Mann, bildung) = 0$ und $Cov(Mann, lohn) > 0$, dann ist β_1 verzerrt geschätzt.
d	Wenn $Cov(Mann, berufserfahrung) > 0$ und $Cov(Mann, lohn) = 0$, dann ist β_2 unterschätzt.

7.	Das Bestimmtheitsmaß $\bar{R}^2$
a	gibt den durch das Modell erklärten Anteil der Variation in der abhängigen Variable an.
b	X steigt nicht unbedingt, wenn einem Modell eine Variable mit einem geringen Erklärungsgehalt hinzugefügt wird.
c	wird durch einen F-Test berechnet.
d	wird direkt aus den Koeffizientenschätzern berechnet.

8.	Ein Koeffizient ist auf dem 10%-Niveau statistisch signifikant. Welche Aussage ist richtig?
a	Der Koeffizient ist zwangsläufig auch auf dem 5%-Niveau statistisch signifikant.
b	Der Koeffizient bildet einen kausalen Effekt ab.
c	X Der Signifikanztest liefert keine Aussage über die Stärke des Zusammenhangs.
d	Da der Koeffizient statistisch signifikant ist, liegt Homoskedastie vor.
9.	Im linearen Wahrscheinlichkeitsmodell
a	gilt MLR.1 nicht.
b	X liegen die beobachteten Werte für y im Intervall $[0,1]$.
c	können keine quadrierten Variablen als Regressoren verwendet werden.
d	liegen die vorhergesagten Werte für y im Intervall $[0,1]$.
10.	Welche Aussage bezüglich der Interpretation eines log-lin-Modells der Form $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$ ist richtig?
a	Durch die Logarithmierung von Variablen erhalten Ausreißer ein größeres Gewicht in der Schätzung.
b	X $\hat{\beta}_1 \cdot 100\%$ gibt die Änderung in $\hat{y}$ an, wenn sich x_1 um 1 Einheit erhöht.
c	Die Koeffizienten eines solchen Modells können nicht sinnvoll interpretiert werden.
d	Die Schätzkoeffizienten bilden eine Elastizität ab.
11.	Das 95%-Konfidenzintervall eines Koeffizienten $\hat{\beta}_1$ lautet $[-0,60; 0,20]$. Welche der Aussagen ist korrekt?
a	Es können keine Aussagen über $\hat{\beta}_1$ getroffen werden.
b	$\hat{\beta}_1$ ist auf dem 1%-Niveau statistisch signifikant von Null verschieden.
c	Die Berechnung des Konfidenzintervalls ist unabhängig von der Stichprobengröße.
d	X $\hat{\beta}_1 = -0,20$.
12.	Was erhöht unter Umständen die Effizienz einer KQ-Schätzung?
a	X Die Aufnahme relevanter Regressoren.
b	Höhere Varianz der Störterme.
c	Die Aufnahme von Ausreißern.
d	Eine Umskalierung der erklärenden Variable.
13.	Ein omitted variable bias im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$
a	X liegt vor, wenn eine relevante erklärende Variable ausgelassen wurde, die mit x_1 korreliert.
b	führt immer zu einer Unterschätzung von β_1 .
c	liegt vor, wenn eine Variable ausgelassen wurde, die mit x_1 korreliert, auch wenn sie irrelevant ist.
d	führt zu Homoskedastie.
14.	Im Falle einer binären abhängigen Variable
a	können keine Polynome erklärender Variablen als Regressoren verwendet werden.
b	X geben die Koeffizienten von Dummyvariablen Gruppenunterschiede in den vorhergesagten Wahrscheinlichkeiten an.
c	können die vorhergesagten Werte ausschließlich 0 und 1 sein.
d	haben Koeffizienten von kontinuierlichen Variablen keine sinnvolle Interpretation.
15.	Der Chow-Test
a	kann genutzt werden, um zu prüfen, ob Heteroskedastie vorliegt.
b	nutzt eine χ^2 -verteilte Teststatistik.
c	kann genutzt werden, um zu untersuchen, ob MLR.3 erfüllt ist.
d	X kann verwendet werden, um zu untersuchen, ob sich Steigungskoeffizienten zwischen zwei Gruppen signifikant unterscheiden.
16.	Gegeben ist das Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, wobei x_i und y_i binäre Variablen (0/1) sind. Zwei weitere Variablen sind definiert als $z_i = 1 - y_i$ und $w_i = 1 - x_i$. Wie verändert sich der Wert des R^2 , wenn y_i durch z_i oder x_i durch w_i umkodiert wird?
a	Das R^2 steigt, wenn nur die erklärende Variable umkodiert wird.
b	Das R^2 steigt, wenn nur die abhängige Variable umkodiert wird.
c	X Das R^2 steigt nicht, wenn eine der Variablen umkodiert wird.
d	Das R^2 steigt nur, wenn beide Variablen umkodiert werden.

17.	Gegeben sei folgendes Modell: $income_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$. In welchem der folgenden Modelle ist dieses Modell genestet?
a	$\log(income_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$
b	$hwage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$
c	$income_i = \beta_0 + \beta_1 educ_i + \beta_2 educ_i^2 + u_i$
d	X $income_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 female_i + u_i$

18.	Durch die Aufnahme eines irrelevanten Regressors in ein multiples lineares Regressionsmodell
a	werden die KQ-Schätzer verzerrt.
b	X steigt die Präzision der Schätzung nicht.
c	verändern sich die Regressionsergebnisse nicht.
d	verlieren unverzerrte Schätzer ihre kausale Interpretierbarkeit.

19.	Im linearen Wahrscheinlichkeitsmodell beschreibt $\hat{y}$ die Wahrscheinlichkeit, dass
a	$y=0$.
b	X $y=1$.
c	$y=1\%$.
d	$x=1$.

20.	Der Standardfehler von $\hat{\beta}_j$
a	steigt bei einer präziseren Schätzung.
b	ist keine Zufallsvariable.
c	ist ein Schätzer für die Varianz σ^2 .
d	X steigt mit steigender Residuenquadratsumme.

21.	Bei einem F-Test mit der Hypothese $\beta_3 = 0$
a	kann die Teststatistik sowohl positiv als auch negativ sein.
b	ist die Anzahl der Zählerfreiheitsgrade größer als die Anzahl der Nennerfreiheitsgrade.
c	kann das Bestimmtheitsmaß R^2 nicht zur Berechnung der Teststatistik verwendet werden.
d	X entspricht die Teststatistik dem quadrierten t-Wert des geschätzten Koeffizienten.

22.	Für die Schätzgleichung $\hat{y}_i = 1,5 - 0,5 \cdot x_i$ beträgt das Residuum für die Beobachtung $(x_1, y_1) = (1, 3)$
a	0.
b	1.
c	X 2.
d	4.

23.	Typ-I-Fehlerwahrscheinlichkeiten
a	sind für F-Tests irrelevant.
b	werden durch t-Werte beschrieben.
c	X beschreiben die Wahrscheinlichkeit, die wahre Nullhypothese zu verwerfen.
d	werden durch die Aufnahme zusätzlicher Kontrollvariablen verringert.

24.	Sie schätzen das Modell $\log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 \log(x_{2i}) + \beta_3 \exp(x_{3i}) + u_i$. Welcher Koeffizient gibt eine Semielastizität an?
a	$\hat{\beta}_0$
b	X $\hat{\beta}_1$
c	$\hat{\beta}_2$
d	$\hat{\beta}_3$

25.	In einer quadratischen Spezifikation $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ hat der Graph der Beziehung zwischen x und $\hat{y}$ die Form eines umgedrehten U, wenn
a	X $\hat{\beta}_1 > 0$ und $\hat{\beta}_2 < 0$.
b	$\hat{\beta}_1 < 0$ und $\hat{\beta}_2 > 0$.
c	$\hat{\beta}_1 > 0$ und $\hat{\beta}_2 > 0$.
d	$\hat{\beta}_1 < 0$ und $\hat{\beta}_2 = 0$.

26.	Bei kleiner Stichprobengröße führen χ^2 -verteilte Störterme zu
a	einem zu hoch geschätzten Achsenabschnittsparameter.
b	verzerrten Steigungsparametern.
c	einer inkonsistenten Schätzung.
d	X falschen Verteilungsannahmen bei t-Tests.

27.	Eine erklärende Variable wird als endogen bezeichnet, wenn sie korreliert mit
a	der abhängigen Variable.
b	anderen unabhängigen Variablen.
c	X dem Störterm.
d	der Konstante.

28.	Interaktionsterme zwischen zwei erklärenden Variablen
a	schließen bei einer Dummy-Variable und einer stetigen Variable konstante partielle Effekte aus.
b	X erlauben bei zwei Dummy-Variablen die Bestimmung unterschiedlicher Achsenabschnitte.
c	generieren bei zwei stetigen Variablen stets konstante partielle Effekte.
d	erlauben bei zwei stetigen Variablen die Bestimmung unterschiedlicher Achsenabschnitte.

29.	Sie möchten überprüfen, ob es im Modell $Lohn_i = \beta_0 + \beta_1 Bildung_i + \beta_2 Alter_i + \varepsilon_i$ signifikante Unterschiede in den Parametern zwischen Männern und Frauen gibt. Wie viele Parameter müssen Sie im Rahmen eines vollständig interagierten Modells schätzen?
a	4
b	5
c	X 6
d	7

30.	Wie viele Dummyvariablen müssen in ein Modell mit Konstante aufgenommen werden, um die kategoriale Variable Schulabschluss (1=Kein Schulabschluss, 2=Hauptschulabschluss, 3=Realschulabschluss, 4=Abitur) vollständig abzubilden?
a	1
b	2
c	X 3
d	4

31.	Sie schätzen das Modell $Lohn_i = \beta_0 + \beta_1 Bildung_i + \beta_2 Alter_i + \beta_3 female_i + \varepsilon_i$ und erhalten $\hat{\beta}_3 = -5$. Wie ändert sich der Koeffizient $\hat{\beta}_3$, wenn Sie im obigen Modell statt der Variable $female_i$ die Variable $male_i = 1 - female_i$ verwenden?
a	Der neue Koeffizient ist mit den gegebenen Werten nicht berechenbar.
b	Der Koeffizient verändert sich durch die Umformung nicht.
c	Der neue Koeffizient ist 6.
d	X Der neue Koeffizient ist 5.

32.	Wenn das Residuum einer KQ-Schätzung $y_i = \beta_0 + \beta_1 x_i + u_i$ für die dritte Beobachtung $\hat{u}_3 = -1$ beträgt,
a	liegt der Datenpunkt für diese Beobachtung in einem (x,y)-Diagramm überhalb der KQ-Regressionsgerade
b	X muss der beobachtete Wert der abhängigen Variable für diese Beobachtung kleiner sein als der vorhergesagte Wert.
c	ist die Annahme $E(u) = 0$ ungültig.
d	ist die Schätzung verzerrt.

33.	Wenn für zwei Zufallsvariablen a und b zutrifft, dass $E(ab) = 0$ und $E(a) = 0$, dann
a	sind a und b unabhängig.
b	X sind a und b unkorreliert.
c	gilt $cov(a, b) > 0$.
d	gilt $cov(a, b) < 0$.

34.	Wenn $\log(x) < 0$, ist
a	$x < 0$
b	$x = 0$
c	X $x > 0$ und $x < 1$
d	$x > 1$

35.	Was ändert sich <u>nicht</u> bei Umskalierung der abhängigen Variablen in einer KQ-Schätzung ?
a	Die Konfidenzintervalle der Parameter.
b	X Das R^2 der Schätzung.
c	Die Residuen $\hat{u}_i$.
d	Die Standardfehler $se(\hat{\beta}_j)$.

36.	Folgender Auszug aus einem Datensatz liegt Ihnen vor. Um welche Datenstruktur handelt es sich?															
	<table><tr><th>Personennr.</th><th>Jahr</th><th>Einkommen</th></tr><tr><td>1</td><td>2018</td><td>2200</td></tr><tr><td>1</td><td>2019</td><td>2400</td></tr><tr><td>2</td><td>2017</td><td>2300</td></tr><tr><td>2</td><td>2019</td><td>2100</td></tr></table>	Personennr.	Jahr	Einkommen	1	2018	2200	1	2019	2400	2	2017	2300	2	2019	2100
Personennr.	Jahr	Einkommen														
1	2018	2200														
1	2019	2400														
2	2017	2300														
2	2019	2100														
a	Zeitreihendaten.															
b	Querschnittsdaten.															
c	Gepoolte Querschnittsdaten.															
d	X Paneldaten.															

37.	Wird die Nullhypothese eines Strukturbruchtests nicht abgelehnt, so
a	X liegen keine signifikante Unterschiede in den Parametern verschiedener Gruppen vor.
b	liegt Heteroskedastie vor.
c	liegt perfekte Multikollinearität vor.
d	liegt ein Endogenitätsproblem vor.

38.	Ein Konfidenzintervall
a	hat für einen positiven Punktschätzer positive Intervallgrenzen.
b	X wird bei sinkender Stichprobengröße breiter.
c	hat für einen negativen Punktschätzer negative Intervallgrenzen.
d	umfasst stets den Wert des R^2 .

39.	Nehmen Sie an, Sie erhalten bei der Schätzung eines linearen multiplen Regressionsmodells ein R^2 von 1. Der Standardfehler der Regression
a	X beträgt 0.
b	beträgt 1.
c	kann im vorliegenden Fall nicht berechnet werden.
d	wird im vorliegenden Fall auf Basis des angepassten $\bar{R}^2$ berechnet.

40.	Konsistenz eines Schätzers bedeutet, dass
a	die Varianz des Schätzers bei zunehmender Beobachtungszahl steigt.
b	die Parameterschätzer auch bei kleinen Stichproben im Mittel stets den wahren Wert angeben.
c	X die geschätzten Parameter umso näher am wahren Wert liegen, je größer die Beobachtungszahl ist.
d	der Schätzer unverzerrt ist.