Bachelorprüfung WS 2015/2016 - MUSTERLÖSUNG

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 5 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 beigefügt)
- Tabellen der statistischen Verteilungen (sind der Klausur beigefügt)
- 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: **[12,5 Punkte]**

Sie möchten den Zusammenhang zwischen der Anzahl der Mitbewohner/innen einer Wohngemeinschaft (x_i) und dem Stromverbrauch in der Wohngemeinschaft (y_i) mittels des Modells $y_i = \beta_0 + \beta_1 x_i + u_i$ schätzen. Hierzu befragen Sie vier Personen und erhalten folgende Antworten:

	Anzahl Mitbewohner	Jährlicher Stromverbrauch (in 1.000 kWh)
Person 1	2	3,7
Person 2	3	3,3
Person 3	3	4,5
Person 4	4	4,7

a) Berechnen Sie $\hat{\beta}_0$ und $\hat{\beta}_1$ mit dem Kleinstquadrateverfahren und zeigen Sie Ihren Rechenweg. (6 Punkte)

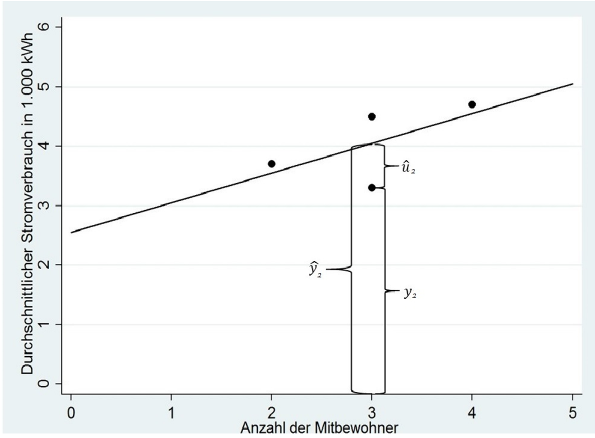
- $\hat{\beta}_1 = \frac{Cov[x,y]}{Var[x]}$ Alternativ: $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

  $\bar{x} = (2+3+3+4)/4 = 3$; $\bar{y} = (3,7+3,3+4,5+4,7)/4 = 4,05$

- $Cov[x,y] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{4-1}[(2-3)(3,7-4,05)+(3-3)(3,3-4,05)+(3-3)(4,5-4,05)+(4-3)(4,7-4,05)] = \frac{1}{3}$
- $Var[x] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{4-1}[(2-3)^2+(3-3)^2+(3-3)^2+(4-3)^2] = \frac{2}{3}$
- $\hat{\beta}_1 = \frac{\frac{1}{3}}{\frac{2}{3}} = 0,5$ und $\hat{\beta}_0 = 4,05 - 0,5 * 3 = 2,55$

b) Zeichnen Sie die Datenpunkte und die geschätzte Regressionsgerade aus Aufgabe a) in ein Koordinatensystem ein. Beschriften Sie die Achsen und machen Sie in der Grafik exemplarisch für Person 2 den wahren Wert y_2, den vorhergesagten Wert $\hat{y}_2$ und das Residuum $\hat{u}_2$ kenntlich. (4,5 Punkte)
Hinweis: Sollten Sie in Aufgabe a) zu keinem Ergebnis gekommen sein, verwenden Sie $\hat{\beta}_0 = 1,0$ *und* $\hat{\beta}_1 = 1,5$.

- korrekt beschriftete Achse
- Konstante korrekt eingezeichnet
- Steigung korrekt eingezeichnet
- y_2, $\hat{y}_2$ und $\hat{u}_2$ richtig eingezeichnet



c) Erläutern Sie kurz in Worten das Prinzip des Kleinstquadrateschätzers. (1 Punkt)

- Der KQ-Schätzer minimiert die Summe der quadrierten Residuen.

d) Interpretieren Sie den Koeffizienten $\hat{\beta}_0$ inhaltlich. (1 Punkt)

- Eine Person, die alleine wohnt (ohne Mitbewohner), verbraucht durchschnittlich 2.550 kWh Strom im Jahr.

Aufgabe 2: **[7,5 Punkte]**

Ein Skigebietbetreiber interessiert sich für die Determinanten der Anzahl verkaufter Skipässe in seinem Skigebiet. Es liegt ein Datensatz mit Informationen für 150 Wintertage vor:

Skipaesse	Zahl der verkauften Skipässe am Tag
Schneehoehe	Schneehöhe (in cm)
Werktag	=1, wenn Werktag, =0 wenn Sonn- und Feiertag

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Skipaesse_i = \beta_0 + \beta_1 Schneehoehe_i + \beta_2 Werktag_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	-12,650	8,324	-1,520	0,131
Schneehoehe	125,349	54,860	2,285	0,024
Werktag	-89,621	32,809	-2,817	0,005

a. Abhängige Variable: *Skipaesse*

a) Interpretieren Sie den geschätzten Koeffizienten der Variable *Schneehoehe* inhaltlich. (1 Punkt)

- `Bei einer 1cm höheren Schneeschicht werden c.p. im Mittel 125,35 Skipässe mehr verkauft.`

b) Wie würde sich der geschätzte Koeffizient der Variable $\hat{\beta}_2$ ändern, wenn die Variable *Werktag* an Sonn- und Feiertagen den Wert 1 und an Werktagen den Wert 0 annimmt? Ist der Koeffizient der umkodierten Variable $\tilde{\beta}_2$ statistisch signifikant? (2 Punkte)

- $\tilde{\beta}_2 = -\hat{\beta}_2$
- `Daher gilt:` $\tilde{\beta}_2 = 89,62$
- `Der Koeffizient ist statistisch signifikant auf dem 1% Niveau.`

c) Eine KQ-Schätzung des Modells ergibt $R^2 = 0,483$. Interpretieren Sie diesen Wert. (1 Punkt)

- `Das Modell erklärt 48,3% der Variation in der Anzahl der verkauften Skipässe.`

d) Welche der Gauß-Markov Annahmen könnte in Bezug auf die Variable *Schneehoehe* im Modell verletzt sein? Begründen Sie. Wird der Koeffizient β_1 dennoch unverzerrt und effizient geschätzt? (3,5 Punkte)

- `Die Linearitätsannahme (Gauß-Markov Annahme 1) könnte verletzt sein. Der Zusammenhang zwischen Schneehöhe und Besucherzahl ist möglicherweise nicht linear.`
- `Begründung: Abnehmender Grenznutzen der Schneehöhe. Wenn kein/wenig Schnee liegt, führt ein zusätzlicher cm Schnee zu einem starken Anstieg der Besucherzahl. Liegt bereits viel Schnee, führt ein zusätzlicher cm Schnee nur zu einem kleinen Anstieg der Besucherzahl.`
- `Der KQ-Schätzer ist weder unverzerrt noch effizient.`
- *`Alternative Antworten möglich.`*

Aufgabe 3: **[15 Punkte]**

Sie wollen die Determinanten der Verkaufspreise von gebrauchten PKWs untersuchen. Hierfür steht Ihnen ein Datensatz über 853 im Jahr 2015 verkaufte Gebrauchtwagen zur Verfügung.

Preis	Verkaufspreis eines gebrauchten PKWs (in Euro)
km	Kilometerstand beim Verkauf (in 100.000 km)
Verbrauch	Kraftstoffverbrauch (in Liter pro 100 km)
kW	Motorleistung (in Kilowatt)

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$log(Preis_i) = \beta_0 + \beta_1 km_i + \beta_2 Verbrauch_i + \beta_3 kW_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	2,942	1,534	1,918	0,055
km	-0,349	0,144	-2,417	0,016
Verbrauch	-0,012	0,011	-1,091	0,276
kW	0,002	0,001	2,000	0,046

a. Abhängige Variable: $log(Preis)$

a) Berechnen und interpretieren Sie inhaltlich den genauen marginalen Effekt des Kilometerstandes auf den Verkaufspreis. Interpretieren Sie den geschätzten Koeffizienten der Variable *km* statistisch. (3 Punkte)

- $\hat{\beta}_1 = -0,349$.
- Der genaue Effekt beträgt $\%\Delta\widehat{Preis} = 100\%[e^{\hat{\beta}_1} - 1] = -29,5\%$
- Inhaltliche Interpretation: Eine Erhöhung des Kilometerstandes beim Verkauf um 100.000km, verringert den Verkaufspreis eines gebrauchten PKWs c.p. im Mittel um 29,5 Prozent.
- Statistische Interpretation: Der Koeffizient ist auf dem 5%-Niveau statistisch signifikant von Null verschieden (p-Wert: $0,016 < 0,05$).

b) Bestimmen Sie das 95%-Konfidenzintervall für β_1 und interpretieren Sie Ihr Ergebnis. (Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle.) (3,5 Punkte)

- Konfidenzintervall: $[\hat{\beta}_1 - t_{\frac{\alpha}{2};n-k-1} \cdot se(\hat{\beta}_1); \hat{\beta}_1 + t_{\frac{\alpha}{2};n-k-1} \cdot se(\hat{\beta}_1)]$
- $\hat{\beta}_1 = -0,349$, $\alpha = 5\%$, $n = 853$, $k = 3$, $se(\hat{\beta}_1) = 0,144$
- $t_{\frac{0,05}{2};853-3-1} = 1,96$
- 95%-Konfidenzintervall um β_1: $[-0,349 - t_{\frac{0,05}{2};853-3-1} \cdot 0,144; -0,349 + t_{\frac{0,05}{2};853-3-1} \cdot 0,144]$
  $=[-0,349 - 1,96 \cdot 0,144; -0,349 + 1,96 \cdot 0,144]$
  $=[-0,631; -0,067]$
- Für wiederholte Stichproben liegt in 95% der Fälle der wahre Wert innerhalb der auf diese Weise berechneten Intervallgrenzen.

c) Sie vermuten, dass eine Erhöhung der Motorleistung eines gebrauchten PKWs um 1kW den Verkaufspreis um weniger als 0,3 Prozent erhöht. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das 5%-Signifikanzniveau an. Wird Ihre Vermutung bestätigt? (4,5 Punkte)

- Testverfahren: einseitiger t-Test
- Hypothesen: $H_0 : \beta_3 \geq 0,003, H_1 : \beta_3 < 0,003$
- Teststatistik: $t = \frac{\hat{\beta}_3 - 0,003}{se(\hat{\beta}_3)} = \frac{0,002-0,003}{0,001} = -1$
- Freiheitsgrade: $n - k - 1 = 853 - 3 - 1 = 849$
- Kritischer Wert: $-t_{\alpha;n-k-1} = -t_{0,05;853-3-1} = -1,645$
- Testentscheidung: Da $t_{empirisch} = -1 > -1,645 = t_{kritisch}$ kann die Nullhypothese auf dem 5%-Niveau nicht verworfen werden. Die Vermutung wird nicht bestätigt.

d) Testen Sie die Koeffizienten von *Verbrauch* und *kW* auf gemeinsame Signifikanz auf dem 1%-Niveau. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung an. Die Residuenquadratsumme (SSR) des restringierten Modells beträgt 2950,26 und des unrestringierten Modells 2940,18. (4 Punkte)

- Testverfahren: F-Test
- H_0: $\beta_1 = \beta_2 = 0$, H_1: mindestens ein $\beta_j \neq 0$ mit $j = 1,2$
- Teststatistik: $F_{emp} = \frac{(S_r - S_{ur})/q}{S_{ur}/(n-k-1)} = \frac{(2950{,}26 - 2940{,}18)/2}{2940{,}18/(853-3-1)} = 1{,}455$
- Freiheitsgrade: $n - k - 1 = 853 - 3 - 1 = 849$
- Kritischer Wert: $F_{kritisch} = F_{0{,}01;2;849} = 4{,}61$
- Testentscheidung: Da $F_{emp} < F_{kritisch}$ kann die Nullhypothese auf dem 1%-Niveau nicht verworfen werden. Die Koeffizienten der Variablen *Verbrauch* und *kW* sind nicht gemeinsam statistisch signifikant von Null verschieden.

Aufgabe 4: **[15 Punkte]**

Sie interessieren sich für die Determinanten von Lebenszufriedenheit. Ihnen steht ein Querschnittsdatensatz für 4.716 Personen mit folgenden Informationen zur Verfügung:

$Zufriedenheit_i$	Allgemeine Lebenszufriedenheit der Person i in Punkten (0-10 von schlecht bis gut)
$Einkommen_i$	Jahresbruttoeinkommen der Person i in 1.000 Euro
$Frau_i$	=1 für Frauen, =0 für Männer
$Verheiratet_i$	=1, wenn die Person verheiratet ist, =0 sonst

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Zufriedenheit_i = \beta_0 + \beta_1 Verheiratet_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	???	???	169,82	,000
Verheiratet	???	???	3,62	,000

a. Abhängige Variable: *Zufriedenheit*

Des Weiteren wissen Sie Folgendes über Ihre Stichprobe:

	Durchschnittliche Lebenszufriedenheit
Unverheiratete Personen	6,708
Verheiratete Personen	6,898

a) Berechnen Sie die Werte für $\widehat{\beta}_0$ und $\widehat{\beta}_1$. (2 Punkte)

- $\widehat{\beta}_0 = 6,708$.
- $\widehat{\beta}_1 = 6,898 - 6,708 = 0,19$.

b) Interpretieren Sie $\widehat{\beta}_1$ sowohl inhaltlich als auch statistisch. (2 Punkte)
Hinweis: Sollten Sie in Ausgabe a) zu keinem Ergebnis gekommen sein, verwenden Sie $\widehat{\beta}_1 = 0,5$.

- Verheiratete haben c.p. im Mittel eine um 0,19 Punkte höhere Lebenszufriedenheit als Unverheiratete.
- Alternativ: Verheiratete haben c.p. im Mittel eine um 0,5 Punkte höhere Lebenszufriedenheit als Unverheiratete.
- Der geschätzte Koeffizient ist auf dem 1%-Niveau statistisch signifikant von Null verschieden (p-Wert:0,000<0,01).

Sie erweitern das Modell um die Variablen $Einkommen_i$ und $Frau_i$ und schätzen nun folgendes Modell mit SPSS:

$$Zufriedenheit_i = \beta_0 + \beta_1 Verheiratet_i + \beta_2 Einkommen_i + \beta_3 Frau_i + \beta_4 Verheiratet_i \cdot Frau_i + u_i$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	6,64	,060	110,17	,000
Verheiratet	,052	,076	,68	,494
Einkommen	,004	,001	5,95	,041
Frau	-,002	,079	-0,03	,974
Verheiratet · Frau	,225	,104	2,16	,031

a. Abhängige Variable: *Zufriedenheit*

c) Interpretieren Sie den geschätzten Koeffizienten $\widehat{\beta}_4$ inhaltlich. (1,5 Punkte)

- Der Unterschied in der Lebenszufriedenheit zwischen Verheirateten und Unverheirateten ist bei Frauen c.p. im Mittel um 0,225 Punkten höher als bei Männern.
- Alternativ: Der Unterschied in der Lebenszufriedenheit zwischen Frauen und Männern ist bei Verheirateten c.p. im Mittel um 0,225 Punkte größer als bei Unverheirateten.

d) Welchen Wert würde $\widehat{\beta}_2$ sowie die t-Statistik von *Einkommen* annehmen, wenn das Einkommen nicht in 1000, sondern in 100 Euro gemessen worden wäre? Interpretieren Sie den geschätzten Koeffizienten $\widehat{\beta}_2$ nach Umskalierung sowohl inhaltlich als auch statistisch. (4 Punkte)

- Umskalierung der unabhängigen Variablen *Einkommen*:
- $\widehat{Zufriedenheit}_i = \widehat{\beta}_0 + \widehat{\beta}_1 Verheiratet_i + \widehat{\beta}_2 Einkommen_i + \widehat{\beta}_3 Frau_i + \widehat{\beta}_4 Verheiratet_i \cdot Frau_i$
  $\Leftrightarrow \widehat{\beta}_0 + \widehat{\beta}_1 Verheiratet_i + \frac{\widehat{\beta}_2}{10} 10 \cdot Einkommen_i + \widehat{\beta}_3 Frau_i + \widehat{\beta}_4 Verheiratet_i \cdot Frau_i$
- $\widetilde{\beta}_2 = \frac{\widehat{\beta}_2}{10} = 0{,}0004$.
- Umskalierung ändert die t-Statistik nicht. Der Wert bleibt bei 5,95.
- Inhaltliche Interpretation: Erhöht sich das Einkommen um 100 Euro, so erhöht sich die durchschnittliche Lebenszufriedenheit c.p. im Mittel um 0,0004 Punkte.
- Statistische Interpretation: Der geschätzte Koeffizient ist auf dem 5%-Niveau statistisch signifikant von Null verschieden (p-Wert:0,041<0,05).

e) Sie möchten überprüfen, ob sich der Heiratseffekt und der Einkommenseffekt im Modell aus Teilaufgabe c) gemeinsam zwischen Frauen und Männern signifikant unterscheiden. Benennen Sie das Testverfahren und stellen Sie die dafür geeigneten Modellgleichungen und Hypothesen auf. Geben Sie weiterhin Teststatistik, Freiheitsgrade, kritischen Wert für das 5%-Signifikanzniveau und Ablehnungsregel an. (Hinweis: Die Teststatistik muss nicht berechnet werden.) (5,5 Punkte)

- Testverfahren: (Schätzung vollständig interagiertes Modell mit anschl.) F-Test
- restringiertes Modell: $Zufriedenheit_i = \beta_0 + \beta_1 Verheiratet_i + \beta_2 Einkommen_i + \beta_3 Frau_i + u_i$
- unrestringiertes Modell: $Zufriedenheit_i = \beta_0 + \beta_1 Verheiratet_i + \beta_2 Einkommen_i + \beta_3 Frau_i + \beta_4 Verheiratet_i \cdot Frau_i + \beta_5 Einkommen_i \cdot Frau_i + u_i$
- Hypothesen: $H_0 : \beta_4 = \beta_5 = 0$, H_1 : mindestens ein $\beta_j \neq 0$ mit $j = 4,5$
- Teststatistik: $\frac{(R^2{}_U - R^2{}_R)/q}{(1-R^2{}_U)/(n-k-1)}$
- Freiheitsgrade: $n-k-1 = 4716-5-1 = 4710$
- Kritischer Wert: $c = F_{\alpha,q,n-k-1} = F_{0,05;2;4716-5-1} = 3{,}00$
- Ablehnungsregel: Wenn $F_{empirisch} > 3{,}00 = F_{kritisch}$ kann die Nullhypothese auf dem 5%-Niveau verworfen werden. d.h. β_4 und β_5 sind gemeinsam signifikant von Null verschieden.

Aufgabe 5 - 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.	Sie schätzen das Modell $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ und Ihr Kommilitone das Modell $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$. In Ihrem Modell wurde β_1 negativ verzerrt geschätzt, wenn für das Modell Ihres Kommilitonen gilt:
a	**X** $cov(x,z) < 0$ und $\beta_2 > 0$.
b	$cov(x,z) > 0$, $\beta_1 < 0$ und $\beta_2 > 0$.
c	$cov(x,z) < 0$ und $\beta_2 < 0$.
d	$cov(x,z) = 0$, $\beta_1 > 0$ und $\beta_2 < 0$.

2.	Ein Typ-2 Fehler liegt vor, wenn
a	H_0 nicht verworfen wird, obwohl sie zutrifft.
b	H_0 verworfen wird, obwohl sie nicht zutrifft.
c	H_0 verworfen wird, obwohl sie zutrifft.
d	**X** H_0 nicht verworfen wird, obwohl sie nicht zutrifft.

3.	Für die Unverzerrtheit des KQ-Schätzers ist es in jedem Fall notwendig, dass
a	Multikollinearität vorliegt.
b	eine Konstante mitgeschätzt wird.
c	$Var(u\|x) = \sigma^2$.
d	**X** $E(u\|x) = 0$.

4.	Die Schätzung des Modells $Arbeitslosenquote = \beta_0 + \beta_1 log(Wirtschaftswachstum) + u$ (wobei $Arbeitslosenquote$ und $Wirtschaftswachstum$ in 0-100% gemessen werden) ergibt $\widehat{\beta}_1 = 8$. Daher gilt:
a	Eine Steigerung des Wirtschaftswachstums um 1%, führt zu einer Erhöhung der Arbeitslosenquote um 8 Prozentpunkte.
b	Eine Steigerung des Wirtschaftswachstums um 1 Prozentpunkt, führt zu einer Erhöhung der Arbeitslosenquote um 8 %.
c	**X** Eine Steigerung des Wirtschaftswachstums um 10%, führt zu einer Erhöhung der Arbeitslosenquote um 0,8 Prozentpunkte.
d	Eine Steigerung des Wirtschaftswachstums um 1%, führt zu einer Erhöhung der Arbeitslosenquote um 8%.

5.	Unter den Annahmen des klassischen linearen Modells folgt der KQ-Schätzer
a	einer F-Verteilung.
b	einer χ^2Verteilung.
c	**X** einer Normalverteilung.
d	keiner Verteilung.

6.	Gegeben ist folgendes Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$. Die aufgestellten Hypothesen lauten $H_0 : \beta_3 \geq -4$ vs. $H_1 : \beta_3 < -4$. Auf einem Signifikanzniveau von 1% führt eine Teststatistik von $-2,519$
a	bei $n = 20$ zur Ablehnung von H_0.
b	**X** bei $n = 25$ zur Ablehnung von H_0.
c	bei $n = 30$ nicht zur Ablehnung von H_0.
d	bei $n = 35$ nicht zur Ablehnung von H_0.

7.	Sie haben Daten über das Angebot an Kindergartenplätzen (pro Bewohner) für Großstädte (>100.000 Einwohner) und Kleinstädte (<100.000 Einwohnern) in Ost- und Westdeutschland. Bei welcher Modellspezifikation kommt es nicht zur Dummy-Variable-Trap?
a	$Plätze_i = \beta_0 + \beta_1 Ost_i + \beta_2 Großstadt_i \cdot Ost_i + \beta_3 Kleinstadt_i \cdot Ost_i + u_i$
b	$Plätze_i = \beta_1 West_i + \beta_2 Ost_i + \beta_3 Kleinstadt_i \cdot West_i + \beta_4 Großstadt_i \cdot West_i + u_i$
c	**X** $Plätze_i = \beta_0 + \beta_1 Großstadt_i \cdot Ost_i + \beta_2 Kleinstadt_i \cdot West_i + \beta_3 Kleinstadt_i \cdot Ost_i + u_i$
d	$Plätze_i = \beta_1 Kleinstadt_i + \beta_2 Kleinstadt_i \cdot West_i + \beta_3 Kleinstadt_i \cdot Ost_i + u_i$

8.	Welche Aussage über die Regressionsgleichung $y_i = \beta_0 + \beta_1 log(x_i) + u_i$ ist richtig?
a	β_1 beschreibt, um wie viel Prozent sich y bei einer Änderung von x um eine Einheit ändert.
b	β_1 kann als Elastizität interpretiert werden.
c	**X** $\beta_1/100$ beschreibt, um welchen absoluten Betrag sich y bei einer Änderung von x um 1% ändert.
d	Die abhängige Variable y ist linear in x.

9.	Für das Modell $Unfall_i = \beta_0 + \beta_1 Alter_i + \beta_2 Alter_i^2 + \varepsilon_i$ ergibt eine KQ-Schätzung $\widehat{\beta}_1$ = -0,210 und $\widehat{\beta}_2$ = 0,002. Bei welchem Alter wird die Unfallhäufigkeit minimiert?
a	26,25
b	**X** 52,5
c	55
d	47,75

10.	Die asymptotische Effizienz
a	setzt Normalverteilung des Schätzers voraus.
b	**X** setzt Konsistenz des Schätzers voraus.
c	ist beim KQ Schätzer in kleinen Stichproben immer gegeben.
d	ist beim KQ Schätzer in großen Stichproben immer gegeben.

11.	Sie führen nacheinander einen rechtsseitigen und einen beidseitigen t-Test durch. Wie unterscheiden sich die kritischen Werte der t-Statistik im Betrag, wenn beide Tests für das gleiche Modell, die gleiche Stichprobe und das gleiche Signifikanzniveau durchgeführt werden?
a	Der kritische Wert des einseitigen Tests ist größer.
b	Der kritische Wert ist in beiden Tests gleich groß.
c	Die Antwort hängt von dem Vorzeichen des Koeffizienten ab.
d	**X** Der kritische Wert des beidseitigen Tests ist größer.

12.	Beim Chow-Test
a	**X** besagt die Nullhypothese, dass der Achsenabschnittsparameter für verschiedene Gruppen identisch ist.
b	folgt die Teststatistik bei Gültigkeit der Alternativhypothese immer einer t-Verteilung.
c	besagt die Alternativhypothese, dass die Steigungsparameter für verschiedene Gruppen identisch sind.
d	folgt die Teststatistik bei Gültigkeit der Nullhypothese immer einer t-Verteilung.

13.	Sie untersuchen den Zusammenhang zwischen Einkommen und Entfernung zum Urlaubsort. Dabei stellen Sie fest, dass mit steigendem Einkommen die Variation der Entfernung zum Urlaubsort zunimmt. Welches Problem liegt hier vor?
a	**X** Heteroskedastie.
b	Simultanität.
c	Multikollinearität.
d	Rechtsschiefe Verteilung.

14.	Mit einem F-Test vergleichen Sie zwei Modelle (M1 und M2) mit der gleichen abhängigen Variable. M1 enthält neben der Konstante ausschließlich die erklärenden Variablen: Bildung, Alter, Geschlecht und Gewicht. M2 berücksichtigt nur Bildung als erklärende Variable. Welche Aussage trifft zu?
a	M1 ist das restringierte Modell.
b	M1 enthält vier Parameter.
c	Die Anzahl der Zählerfreiheitsgrade ist $q = 2$.
d	**X** Das restringierte Modell berücksichtigt einen Steigungsparameter.

15.	Eine Dummy-Variable-Trap
a	ist bei einer großen Stichprobe kein Problem.
b	**X** lässt sich durch Aufnahme einer relevanten Dummy-Variablen nicht lösen.
c	entsteht durch einen Interaktionsterm zwischen zwei Dummy-Variablen.
d	lässt sich immer durch die Schätzung ohne Konstante lösen.

16.	Welche Aussage über Konsistenz ist korrekt?
a	**X** Die asymptotische Eigenschaft der Konsistenz betrachtet, wie weit der Wahrscheinlichkeitsgrenzwert eines Schätzers bei unendlich großen Stichproben vom wahren Wert entfernt liegt.
b	Bei konsistenten Schätzverfahren steigt die Varianz des Schätzers mit der Größe der Stichprobe.
c	Ein konsistenter Schätzer trifft unabhängig von der Größe der Stichprobe immer im Mittel den wahren Wert.
d	Ein inkonsistenter Schätzer ist immer auch verzerrt.

17.	Durch eine Erhöhung der Stichprobengröße
a	**X** kann das Konfidenzintervall enger werden.
b	steigt das R^2.
c	können Heteroskedastieprobleme gelöst werden.
d	können sich die Standardfehler der Koeffizienten nicht ändern.

18.	Sie schätzen das Modell $Lebenserwartung_i = \beta_0 + \beta_1 Mann_i + u_i$ und erhalten $\widehat{\beta}_0$ = 80 und $\widehat{\beta}_1$ = -5. Für die Schätzung des Modells $Lebenserwartung_i = \gamma_0 + \gamma_1 Frau_i + u_i$ muss dann Folgendes gelten:
a	$\widehat{Lebenserwartung}_i = 85$ für $Frau_i = 1$.
b	$\widehat{\gamma}_1 = \widehat{\beta}_0 + \widehat{\beta}_1$.
c	$\widehat{\gamma}_0 = 80$.
d	**X** $\widehat{\gamma}_0 + \widehat{\gamma}_1 = \widehat{\beta}_0$.

19.	Mithilfe des p-Wertes lassen sich Aussagen zur
a	ökonomischen Signifikanz des Koeffizienten treffen.
b	Varianz der Variable treffen.
c	**X** statistischen Signifikanz des Koeffizienten treffen.
d	Konsistenz des Schätzers treffen.

20.	Interaktionsterme
a	haben immer einen Erwartungswert von null.
b	**X** erlauben bei stetigen Variablen die Schätzung von nicht-konstanten partiellen Effekten.
c	müssen mindestens eine stetige Variable enthalten.
d	sind das Produkt dreier unterschiedlicher Variablen.

21.	Die Interpretation kausaler Effekte in einem linearen Regressionsmodell
a	ist nicht möglich.
b	**X** ist nur verlässlich, wenn alle relevanten Faktoren im Modell konstant gehalten werden.
c	entspricht der Korrelation zwischen der abhängigen und der erklärenden Variablen.
d	setzt voraus, dass $R^2 > 0,5$.

22.	Wenn im linearen Modell $log(Lohn) = \beta_0 + \beta_1 log(Bildung) + u$ die Variable Bildung nicht in Jahren, sondern in Monaten gemessen würde,
a	steigt der geschätzte Koeffizient für Bildung auf das 12-fache an.
b	sinkt die geschätzte Konstante um $log(12)$.
c	**X** steigt die geschätzte Konstante um $\widehat{\beta}_1 * log(12)$ an.
d	sinkt der geschätzte Koeffizient für Bildung auf das $\frac{1}{12}$-fache.

23.	Ein Konfidenzintervall wird
a	**X** mit steigendem Signifikanzniveau α enger.
b	mit steigendem Standardfehler $se(\beta_j)$ enger.
c	mit steigendem Schätzwert $\hat{\beta}_j$ breiter.
d	mit kleinerem Schätzwert $\hat{\beta}_j$ enger.

24.	Sie schätzen den Zusammenhang zwischen der Fertilität von Akademikerinnen und Nicht-Akademikerinnen in Deutschland und Frankreich mit folgendem Modell: $Geburten_i = \beta_1 Deutschland_i + \beta_2 Akademikerin_i + \beta_3 Frankreich_i + \beta_4 Akademikerin_i \cdot Deutschland_i + u_i$. Was misst der geschätzte Parameter β_2?
a	Den Mittelwert der Geburten für Französinnen.
b	**X** Den Mittelwertunterschied der Geburten von Akademikerinnen und Nicht-Akademikerinnen in Frankreich.
c	Den Mittelwert der Geburten von Akademikerinnen in Frankreich.
d	Den Mittelwertunterschied der Geburten von Nicht-Akademikerinnen in Deutschland und Frankreich.

25.	Leitet man den Kleinstquadrate-Schätzer für das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$ her, ergeben sich durch Minimieren der quadrierten Residuen höchstens
a	2 Bedingungen erster Ordnung.
b	k Bedingungen erster Ordnung.
c	**X** $k+1$ Bedingungen erster Ordnung.
d	$k-1$ Bedingungen erster Ordnung.

26.	Die Aufnahme irrelevanter Variablen in ein Regressionsmodell
a	verringert den Wert des R^2.
b	erhöht den Wert des angepassten Bestimmtheitsmaßes.
c	führt zu verzerrten KQ-Schätzern.
d	**X** verringert die Effizienz der Schätzung.

27.	Im linearen Modell $log(y_i) = \beta_0 + \beta_1 x_i + \beta_2 {x_i}^2 + u_i$ mit $\beta_1 > 0$ und $\beta_2 > 0$
a	ist der marginale Effekt von x auf y konstant.
b	nimmt y in x zunächst ab und nimmt nach dem Minimum zu.
c	sinkt y kontinuierlich mit steigendem x ab, für $x > 0$.
d	**X** ist der marginale Effekt von x auf y für $x > 0$ positiv.

28.	Für einen t-Test mit $H_1 : \beta_1 < 0$ ergibt sich ein kritischer Wert von $t_{0,10;35-4-1} = 1,310$ und eine Teststatistik mit $t_{\hat{\beta}_1} = -1,189$. Welche Interpretation ist richtig?
a	Die Nullhypothese kann auf dem 10%-Signifikanzniveau verworfen werden.
b	**X** Die Nullhypothese kann auf dem 10%-Signifikanzniveau nicht verworfen werden.
c	β_1 ist signifikant von Null verschieden.
d	Die Alternativhypothese kann auf dem 10%-Signifikanzniveau angenommen werden.

29.	Das R^2 kann als Maß für einen Modellvergleich <u>nicht</u> herangezogen werden, wenn
a	die zu vergleichenden Modelle genestet sind.
b	in beiden Modellen die gleiche Anzahl an Parametern verwendet wird.
c	**X** die Modelle unterschiedliche abhängige Variablen verwenden.
d	eines der Modelle ein $R^2 < 0,5$ hat.

30.	Das Auslassen der Variable z_i im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ führt zu verzerrten Parameterschätzern, wenn z
a	mit x korreliert ist.
b	mit y korreliert ist.
c	mit y unkorreliert ist.
d	**X** mit x und y korreliert ist.

31.	Wird im Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ mit $\beta_0 \neq 0$ die Konstante β_0 irrtümlich nicht mitgeschätzt, so
a	ist gewährleistet, dass der Mittelwert der Residuen Null ist.
b	**X** so spricht man von einer Regression durch den Ursprung.
c	kann das R^2 negativ werden.
d	sind die Steigungsparameter unverzerrt.

32.	Das angepasste Bestimmtheitsmaß $\overline{R}^2$
a	beschreibt den Anteil der Variation in y der durch das Modell erklärt wird.
b	dient zum Vergleich von Modellen mit der gleichen Anzahl erklärender Variablen und unterschiedlichen abhängigen Variablen.
c	**X** kann durch die Aufnahme eines nicht-signifikanten Regressors steigen.
d	nimmt ausschließlich positive Werte an.

33.	Für einen geschätzten Koeffizienten erhalten Sie im Rahmen eines Signifikanztests einen t-Wert von 2,500 und einen p-Wert von 0,013. Welche der folgenden Aussagen kann damit getroffen werden?
a	Der Koeffizient der betrachteten Variable ist nicht signifikant auf dem 5%-Niveau.
b	Der Koeffizient der betrachteten Variable ist signifikant auf dem 1%-Niveau.
c	**X** Der Koeffizient der betrachteten Variable ist signifikant auf dem 10%-Niveau.
d	Der Erklärungsgehalt des Modells steigt durch die Aufnahme dieser Variable um 1,3%.

34.	Gepoolte Querschnittsdaten
a	beobachten die gleichen Individuen über mehrere Zeitpunkte.
b	werden auch als Zeitreihendaten bezeichnet.
c	werden an einem Zeitpunkt erhoben.
d	**X** können genutzt werden, um die Wirkung von Reformen zu betrachten.

35.	Im linearen Wahrscheinlichkeitsmodell
a	nimmt die abhängige Variable y Werte außerhalb des Intervalls (0,1) an.
b	**X** gibt $1-\widehat{y}$ die geschätzte Wahrscheinlichkeit des Ereignisses $y=0$ an.
c	liegt per Definition perfekte Multikollinearität vor.
d	sind die erklärenden Variablen stetig skaliert.

36.	Im Modell $log(y) = \beta_0 + \beta_1 x_1 + \beta_2 log(x_2) + u$ mit $\beta_1 > 0$ und $\beta_2 > 0$
a	**X** ist $log(y)$ linear im Parameter β_2.
b	führt eine 1%ige Senkung von x_2 zu einer Steigerung von y um β_2%.
c	gilt für die Vorhersage der abhängigen Variablen $\widehat{y} = exp(\widehat{log(y)})$
d	führt eine Steigung von x_1 um eine Einheit zu einer Steigerung von y um $0,01\beta_1$%.

37.	Gilt $E[u\|x] \neq 0$, dann
a	**X** können ausgelassene Variablen den KQ-Schätzer verzerren.
b	ist der KQ-Schätzer erwartungstreu.
c	kann der KQ-Schätzer kausal interpretiert werden.
d	muss die Konstante den Wert 0 annehmen.

38.	Logarithmieren der abhängigen Variablen
a	**X** schließt Beobachtungen mit $y = 0$ von der Schätzung aus.
b	steigert das Gewicht von Ausreißerbeobachtungen.
c	erhöht die Streuung der abhängigen Variablen.
d	führt zu heteroskedastischen Störtermen.

39.	Für die Regression $y_i = \beta_0 + \beta_1 x_i + u_i$ ist $Var(\hat{\beta}_1)$ umso kleiner,
a	je größer σ_u^2.
b	**X** je größer die Streuung von x.
c	je kleiner die Stichprobe.
d	je größer β_0.

40.	Homoskedastie im linearen Modell $y_i = \beta_0 + \beta_1 x_i + u_i$
a	ist formal definiert als $\sum_{i=1}^{n} x_i \hat{u}_i = 0$.
b	**X** liegt vor, wenn die Varianz von u nicht von x abhängt.
c	ist nur in kleinen Stichproben ein Problem.
d	erzeugt verzerrte Schätzer.