Bachelorprüfung WS 2014/15 - 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: **[16 Punkte]**

Sie interessieren sich für die Determinanten von Hauspreisen. Ihr Datensatz enthält für 506 Häuser folgende Informationen:

Ln_Preis	Hauspreis (logarithmiert)
Ln_Konz	Feinstaub-Konzentration in der Umgebung (logarithmiert)
Zimmer	Anzahl der Zimmer
Laerm	Fluglärm (gemessen in Dezibel)
Baujahr	Baujahr des Hauses

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Ln\_Preis_i = \beta_0 + \beta_1 Ln\_Konz_i + \beta_2 Zimmer_i + \beta_3 Laerm + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	10,359	0,219	47,29	0,000
Ln_Konz	-0,645	0,063	-10,31	0,000
Zimmer	0,261	0,019	13,75	0,000
Laerm	-0,051	0,006	-8,58	0,000

a. Abhängige Variable: *Ln_Preis*

a) Interpretieren Sie den geschätzten Koeffizienten von *Ln_Konz* sowohl inhaltlich als auch statistisch. (2 Punkte)

- Eine Erhöhung der Feinstaub-Konzentration von 1% geht c.p. im Mittel mit einer Hauspreisreduktion von 0,645% einher. [1P]
- Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden auf dem 1% Niveau (p-Wert:0,000<0,01). [1P]

b) Bestimmen Sie das 99%-Konfidenzintervall für β_2 und interpretieren Sie Ihr Ergebnis. Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (3 Punkte)

- $[0,261 - t_{\frac{0,01}{2};506-3-1} \cdot 0,019; 0,261 + t_{\frac{0,01}{2};506-3-1} \cdot 0,019] = [0,261 - 2,756 \cdot 0,019; 0,261 + 2,756 \cdot 0,019;] = [0,209; 0,313]$ [2P]
- Für wiederholte Stichproben liegt in 99% der Fälle der wahre Wert innerhalb der auf diese Weise berechneten Intervallgrenzen. [1P]

c) Testen Sie auf dem 10%-Signifikanzniveau, ob bei einer Fluglärmerhöhung um 1 Dezibel der Hauspreis um mehr als 5% sinkt. Geben Sie Testverfahren, Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

- Testverfahren: einseitiger t-Test [0,5P]
- Hypothesen: $H_0: \beta_3 \geq -0,05, H_1: \beta_3 < -0,05$ [1P]
- Teststatistik: $t = \frac{\hat{\beta}_3 + 0,05}{se(\hat{\beta}_3)} = \frac{-0,051+0,05}{0,006} = -0,1667$ [1P]
- Kritischer Wert: $t_{\alpha;n-k-1} = t_{0,1;506-3-1} = -1,282$ [1P]
- Testentscheidung: Da $t_{empirisch} = -0,1667 > -1,282 = t_{kritisch}$ kann die Nullhypothese auf dem 10% Niveau nicht verworfen werden. [1P]

d) Welchen Wert erhält man für $\hat{\beta}_3$, wenn der Fluglärm nicht in Dezibel, sondern in Bel gemessen wird?
Hinweis: 1 Bel $\widehat{=}$ 10 Dezibel (1,5 Punkte)

- Umskalierung der unabhängigen Variable *Laerm*: $\widehat{Ln\_Preis}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot Ln\_Konz_i + \hat{\beta}_2 \cdot Zimmer_i + 10 \cdot \hat{\beta}_3 \cdot \frac{Laerm}{10}$
- $\hat{\beta}_{3,neu} = 10 \cdot \hat{\beta}_3 = 10 \cdot (-0,051) = -0,51$ [1,5P]

e) Sie vermuten, dass der Hauspreis ebenso wie der Effekt von *Laerm* vom Baujahr der Wohnung abhängt. Wie lautet die Schätzgleichung, mit der Sie diese Vermutung testen können? (Keine Berechnung notwendig) (2 Punkte)

- $Ln\_Preis_i = \beta_0 + \beta_1 Ln\_Konz_i + \beta_2 Zimmer_i + \beta_3 Laerm + \beta_4 Baujahr_i + \beta_5 Laerm \cdot Baujahr_i + u_i$ [2P]

f) Das korrigierte Bestimmtheitsmaß $\overline{R^2}$ beträgt 0,573 und die Gesamtvariation $SST = \sum_{i=1}^{506}(y_i - \bar{y})^2$ beträgt 84,582. Berechnen Sie den Standardfehler der Regression. Zeigen Sie Ihren Rechenweg. (3 Punkte)

- $\overline{R^2} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$
- $SER = \hat{\sigma} = \sqrt{\frac{(1-\overline{R^2}) \cdot SST}{n-1}}$ [2P]
- $SER = \sqrt{\frac{(1-0,573) \cdot 84,582}{506-1}} = 0,267$ [1P]

Aufgabe 2: **[13 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen der Anzahl an leiblichen Kindern und Bildung. Im Februar 2015 sammeln Sie folgende Daten für 1129 Frauen:

Kinder	Anzahl an leiblichen Kindern
Bildung	Bildungsjahre der Frau
Alter	Alter der Frau (berechnet als 2015 - Geburtsjahr der Frau)
Alter2	Alter der Frau (quadriert)
Land	Dummyvariable (=1, wenn Frau auf dem Land lebt; =0, wenn Frau in der Stadt lebt)

Sie schätzen folgendes lineares Regressionsmodell mit SPSS:

$$Kinder_i = \beta_0 + \beta_1 Bildung_i + \beta_2 Alter_i + \beta_3 Alter2_i + u_i$$

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	209,114	3	69,705	27,26	,000(a)
	Residuen	2876,395	1125	2,557		
	Gesamt	3085,509	1128	2,735		

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	-7,676	3,103	-2,47	0,014
$Bildung$	-0,131	0,018	-7,23	0,000
$Alter$	0,533	0,141	3,77	0,000
$Alter2$	-0,006	0,002	-3,60	0,000

a. Abhängige Variable: *Kinder*

a) Wie hoch ist die vorhergesagte Kinderzahl einer 33-jährigen Frau mit 18 Bildungsjahren? (1,5 Punkte)

- $\widehat{Kinder_i} = -7,676 - 0,131 \cdot 18 + 0,533 \cdot 33 - 0,006 \cdot 33^2 = 1,021$ [1,5P]

b) Wie groß ist der marginale Effekt des Alters auf die Kinderanzahl einer 37-jährigen Frau? Ab welchem Alter wird der marginale Effekt des Alters negativ? (3 Punkte)

- Marginaler Effekt: $\frac{\Delta \widehat{Kinder_i}}{\Delta Alter_i} \approx 0,533 - 2 \cdot 0,006 \cdot Alter_i$
- Marginaler Effekt (Alter 37): $0,533 - 0,012 \cdot Alter_i = 0,533 - 0,012 \cdot 37 = 0,089$ [1,5P]
- $0,533 - 0,012 \cdot Alter_i = 0 \Leftrightarrow Alter_i = \frac{0,533}{0,012} = 44,42$ (ab einem Alter von ca. 45 Jahren) [1,5P]

c) Ein Kommilitone hat die Idee, das Geburtsjahr der Frau als Kontrollvariable in das Modell aufzunehmen. Erläutern Sie knapp verbal, welche Konsequenzen dies für Ihre Schätzung hätte. (2 Punkte)

- Das Geburtsjahr lässt sich als Linearkombination des Alters und der Konstanten darstellen, demnach besteht das Problem der perfekten Multikollinearität. Die Gauss-Markov-Annahme MLR.3 ist somit verletzt. Das Modell lässt sich nur schätzen, wenn entweder das Alter oder das Geburtsjahr verwendet wird. [2P]

d) Sie vermuten, dass es Unterschiede in den Regressionsparametern für Frauen auf dem Land und Frauen in der Stadt gibt.
Sie schätzen Ihr Modell lediglich für Frauen, die auf dem Land leben, und erhalten folgenden Anova-Block:

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	68,165	3	22,722	8,31	,000(a)
	Residuen	967,591	354	2,733		
	Gesamt	1035,757	357	2,901		

Anschließend schätzen Sie Ihr Modell lediglich für Frauen, die in der Stadt leben, und erhalten folgenden Anova-Block:

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	144,524	3	48,175	19,40	,000(a)
	Residuen	1904,900	767	2,484		
	Gesamt	2049,424	770	2,662		

Um Ihre Vermutung zu testen, führen Sie einen Chow-Test (Strukturbruchtest) auf dem 1%-Signifikanzniveau durch. Beschreiben Sie knapp das Vorgehen und geben Sie Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung an. (6,5 Punkte)

- Vorgehen: 3 Modelle müssen geschätzt werden: 1) gepooltes Modell schätzen, 2) Modell für Frauen aus dem Dorf schätzen, 3) Modell für Frauen aus der Stadt schätzen $\Rightarrow$ jeweils SSR berechnen. [1P]
- Hypothesen: $H_0 : \beta_{1,0} = \beta_{2,0}, \beta_{1,1} = \beta_{2,1}, \beta_{1,2} = \beta_{2,2}, \beta_{1,3} = \beta_{2,3}$
  H_1 :mindestens ein β_j mit $j = 0,1,2,3$ unterscheidet sich zwischen den Gruppen $g = 1$ und $g = 2$ [1P]
- Teststatistik: $F = \frac{(SSR_p-(SSR_1+SSR_2))/q}{(SSR_1+SSR_2)/(n-2(k+1))} = \frac{(2876,395-(967,591+1904,900))/4}{(967,591+1904,900)/(1129-2(3+1))} = \frac{0,976}{2,562} = 0,381$ [1,5P]
  wobei $q = 4 \,\widehat{=}$ Anzahl der Restriktionen/zu testenden Parameter (alle Parameter inkl. der Konstanten) und $k = 3 \,\widehat{=}$ Anzahl der Variablen im Ausgangsmodell
- Kritischer Wert c: $c = F_{\alpha,q,n-2(k+1)} = F_{1\%,4,1121} = 3,32$ [1P]
- Vergleich Teststatistik mit kritischem Wert: $F = 0,381 < 3,32 = c$ [1P]
- Testentscheidung: Die Nullhypothese kann nicht verworfen werden, somit unterscheiden sich auf einem Signifikanzniveau von 1% die Parameter zwischen den Gruppen nicht signifikant. [1P]

Aufgabe 3: **[10 Punkte]**

Sie interessieren sich für die Determinanten von Fernbusreisen. Hierzu befragen Sie Personen zwischen 18 und 70 Jahren. Ihr Datensatz enthält folgende Informationen:

Reisen	Anzahl an Fernbusreisen pro Monat
Fernbeziehung	Dummyvariable (=1, wenn Person eine Fernbeziehung hat; =0 sonst)
Einkommen	Einkommen (gemessen in 1000 €)
*Alter*_18_30	Dummyvariable (=1, wenn Person zwischen 18 und 30 Jahre alt ist; =0 sonst)
*Alter*_31_50	Dummyvariable (=1, wenn Person zwischen 31 und 50 Jahre alt ist; =0 sonst)
*Alter*_51+	Dummyvariable (=1, wenn Person 51 Jahre oder älter ist; =0 sonst)

Sie schätzen folgendes Regressionsmodell mit SPSS und erhalten untenstehenden Output:

$$Reisen_i = \beta_0 + \beta_1 Fernbeziehung_i + \beta_2 Einkommen_i + \beta_3 Alter\_18\_30 + \beta_4 Alter\_31\_50 + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	0,133	0,094	1,417	0,156
Fernbeziehung	0,844	0,077	10,961	0,000
Einkommen	-0,388	0,041	-9,463	0,000
*Alter*_18_30	0,009	0,008	1,182	0,237
*Alter*_31_50	1,014	1,060	0,957	0,170

a. Abhängige Variable: *Reisen*

a) Interpretieren Sie die geschätzten Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_3$ inhaltlich. (2 Punkte)

- $\hat{\beta}_1$: Personen, die eine Fernbeziehung führen, führen c.p. im Durchschnitt 0,844 Busreisen pro Monat mehr aus als Personen, die keine Fernbeziehung führen. [1P]
- $\hat{\beta}_3$: Personen zwischen 18 und 30 Jahren führen c.p. im Durchschnitt 0,009 Busreisen pro Monat mehr aus als Personen, die mindestens 51 Jahre alt sind. [1P]

b) Johanna ist 20 Jahre alt, hat im Monat 700€ zur Verfügung und führt eine Fernbeziehung. Franz ist 49 Jahre alt, hat im Monat 2300€ zur Verfügung und führt keine Fernbeziehung. Wieviel € müsste Franz pro Monat weniger zur Verfügung haben, um den Fernbus genauso häufig wie Johanna zu nutzen? (4,5 Punkte)

- Unterschied in der Fernbusnutzung zwischen Johanna und Franz:
  $\Delta Reisen = (0,133 + 0,844 - 0,388 \cdot 0,7 + 0,009) - (0,133 - 0,388 \cdot 2,3 + 1,014) = 0,7144 - 0,2546 = 0,4598$ [2P]
  $\Rightarrow$ Bei gegebenen Charakteristika nutzt Johanna den Fernbus 0,4598 Fahrten pro Monat häufiger als Franz.
- Ausgleich durch Einkommen von Franz:
  $\Delta Reisen = \hat{\beta}_2 \Delta Einkommen \;\Rightarrow\; \frac{0,4598}{-0,388} = -1,185$ [2,5P]
  $\Rightarrow$ Franz müsste 1185 € weniger verdienen, um den Fernbus genauso häufig wie Johanna zu nutzen.

c) Geben Sie ein Beispiel für eine Situation, in der β_2 nicht den kausalen Effekt der Variable *Einkommen* auf die Variable *Reisen* beschreibt. Begründen Sie kurz Ihre Antwort. (2 Punkte)

- Eine solche Situation liegt vor, wenn es einen Einflußfaktor gibt, der sowohl *Reisen* beeinflusst [0,5P] als auch mit *Einkommen* korreliert [0,5P]. Beispiel: Person besitzt eigenes Auto 1/0 [1P].

d) Sie möchten testen, ob die Parameter β_3 und β_4 gemeinsam signifikant sind. Bennenen Sie das Testverfahren und formulieren Sie Null- und Alternativhypothese. (1,5 Punkte)

- F-Test auf gemeinsame Signifikanz [0,5P]
- $H_0 : \beta_3 = \beta_4 = 0$ [0,5P] und H_1 : mindestens ein Parameter $\neq 0$ [0,5P]

Aufgabe 4: **[11 Punkte]**

a) Zeigen Sie durch schrittweises Herleiten, dass folgende Gleichung gilt: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. (2 Punkte)

- $\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n \cdot \bar{x} = \sum_{i=1}^{n} x_i - n \cdot \frac{1}{n} \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0.$ [2P]

b) Betrachten Sie folgende Modellgleichungen:

Modellgleichung I: $y_i = \beta_0 + \beta_1 x_i + u_i$
Modellgleichung II: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + u_i$

Beschreiben Sie knapp die beiden Situationen, bei welchen jeweils β_1 in Modellgleichung I unterschätzt (negativ verzerrt) wäre. (2 Punkte)

- Situation I: $Cov(x_i, z_i) > 0$ und $\beta_2 < 0$ [1P]
- Situation II: $Cov(x_i, z_i) < 0$ und $\beta_2 > 0$ [1P]

c) Betrachten Sie das einfache lineare Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$. Ermitteln Sie mit Hilfe untenstehender Werte die geschätzten Regressionskoeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$. (3 Punkte)

Mittelwert von x	5,07
Mittelwert von y	49,60
Standardabweichung von x	1,83
Standardabweichung von y	21,44
Kovarianz von x und y	0,27

- $\hat{\beta}_1 = \frac{0{,}27}{1{,}83^2} = 0{,}0806$ [1,5P]
- $\hat{\beta}_0 = 49{,}60 - 0{,}0806 \cdot 5{,}07 = 49{,}19$ [1,5P]

d) Sie interessieren sich für die Determinanten von Museumsbesuchen. Hierzu befragen Sie Männer und Frauen aus den Bundesländern Bayern und Rheinland Pfalz und bilden anschließend folgende Variablen:

Museum	Anzahl Museumsbesuche pro Jahr
Mann	Dummyvariable (=1, wenn Person männlich; =0, wenn Person weiblich)
Bayern	Dummyvariable (=1, wenn Person in Bayern lebt; =0 sonst)

Nachfolgende Kreuztabelle enthält Mittelwerte der Variable *Museum* nach Geschlecht und Wohnregion:

	Männer	Frauen
Bayern	3,5	5,5
Rheinland Pfalz	2,5	4,5

Sie schätzen das Regressionsmodell $Museum_i = \beta_0 + \beta_1 Mann_i + \beta_2 Bayern_i + u_i$.
Ermitteln Sie $\hat{\beta}_0$, $\hat{\beta}_1$ und $\hat{\beta}_2$. (4 Punkte)

- $\hat{\beta}_0 = 4,5$ [1P]
- $\hat{\beta}_1 = 2,5 - 4,5 = -2$ [1,5P]
- $\hat{\beta}_2 = 3,5 - 2,5 = 1$ [1,5P]

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.	Gilt $E[u\|x] \neq 0$, dann
a	können ausgelassene Variablen den KQ-Schätzer nicht verzerren.
b	kann der KQ-Schätzer kausal interpretiert werden.
c	ist die ceteris paribus Interpretation angemessen.
d	**X** ist der KQ-Schätzer verzerrt.

2.	Sie betrachten zwei Zufallsvariablen X und Y. Welche der folgenden Aussagen ist korrekt?
a	**X** Sind X und Y negativ korreliert, so ist die Kovarianz von X und Y negativ.
b	Sind X und Y unabhängig, so ist der Korrelationskoeffizient gleich Eins.
c	Sind X und Y unabhängig, so ist der Erwartungswert von X und der Erwartungswert von Y gleich Null.
d	Sind X und Y unkorreliert, so sind X und Y auch unabhängig.

3.	Wenn Sie in einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ die Variable x_i durch $\frac{1}{3} \cdot x_i$ ersetzen,
a	verringert sich der Wert des geschätzten Steigungsparameters um den Faktor 3.
b	**X** verdreifacht sich der Wert des geschätzten Steigungsparameters.
c	verdreifacht sich der Wert der geschätzten Konstante sowie des geschätzten Steigungsparameters.
d	verringert sich der Wert der geschätzten Konstante um den Faktor 3.

4.	Schätzt man ein einfaches lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ mit der KQ-Methode, so
a	ist das korrigierte Bestimmtheitsmaß $\bar{R}^2$ stets positiv.
b	**X** werden die quadrierten vertikalen Abstände der Datenpunkte zur geschätzten Regressionsgerade minimiert.
c	ist die Summe der Residuen ungleich 0.
d	liegt jeder Punkt (y_i, x_i) auf der geschätzten Regressionsgerade.

5.	Man testet, ob ein Steigungsparameter signifikant von 0 verschieden ist, indem man
a	**X** überprüft, ob das Konfidenzintervall den Wert 0 einschließt oder nicht.
b	überprüft, ob das Konfidenzintervall den Schätzwert einschließt oder nicht.
c	überprüft, ob das Konfidenzintervall den Standardfehler einschließt oder nicht.
d	den Schätzwert mit dem kritischen Wert der t-Verteilung vergleicht.

6.	Gepoolte Querschnittsdaten
a	betrachten wiederholte Messungen für jede Beobachtungseinheit.
b	**X** betrachten für jede Beobachtungseinheit eine Messung.
c	sind Kombinationen von Zeitreihenerhebungen zu verschiedenen Erhebungszeitpunkten.
d	werden auch als Paneldaten bezeichnet.

7.	Für das Modell $lohn_i = \beta_0 + \beta_1 alter_i + \beta_2 alter_i^2 + u_i$ ergibt eine KQ-Schätzung $\hat{\beta}_1 = 0,31$ und $\hat{\beta}_2 = -0,005$. In welchem Alter wird der durchschnittliche Lohn maximiert?
a	30.
b	**X** 31.
c	32.
d	33.

8.	Ein Interaktionsterm
a	muss eine Dummyvariable beinhalten.
b	ist eine Linearkombination mehrerer Variablen.
c	ist die Summe von mindestens zwei Variablen.
d	**X** ist das Produkt von mindestens zwei Variablen.

9.	Das Problem perfekter Multikollinearität
a	kann durch Vergrößerung der Stichprobe reduziert werden.
b	führt zu inkonsistenten Schätzergebnissen.
c	führt zu ineffizienten Schätzergebnissen.
d	**X** Keine der Antworten ist korrekt.

10.	Eine KQ-Schätzung liefert $\hat{y}_i = 5 - 1,5x_i^2 + 2x_i$. Welchen Wert hat das Residuum für die Beobachtung $(y_5;x_5) = (5;-1)$?
a	0,5.
b	**X** 3,5.
c	-0,5.
d	-2,5.

11.	Sie beobachten das Gewicht für Männer und Frauen in den USA und Deutschland. Sie schätzen folgendes Modell: $Gewicht_i = \beta_1 Frau_i + \beta_2 Mann_i + \beta_3 USA_i + \beta_4 Frau_i \cdot USA_i + u_i$. Was misst der geschätzte Parameter für β_3?
a	Den Mittelwertunterschied des Gewichtes zwischen Männern und Frauen aus Deutschland.
b	**X** Den Mittelwertunterschied des Gewichtes für Männer zwischen den USA und Deutschland.
c	Den Mittelwert des Gewichtes für Männer aus den USA.
d	Den Mittelwert des Gewichtes für Männer.

12.	Wie hoch ist die Gesamtvariation SST über alle Beobachtungen gegeben $(y_1;x_1) = (0;2), (y_2;x_2) = (1;2), (y_3;x_3) = (2;2)$?
a	0.
b	1.
c	**X** 2.
d	3.

13.	Der Standardfehler eines geschätzten Koeffizienten $se(\hat{\beta}_j)$ im multiplen Modell
a	gibt an, wie stark die Beobachtungen um die Regressionsgerade streuen.
b	wird bei hoher Multikollinearität im SPSS-Output nicht ausgewiesen.
c	**X** ist eine Zufallsvariable mit Verteilungseigenschaften.
d	ist bei Vorliegen von normalverteilten Störtermen konsistent geschätzt.

14.	Sie schätzen folgende Konsumfunktion mit KQ: $log(konsum_i) = \beta_0 + \beta_1 log(einkommen_i) + u_i$. (*Hinweis*: *konsum* und *einkommen* sind in Euro gemessen.) $\hat{\beta}_1$ beträgt 0,5. Welche Aussage ist richtig?
a	Wenn das Einkommen um 2 Euro ansteigt, steigt der Konsum um 1 Euro.
b	Wenn das Einkommen um 2% ansteigt, steigt der Konsum um 10%.
c	**X** Wenn das Einkommen um 2% ansteigt, steigt der Konsum um 1%.
d	Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 5%.

15.	Sie haben Daten über Löhne von Männern und Frauen aus Ost- und Westdeutschland. Bei welcher Modellspezifikation kommt es zum Dummy-Variable-Trap?
a	**X** $Lohn_i = \beta_1 Mann_i + \beta_2 Frau_i + \beta_3 (Mann_i \cdot Ost_i) + \beta_4 (Mann_i \cdot West_i) + u_i$
b	$Lohn_i = \beta_0 + \beta_1 Mann_i + \beta_2 (Mann_i \cdot Ost_i) + \beta_3 (Frau_i \cdot Ost_i) + u_i$
c	$Lohn_i = \beta_0 + \beta_1 (Mann_i \cdot Ost_i) + \beta_2 (Frau_i \cdot Ost_i) + \beta_3 (Mann_i \cdot West_i) + u_i$
d	Keine der Antworten ist korrekt.

16.	Wenn eine diskrete Zufallsvariable X die Werte -3, 6, 5 und 2 mit konstanter Wahrscheinlichkeit $\frac{1}{4}$ annimmt, so gilt
a	**X** $E(X) = \frac{5}{2}$.
b	$E(X) = 9$.
c	$E(X^2) = 74$.
d	$E(X^2) = 10$.

17.	Um die Präzision einer Schätzung zu erhöhen
a	sollte man mit Konstante schätzen.
b	sollte man ohne Konstante schätzen.
c	sollte man die Stichprobengröße verringern.
d	**X** sollte man erklärende Variablen mit hoher Variation verwenden.

18.	Im einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ berechnet sich der Steigungsparameter $\hat{\beta}_1$ als Verhältnis der
a	Kovarianz von x und y zur Varianz von y.
b	**X** Kovarianz von x und y zur Varianz von x.
c	Kovarianz von x und y zur Standardabweichung von x.
d	Kovarianz von x und y zum Standardfehler von x.

19.	Im linearen Wahrscheinlichkeitsmodell
a	gibt die Prognose $\hat{y}$ die Wahrscheinlichkeit für das Ereignis $y = 0$ an.
b	wird die abhängige Variable in Prozentpunkten gemessen.
c	**X** wird eine binär kodierte abhängige Variable verwendet.
d	können kausale Effekte nicht identifiziert werden.

20.	Wird ein irrelevanter Regressor (d.h. ein Regressor, dessen wahrer Koeffizient exakt gleich 0 ist) einem multiplen Regressionsmodell hinzugefügt, so
a	sinkt die Varianz des Störterms.
b	steigt die Präzision der Schätzung.
c	**X** sinkt das korrigierte Bestimmtheitsmaß $\bar{R}^2$.
d	verändern sich die Regressionsergebnisse nicht.

21.	Ein Schätzer $\hat{\beta}_1$ für den unbekannten Parameter β_1 im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ ist unverzerrt, wenn
a	$N \to \infty$.
b	$plim(\hat{\beta}_1) = \beta_1$.
c	u normalverteilt ist.
d	**X** $E[\hat{\beta}_1] = \beta_1$.

22.	Wird die Nullhypothese eines Chow-Tests abgelehnt, so
a	liegt ein Endogenitätsproblem vor.
b	liegen keine signifikanten Unterschiede im Erklärungsgehalt für verschiedene Gruppen vor.
c	fallen Regressionskoeffizienten für verschiedene Gruppen gleich aus.
d	**X** fallen Regressionskoeffizienten für verschiedene Gruppen unterschiedlich aus.

23.	Im Modell $log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 log(x_{2i}) + u_i$
a	wird β_1 als Elastizität interpretiert.
b	**X** wird β_2 als Elastizität interpretiert.
c	wird β_2 als Semi-Elastizität interpretiert.
d	werden sowohl β_1 als auch β_2 als Semi-Elastizitäten interpretiert.

24.	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 nicht mitgeschätzt, so
a	kann man das R^2 als Maß der Schätzgüte verwenden.
b	sinkt die Zahl der Freiheitsgrade.
c	ändern sich die Konfidenzintervalle der Steigungsparameter nicht.
d	**X** keine der Antworten ist korrekt.

25.	Ein nicht-linearer Zusammenhang zwischen zwei Variablen
a	kann mit der Kovarianz korrekt abgebildet werden.
b	**X** kann mit dem Korrelationskoeffizienten nicht korrekt abgebildet werden.
c	kann nicht in einem multiplen Regressionsmodell abgebildet werden.
d	Keine der Antworten ist korrekt.

26.	Ein Typ 2-Fehler tritt auf, wenn man die Nullhypothese
a	**X** nicht verwirft, obwohl diese falsch ist.
b	verwirft, obwohl diese zutrifft.
c	verwirft, obwohl diese falsch ist.
d	nicht verwirft, obwohl diese zutrifft.

27.	Wenn die Störterme einer multiplen Regression nicht normalverteilt sind, dann
a	ist der KQ-Schätzer inkonsistent.
b	sind t- und F-Tests auch für kleine Stichproben gültig.
c	**X** sind t- und F-Tests für kleine Stichproben nicht gültig.
d	können keine Hypothesentests durchgeführt werden.

28.	Das korrigierte Bestimmtheitsmaß $\bar{R}^2$ im einfachen linearen Regressionsmodell
a	**X** berücksichtigt die zur Schätzung benötigten Freiheitsgrade.
b	nimmt ausschließlich negative Werte an.
c	nimmt ausschließlich positive Werte an.
d	gibt das Verhältnis von unerklärter Variation zur Gesamtvariation an.

29.	Ein Omitted Variable Bias entsteht im Modell $y_i = \beta_0 + \beta_1 x_{1i} + u_i$, wenn eine Variable x_2
a	den Erwartungswert $E[x_2] \neq 0$ hat.
b	mit β_1 korreliert.
c	mit β_1 und u korreliert.
d	**X** mit x_1 und u korreliert.

30.	Der marginale Effekt von x_1 im Modell $y_i = \beta_0 + \beta_1 log(x_{1i}) + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i$ lautet
a	$\beta_1 + \beta_3 x_2$.
b	$\beta_1 x_1 + \beta_3 x_2$.
c	**X** $\beta_1 \frac{1}{x_1} + \beta_3 x_2$.
d	$\beta_1 \frac{1}{x_1} + \beta_3 \frac{1}{x_1} x_2$.

31.	Bei konsistenten Schätzverfahren
a	ist der Schätzer unverzerrt.
b	**X** kann der Erwartungswert des Schätzers vom wahren Wert abweichen.
c	liegt der Wahrscheinlichkeitsgrenzwert des Schätzers umso näher am wahren Wert, je kleiner die Stichprobe ist.
d	ist der Dummy-Variable-Trap bei steigender Stichprobengröße kein Problem.

32.	In einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$
a	ist β_1 positiv, wenn $corr(x,y) < 0$.
b	**X** ist β_1 gleich Null, wenn $corr(x,y) = 0$.
c	ist β_1 negativ, wenn $corr(x,y) > 0$.
d	keine der Antworten ist korrekt.

33.	Unter den Gauss-Markov Annahmen
a	**X** ist der KQ-Schätzer konsistent.
b	ist KQ das beste, erwartungstreue, nichtlineare Schätzverfahren.
c	folgt der KQ-Schätzer der F-Verteilung.
d	ist die Varianz des Störterms heteroskedastisch.

34.	Eine Typ 1-Fehlerwahrscheinlichkeit beschreibt die Wahrscheinlichkeit,
a	eine zutreffende Nullhypothese nicht zu verwerfen.
b	**X** eine zutreffende Nullhypothese zu verwerfen.
c	eine falsche Nullhypothese zu verwerfen.
d	eine falsche Nullhypothese nicht zu verwerfen.

35.	Der Standardfehler der Regression (SER) ist ein Schätzer für
a	die Streuung der Konstante.
b	die Streuung der einzelnen Steigungsparameter.
c	**X** die Streuung der auf x bedingten Störterme.
d	die Streuung der auf x bedingten Variablen.

36.	F-Tests
a	können nicht verwendet werden, um Hypothesen für einzelne Parameter zu testen.
b	**X** können verwendet werden, um Hypothesen für mehrere Parameter zu testen.
c	können bei mehr als 10 Nennerfreiheitsgraden nicht durchgeführt werden.
d	können bei mehr als 120 Zählerfreiheitsgraden nicht durchgeführt werden.

37.	Durch Umskalierung der abhängigen Variable
a	**X** verändert sich das Konfidenzintervall der Konstante.
b	ändert sich der Wert des korrigierte Bestimmtheitsmaßes $\bar{R}^2$.
c	ändert sich die Anzahl der verwendeten Freiheitsgrade.
d	können die p-Werte sinken.

38.	In einem einfachen linearen Regressionsmodell gibt der Steigungsparameter die Semi-Elastizität an, wenn
a	lediglich die unabhängige Variable logarithmiert ist.
b	**X** lediglich die abhängige Variable logarithmiert ist.
c	sowohl abhängige als auch unabhängige Variable logarithmiert sind.
d	weder abhängige noch unabhängige Variable logarithmiert sind.

39.	Wenn Homoskedastie vorliegt
a	ist die Varianz des Fehlerterms u für alle Beobachtungen unterschiedlich.
b	**X** können die Parameter unverzerrt geschätzt werden.
c	sind die Standardfehler der Parameter normalverteilt.
d	ist das korrigierte Bestimmtheitsmaß $\bar{R}^2$ nicht mehr gültig.

40.	Die Modelle $y_i = \beta_0 + \beta_1 \sqrt{x_{1i}} + u_i$ und $y_i = \beta_0 + \beta_1 \sqrt{x_{1i}} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + u_i$
a	sind multiple lineare Modelle.
b	**X** sind genestet.
c	sind nicht linear in den Parametern.
d	lassen sich mit der KQ-Methode nicht schätzen.