Bachelorprüfung SS 2015 - 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: **[8 Punkte]**

a) Stellen Sie formal für das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ das Minimierungsproblem für die Herleitung der KQ-Schätzer auf. (1,5 Punkte)

$$min \sum_{i=1}^{N} \hat{u}_i^2$$
$$= \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$$

b) Stellen Sie formal für das Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ die Bedingungen erster Ordnung auf. (2 Punkte)

- $\sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i}) = 0$
- $\sum_{i=1}^{N} x_i \cdot (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i}) = 0$

c) Betrachten Sie das einfache lineare Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$. Ermitteln Sie mit Hilfe untenstehender Werte den geschätzten Regressionskoeffizienten $\hat{\beta}_0$. Nehmen Sie an, dass der geschätzte Wert für den Steigungsparameter 0,5 beträgt (2,5 Punkte).

y	1	2	3
x	5	5	8

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $\bar{y} = 2$
- $\bar{x} = 6$
- $\hat{\beta}_0 = 2 - 0,5 \cdot 6 = -1$

d) Berechnen Sie die Kovarianz der Zufallsvariablen y und x aus Aufgabenteil c). (2 Punkte)

- $Cov(y,x) = E(x-\bar{x})(y-\bar{y}) = \frac{1}{n-1} \sum_{i=1}^{n} x_i (y_i - \bar{y})$
  $= \frac{1}{2}(5 \cdot (1-2) + 5 \cdot (2-2) + 8 \cdot (3-2)) = \frac{1}{2}(-5+0+8) = 1,5$

Aufgabe 2: **[10 Punkte]**

Sie untersuchen den Zusammenhang zwischen der durchschnittlichen Lebenserwartung und dem Wohlstand. Es stehen Ihnen folgende Informationen über 200 Länder zur Verfügung: Lebenserwartung in Jahren (*LE*), Bruttoinlandsprodukt pro Kopf in USD (*BIP*), Arbeitslosenquote in Prozentpunkten (*ALQ*). Betrachten Sie folgende Modellspezifikationen:

A: $LE_i = \beta_0 + \beta_1 BIP_i + u_i$

B: $LE_i = \beta_0 + \beta_1 BIP_i + \beta_2 ALQ_i + e_i$

C: $log(LE_i) = \beta_0 + \beta_1 BIP_i + \beta_2 ALQ_i + \varepsilon_i$

D: $log(LE_i) = \beta_0 + \beta_1 BIP_i + \beta_2 BIP_i^2 + \mu_i$

a) Sie schätzen das Modell C mittels KQ und erhalten $\hat{\beta}_2 = -0,011$. Interpretieren Sie den geschätzten Koeffizienten inhaltlich. (1 Punkt)

> Steigt die Arbeitslosenquote um einen Prozentpunkt, so sinkt die Lebenserwartung c.p. im Mittel um 1,1%.

b) Zum Vergleich der Modellspezifikationen möchten Sie das Bestimmtheitsmaß R^2 und das korrigierte $\overline{R^2}$ heranziehen. Sie fragen Ihren Kommilitonen um Rat und erhalten folgende Vorschläge, welches Maß für die folgenden Modellvergleiche verwendet werden kann:

i. R^2 für Vergleich zwischen A und B
ii. R^2 für Vergleich zwischen C und D
iii. $\overline{R^2}$ für Vergleich zwischen B und C

Entscheiden Sie jeweils, ob Ihr Kommilitone Recht hat. Begründen Sie knapp Ihre Antwort. (4,5 Punkte)

> i. R^2 geeignet , da Modelle A und B genestet sind
>
> i. (Alternativ) R^2 kann nicht verwendet werden, da keine Freiheitsgradkorrektur vorgenommen wird.
>
> ii. R^2 geeignet, da Modelle C und D die gleiche Anzahl an Parametern berücksichtigen (obwohl die Modelle nicht genestet sind)
>
> iii. $\overline{R^2}$ nicht geeignet , da Modelle B und C unterschiedliche abhängige Variablen haben (*alternative Begründung:* da sich die SST unterscheiden)

c) Eine KQ-Schätzung des Modells B ergibt $\overline{R^2} = 0,675$. Berechnen Sie das R^2 aus dieser Schätzung. (3,5 Punkte)

> Umformen $\overline{R^2} = 1-(1-R^2)\cdot\frac{n-1}{n-k-1}$ zu $R^2 = 1-(1-\overline{R^2})\cdot\frac{n-k-1}{n-1}$
>
> Einsetzen: $R^2 = 1-(1-0,675)\cdot\frac{200-2-1}{200-1}$ (jeweils 0,5P für $\overline{R^2}$, n=200, k=2)
>
> Ausrechnen: $R^2 = 0,678$

d) Eine KQ-Schätzung des Modells C ergibt $R^2 = 0,728$. Interpretieren Sie diesen Wert. (1 Punkt)

> Das Modell erklärt 72,8% der Variation in der log-Lebenserwartung.

Aufgabe 3: **[16 Punkte]**

Sie wollen untersuchen, wie stark der Lernaufwand sich auf Mathenoten auswirkt. Hierfür steht Ihnen ein Datensatz von 531 Zehntklässlern von Nürnberger Gymnasien aus dem Jahr 2012 zur Verfügung.

$Note$	Mathenote am Ende des Schuljahres (von 1="sehr gut" bis 6="ungenügend")
$StdLernen$	wöchentlicher Lernaufwand für Mathe in Stunden
$StdLernen2$	wöchentlicher Lernaufwand für Mathe in Stunden (quadriert)
$EuroNachhilfe$	Gesamtausgaben für Mathenachhilfe im betreffenden Schuljahr in Euro
$LogEuroNachhilfe$	Gesamtausgaben für Mathenachhilfe im betreffenden Schuljahr in Euro (in logarithmierter Form)

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Note_i = \beta_0 + \beta_1 StdLernen_i + \beta_2 StdLernen2_i + \beta_3 LogEuroNachhilfe + u_i \quad (1)$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	2,942	1,534	1,918	0,056
$StdLernen$	-0,359	0,121	-2,967	0,003
$StdLernen2$	0,012	0,011	1,091	0,276
$LogEuroNachhilfe$	-0,053	0,078	0,681	0,496

a. Abhängige Variable: $Note$

a) Berechnen Sie die erwartete Note für einen Schüler, der 2 Stunden pro Woche Mathe lernt und für den im Laufe des Schuljahres 700 Euro für Nachhilfe ausgegeben wurden. (2 Punkte)

$$\widehat{Note_i} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 2 + \hat{\beta}_2 \cdot 2^2 + \hat{\beta}_3 \cdot ln(700) = 2,942 - 0,718 + 0,048 - 0,347 = 1,925$$

b) Berechnen Sie die erwartete Notendifferenz, wenn ein Schüler statt 2 Stunden pro Woche, 3 Stunden pro Woche lernt. Interpretieren Sie das Ergebnis inhaltlich. (3 Punkte)

- 2 Stunden lernen: $\hat{\beta}_1 \cdot StdLernen + \hat{\beta}_2 \cdot StdLernen2_{i0} = -0,359 \cdot 2 + 0.012 \cdot 2^2 = -0,670$
  3 Stunden lernen: $\hat{\beta}_1 \cdot StdLernen + \hat{\beta}_2 \cdot StdLernen2_{i1} = -0,359 \cdot 3 + 0.012 \cdot 3^2 = -0,969$
  Differenz: $-0,299$
- Die erwartete Note verbessert sich um ca. 0,3 Notenstufen.

c) Ab welchem wöchentlichen Lernaufwand wird es, laut dieser Schätzung, für einen Schüler kontraproduktiv, noch mehr zu lernen? (1,5 Punkte)

- $\frac{\partial \widehat{Note}}{\partial StdLernen} = 0 \Leftrightarrow \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot StdLernen_i = 0 \Leftrightarrow -0.359 + 2 \cdot 0,012 \cdot StdLernen = 0$

  $\Leftrightarrow StdLernen = 14,958$ Ab einem wöchentlichen Lernaufwand von 14,958 Stunden.

d) Welche Werte würden die Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ annehmen, wenn der Lernaufwand in Minuten angegeben würde? (Runden Sie auf die dritte Nachkommastelle.) (2 Punkte)

- $\frac{\hat{\beta}_1}{60} = \frac{-0{,}359}{60} = -0{,}006$
- $\frac{\hat{\beta}_2}{60^2} = \frac{0{,}012}{3600} = 0{,}000$

e) Interpretieren Sie den Koeffizienten von $LogEuroNachhilfe$ inhaltlich und statistisch. (2 Punkte)

- Inhaltliche Interpretation: Erhöhung der Ausgaben für Mathenachhilfe von 1%, verringert die Note (verbessert sie also) c.p. im Mittel um 0,00053 Einheiten.
- Statistische Interpretation: $p_{\hat{\beta}_3} = 0{,}496 > 0{,}1$. Der Koeffizient ist am 10%-Niveau nicht statistisch signifikant von Null verschieden.

f) Welche Auswirkungen hat es auf die Schätzung, wenn es in der Stichprobe Schüler gibt, die keine Nachhilfe bekommen. Beschreiben Sie das Problem verbal und erläutern Sie was passiert, wenn man es unberücksichtigt lässt? (1 Punkt)

- Wenn ein Schüler keine Nachhilfe bekommt, hat die Variable $EuroNachhilfe$ den Wert Null. Man kann den Wert Null nicht logarithmieren und so fallen diese Beobachtungen aus der Schätzung bzw. können nicht berücksichtigt werden.

g) Sie wollen testen, ob die Koeffizienten der Variablen $StdLernen$ und $StdLernen2$ gemeinsam signifikant sind. Benennen Sie das Testverfahren, stellen Sie die dafür geeigneten Hypothesen auf und geben Sie den kritischen Wert der Teststatistik am 10%-Signifikanzniveau an. (Hinweis: Die Teststatistik muss nicht berechnet werden.) (2,5 Punkte)

- Testverfahren: F-Test
- H_0: $\beta_1 = \beta_2 = 0$
- H_1: mindestens ein $\beta_j \neq 0$ mit $j = 1,2$
- $F_{kritisch} = F_{0,1;2;528} = 2{,}30$

h) Sie schätzen das neue Modell

$$Note_i = \beta_0 + \beta_1 LogEuroNachhilfe_i + u_i \quad (2)$$

Sie befürchten, die Schätzergebnisse könnten verzerrt sein, weil das Modell das Einkommen der Eltern nicht berücksichtigt. Erläutern Sie anhand des hypothetischen Koeffizienten des Elterneinkommens ($\hat{\beta}_2 < 0$) und der von Ihnen unterstellten Kovarianz zwischen Elterneinkommen ($EltEK$) und $LogNachhilfe$, in welche Richtung der Koeffizient von $LogNachhilfe$ in der Schätzung von Modell (2) verzerrt ist. (2 Punkte)

- $Cov(EltEK, LogNachhilfe) > 0$ (Alternativ: $Cov(EltEK, LogNachhilfe) > 0$) und $\hat{\beta}_2 < 0$
- $\oplus \cdot \ominus = \ominus$ (oder alternative Erläuterung)
  Der Koeffizient ist nach unten verzerrt.

Aufgabe 4: **[16 Punkte]**

Sie interessieren sich für die Determinanten von wöchentlicher Telefonierzeit. Ihr Datensatz aus dem Jahr 2014 enthält folgende Informationen für 732 Studierende:

$Telefonierzeit$	Durchschnittliche wöchentliche Telefonierzeit (in Minuten)
$Freunde$	Anteil der Freunde, die höchstens 20km vom Wohnort entfernt wohnen (0-100%)
$Frau$	Geschlechtsdummy (1=Frau, 0=Mann)
$Freizeit$	Freizeit pro Woche (in Stunden)
$Flatrate$	Dummyvariable zum Handytarif (1=Tarif mit Flatrate, 0=Tarif ohne Flatrate)

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Telefonierzeit_i = \beta_0 + \beta_1 Freunde_i + \beta_2 Frau_i + \beta_3 Freizeit_i + \beta_4 Flatrate_i + \beta_5 Frau_i \cdot Freizeit_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	,594	2,479	,240	,811
$Freunde$	-2,761	1,295	-2,132	,033
$Frau$	72,360	26,520	2,729	,007
$Freizeit$	,918	,123	7,463	,000
$Flatrate$	36,852	24,568	1,932	,054
$Frau \cdot Freizeit$	3,562	1,245	2,861	0,004

a. Abhängige Variable: $Telefonierzeit$

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

- Ein Anstieg des Anteils der Freunde im Umkreis von 20km um einen Prozentpunkt führt c.p. im Mittel zu einer Verkürzung der Telefonierzeit um 2,761 Minuten.
- Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden auf dem 5%-Niveau (p-Wert:0,033<0,05).

b) Bestimmen Sie das 99%-Konfidenzintervall für β_1 und interpretieren Sie Ihr Ergebnis. (Runden Sie alle Zwischenschritte auf die <u>dritte</u> Nachkommastelle.) (3,5 Punkte)

- $t_{\frac{0,01}{2};732-5-1} = 2,576$
- $[-2,761 - t_{\frac{0,01}{2};732-5-1} \cdot 1,295; -2,761 + t_{\frac{0,01}{2};732-5-1} \cdot 1,295]$
- $=[-2,761 - 2,576 \cdot 1,295; -2,761 + 2,576 \cdot 1,295]$
- $=[-6,097; 0,575]$
- Für wiederholte Stichproben liegt in 99% der Fälle der wahre Wert innerhalb der auf diese Weise berechneten Intervallgrenzen.

c) Testen Sie auf dem 10%-Signifikanzniveau, ob Studenten mit Telefonflatrate in der Woche mindestens 30 Minuten länger telefonieren als Studenten ohne Flatrate. Geben Sie Testvefahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung an. Trifft die Aussage zu? (4,5 Punkte)

- Testverfahren: einseitiger t-Test
- Hypothesen: $H_0 : \beta_4 \leq 30, H_1 : \beta_4 > 30$
- Teststatistik: $t = \frac{\hat{\beta}_4 - 30}{se(\hat{\beta}_4)} = \frac{36{,}852-30}{24{,}568} = 0{,}279$
- Kritischer Wert: $t_{\alpha;n-k-1} = t_{0,10;732-5-1} = 1{,}282$
- Testentscheidung: Da $t_{empirisch} = 0{,}279 < 1{,}282 = t_{kritisch}$ kann die Nullhypothese auf dem 10%-Niveau nicht verworfen werden. Die Aussage trifft nicht zu.

d) Mit welcher Vermutung lässt sich die Aufnahme des Interaktionsterms ($Frau_i \cdot Freizeit_i$) in das Modell begründen? Geben Sie zwei Möglichkeiten an. (2 Punkte)

- Es wird vermutet, dass sich eine Stunde mehr Freizeit abhängig vom Geschlecht unterschiedlich auf die Telefonierzeit auswirkt.
- Es wird vermutet, dass der Geschlechterunterschied in der Telefonierzeit abhängig von der Freizeit ist.

e) Berechnen Sie den marginalen Effekt der Freizeit auf die Telefonierzeit für Männer und Frauen und interpretieren Sie jeweils das Ergebnis. (4 Punkte)

- Marginaler Effekt der Freizeit: $\frac{\Delta \widehat{Telefonierzeit}}{\Delta Freizeit} = (\hat{\beta}_3 + \hat{\beta}_5 Frau_i)$
- Marginaler Effekt der Freizeit für Frauen: $(\hat{\beta}_3 + \hat{\beta}_5 \cdot 1) = 0{,}918 + 3{,}562 = 4{,}480$
- Marginaler Effekt der Freizeit für Männer: $(\hat{\beta}_3 + \hat{\beta}_5 \cdot 0) = \hat{\beta}_3 = 0{,}918$
- Interpretation: Für Frauen geht eine Stunde mehr Freizeit in der Woche c.p. im Mittel einher mit 4,48 Minuten mehr Telefonierzeit in der Woche. Für Männer ist eine Stunde mehr Freizeit c.p. im Durchschnitt verbunden mit 0,92 Minuten mehr Telefonierzeit.

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.	Unverzerrtheit und Effizienz als Eigenschaften des KQ-Schätzverfahrens
a	werden in Abhängigkeit von der Stichprobengröße abgeleitet.
b	können für kleine Stichproben nicht nachgewiesen werden.
c	**X** werden als finite sample properties bezeichnet.
d	werden als large sample properties bezeichnet.

2.	Wenn die Störterme einer multiplen Regression asymptotisch normalverteilt sind, dann
a	können keine Hypothesentests durchgeführt werden.
b	ist der KQ-Schätzer inkonsistent.
c	sind t- und F-Tests in einer beliebigen Stichprobengröße gültig.
d	**X** sind t- und F-Tests für große Stichproben approximativ gültig.

3.	Bei konsistenten Schätzverfahren
a	ist perfekte Multikollinearität bei steigender Stichprobengröße kein Problem.
b	ist der Erwartungswert des Schätzers gleich dem wahren Wert.
c	**X** liegt der Wahrscheinlichkeitsgrenzwert des Schätzers umso näher am wahren Wert, je größer die Stichprobe ist.
d	ist der Schätzer unverzerrt.

4.	Unter den Gauss-Markov Annahmen
a	sind die Störterme normalverteilt.
b	folgt der KQ-Schätzer der χ^2-Verteilung.
c	ist der KQ-Schätzer inkonsistent.
d	**X** ist KQ das beste, erwartungstreue, lineare Schätzverfahren.

5.	Gilt $E[u\|x] = 1$, 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 1 annehmen.

6.	Schätzt man ein multiples lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ mit der KQ-Methode, so
a	ergeben sich 2 Bedingungen erster Ordnung.
b	beschreibt β_2 den Partialeffekt von x_2 bei gegebenen β_1 und β_0.
c	**X** beschreibt β_1 den Partialeffekt von x_1 bei gegebenem x_2.
d	liegt jeder Punkt (y_i, x_{1i}, x_{2i}) auf der geschätzten Regressionsgerade.

7.	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.

8.	Schätzt man ein korrekt spezifiziertes multiples lineares Regressionsmodell mit der KQ-Methode, so
a	ist die Stichprobenvarianz jeder unabhängigen Variable Null.
b	**X** ist die Stichprobenkovarianz zwischen jeder unabhängigen Variable und dem Vorhersagefehler Null.
c	ist die Stichprobenkorrelation zwischen den vorhergesagten Werten und den KQ-Residuen Eins.
d	ist der Stichprobenmittelwert der quadrierten vertikalen Abstände der Datenpunkte zur geschätzten Regressionsgerade Eins.

9.	Um die Präzision einer Schätzung zu erhöhen, sollte man
a	**X** die Stichprobengröße vergrößern.
b	die Stichprobengröße verringern.
c	mit Konstante schätzen.
d	ohne Konstante schätzen.

10.	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	ist gewährleistet, dass der Mittelwert der Residuen Null ist.
b	**X** so spricht man von einer Regression durch den Ursprung.
c	kann man das R^2 als Maß der Schätzgüte verwenden.
d	sind die Steigungsparameter unverzerrt.

11.	Wird das Modell $y_i = \beta_0 + \beta_1 log(x_i) + \beta_2 x_i + u_i$ geschätzt, so
a	besteht das Problem perfekter Multikollinearität.
b	lässt sich der marginale Effekt von x_i angeben mit $\beta_1 \frac{1}{x}$.
c	lässt sich der marginale Effekt von x_i angeben mit $\frac{1}{\beta_1} + \beta_2$.
d	**X** können Beobachtungen mit $x_i = 0$ nicht berücksichtigt werden.

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

13.	Im Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ besteht das Problem perfekter Kollinearität, wenn
a	**X** $x_1 + x_2 = 1$.
b	$x_1 = x_2^2$.
c	$Var(x_1) = Var(x_2)$.
d	$Cov(x_1, x_2) = 0$.

14.	Wenn Heteroskedastie vorliegt, ist
a	eine KQ-Schätzung nicht durchführbar.
b	der KQ-Schätzer BLUE.
c	der KQ-Schätzer ineffizient und verzerrt.
d	**X** der KQ-Schätzer ineffizient, aber erwartungstreu, wenn $E[u_i \| x_i] = E[u_i] = 0$.

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

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

17.	Die t-Verteilung konvergiert gegen die Normalverteilung bei steigender
a	**X** Stichprobengröße.
b	Kovarianz von abhängiger Variable und den Regressoren.
c	Anzahl von Regressoren.
d	Varianz der abhängigen Variable.

18.	Die Varianz der Zufallsvariable Y mit dem Mittelwert μ_y^2 lässt sich errechnen mit der Formel
a	$\sum_{i=1}^{n} y_i(x_i - \bar{x})$
b	$\frac{1}{\mu_y} \cdot \sum_{i=1}^{n} y_i$
c	**X** $E(Y^2) - \mu_y^2$
d	$(Y - \mu_y)/\sigma_y$

19.	Welcher der folgenden Ausdrücke über die Zufallsvariable Z mit der Varianz σ_z^2 und die Konstante c ist wahr?
a	$Var(Z \cdot c) = \sigma_z^2 \cdot c$
b	**X** $Var(Z \cdot c) = \sigma_z^2 \cdot c^2$
c	$Var(Z \cdot c) = \sigma_z \cdot c$
d	$Var(Z \cdot c) = \sigma_z \cdot c^2$

20.	Paneldatensätze
a	bezeichnet man auch als gepoolte Querschnitte.
b	müssen für alle Beobachtungseinheiten die gleiche Anzahl an Beobachtungen haben.
c	haben weniger Beobachtungen als Querschnittsdatensätze.
d	**X** ermöglichen die Messung von Effekten, die mit zeitlicher Verzögerung auftreten.

21.	Mit welchen Regressionen kann man das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ auf Multikollinearität testen?
a	Eine Regression von $\hat{\beta}_1$ auf $\hat{\beta}_2$ und umgekehrt.
b	Eine Regression von $\hat{u}_i$ auf $\hat{y}_i$ und umgekehrt.
c	**X** Eine Regression von x_{1i} auf x_{2i} und umgekehrt.
d	Eine Regression von y_i auf x_{1i} und eine Regression von y_i auf x_{2i}.

22.	Die Konsistenz der Schätzung $y_i = \beta_0 + \beta_1 x_i + u_i$ könnte beeinträchtigt sein, wenn
a	Heteroskedastie vorliegt.
b	**X** die Variable z, für die gilt $Cov(y,z) \neq 0$ und $Cov(x,z) \neq 0$, nicht berücksichtigt wird.
c	die Variable z, für die gilt $Cov(y,z) = 0$ und $Cov(x,z) \neq 0$, nicht berücksichtigt wird.
d	$\sigma_u^2 > 0$.

23.	Wenn für die Regression $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ gilt $\hat{\beta}_1 < 0$ und $\hat{\beta}_2 > 0$, dann
a	steigt y zunächst mit x und fällt nach dem Maximum.
b	**X** fällt y zunächst mit x und steigt nach dem Minimum.
c	steigt y kontinuierlich mit x, für $x \in [-\infty, \infty]$.
d	fällt y kontinuierlich mit x, für $x \in [-\infty, \infty]$.

24.	Die Aufnahme irrelevanter Variablen in ein Regressionsmodell
a	verringert den Wert des R^2.
b	erhöht den Wert des angepassten Bestimmtheitsmaßes.
c	verringert die Konfidenzintervalle um die Parameterschätzer.
d	**X** verringert die Effizienz der Schätzung.

25.	Für das Modell $y_i = \beta_0 + \beta_1 log(x_i) + u_i$, kann man den Steigungsparameter interpretieren als
a	Semielastizität der abhängigen Variable in Bezug auf die erklärende Variable.
b	Elastizität der abhängigen Variable in Bezug auf die erklärende Variable.
c	**X** absolute Änderung der abhängigen Variable, wenn die erklärende Variable um 1% steigt.
d	absolute Änderung der abhängigen Variable, wenn die erklärende Variable um einen Prozentpunkt steigt.

26.	Wenn das Residuum einer KQ-Schätzung für die sechste Beobachtung $\hat{u}_6 = -3$,
a	muss der empirische Wert der abhängigen Variable für diese Beobachtung größer sein als der Erwartungswert.
b	**X** muss der empirische Wert der abhängigen Variable für diese Beobachtung kleiner sein als der Erwartungswert.
c	ist die Annahme $E(u) = 0$ ungültig.
d	ist die Schätzung verzerrt.

27.	Für einen t-Test mit $H_1 : \beta_1 < 0$ ergibt sich ein kritischer Wert von $t_{0.10,30-4-1} = 1,316$ und eine Teststatistik mit $t_{\hat{\beta}_1} = -1,257$. 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	Die Alternativhypothese kann auf dem 10%-Signifikanzniveau nicht angenommen werden.
d	Die Alternativhypothese kann auf dem 10%-Signifikanzniveau angenommen werden.

28.	Sie schätzen folgendes Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$. Dabei ist x_{2i} eine 0/1 kodierte Dummyvariable. Wie ändert sich der Koeffizient $\hat{\beta}_2$ (zuvor $\hat{\beta}_2 = 0,74$), wenn Sie diese Variable umkodieren (1/0)?
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 0,26.
d	**X** Der neue Koeffizient ist -0,74.

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

30.	Wird die Nullhypothese des Chow-Testes verworfen, dann
a	liegt perfekte Multikollinearität zwischen mindestens zwei Steigungsparametern vor.
b	ist dies ein Hinweis für Heteroskedastie in den Daten.
c	ist mindestens ein Koeffizient einer erklärenden Variable signifikant von Null verschieden.
d	**X** unterscheiden sich die Steigungsparameter signifikant für verschiedene Gruppen.

31.	Durch eine Erhöhung der Stichprobengröße,
a	**X** wird das Konfidenzintervall enger.
b	steigt das R^2.
c	wird ein Problem mit Heteroskedastie gelöst.
d	verändern sich die Standardfehler der Koeffizienten nicht.

32.	Sie führen nacheinander einen rechtsseitigen und einen beidseitigen t-Test durch. Wie unterscheiden sich die kritischen Werte, 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	**X** Der kritische Wert des beidseitigen Tests ist größer.
c	Der kritische Wert ist in beiden Tests gleich groß.
d	Die Antwort hängt von dem Vorzeichen des Koeffizienten ab.

33.	Für einen geschätzten Koeffizienten erhalten Sie im Rahmen eines Signifikanztests einen t-Wert von 1,750 und einen p-Wert von 0,081. Welche der folgenden Aussagen kann damit über eine t-verteilte Zufallsvariable getroffen werden?
a	Der Koeffizient der betrachteten Variable ist signifikant auf dem 5%-Niveau.
b	**X** Der Koeffizient der betrachteten Variable ist signifikant auf dem 10%-Niveau.
c	Der Koeffizient ist mit einer Irrtumswahrscheinlichkeit von 8,1% insignifikant.
d	Der Erklärungsgehalt des Modells steigt durch die Aufnahme dieser Variable um 1,75%.

34.	Mit einem F-Test vergleichen Sie zwei Modelle (M1 und M2) mit der gleichen abhängigen Variable. M1 enthält die erklärende Variablen: Temperatur, Alter, Einkommen und Gewicht. M2 berücksichtigt nur Einkommen als erklärende Variable. Welche Aussage trifft zu?
a	M1 ist das restringierte Modell.
b	M1 enthält drei Regressanden.
c	**X** Das restringierte Modell berücksichtigt einen Steigungsparameter.
d	Das unrestringierte Modell enthält drei Achsenabschnittsparameter.

35.	Die F-Verteilung
a	ist linksschief.
b	entspricht asymptotisch der t-Verteilung.
c	kann negative Werte annehmen.
d	**X** ist nicht symmetrisch.

36.	Wie viele Dummyvariablen müssen in das Modell 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

37.	In dem Modell $y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$ mit zwei 0/1 kodierten, unkorrelierten Dummyvariablen D_{1i} und D_{2i}
a	weist $\hat{\beta}_2$ für Beobachtungen mit $D_{1i} = 1$ und $D_{2i} = 1$ den Mittelwert von y_i aus.
b	ist $\hat{\beta}_1 + \hat{\beta}_2 = 1$.
c	**X** liegt kein *Dummy-Variable-Trap* vor.
d	muss zusätzlich der Interaktionsterm $D_{1i} \cdot D_{2i}$ aufgenommen werden.

38.	Ein Typ-1 Fehler liegt vor, wenn
a	H_0 nicht verworfen wird, obwohl sie zutrifft.
b	H_0 verworfen wird, obwohl sie nicht zutrifft.
c	**X** H_0 verworfen wird, obwohl sie zutrifft.
d	H_0 nicht verworfen wird, obwohl sie nicht zutrifft.

39.	Wenn eine Schätzung vom *Dummy-Variable-Trap* betroffen ist,
a	**X** liegt ein Problem mit Multikollinearität vor.
b	sind die Koeffizienten dennoch eindeutig bestimmbar.
c	muss eine zusätzliche Dummy-Variable in das Modell aufgenommen werden.
d	muss ein Interaktionsterm aufgenommen werden.

40.	Sie möchten überprüfen, ob es im Modell $y_i = \beta_0 + \beta_1 Einkommen + \beta_2 Alter + \beta_3 Mann + \varepsilon_i$ signifikante Unterschiede in den Steigungsparametern zwischen Männern und Frauen gibt. Wie viele Parameter müssen Sie im Rahmen eines vollständig interagierten Modells schätzen?
a	4
b	**X** 6
c	7
d	8