Bachelorprüfung WS 2013/14 - 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.**
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 Autounfällen. Ihr Datensatz enthält folgende Informationen aus dem Jahr 2013 für 50 verschiedene Länder:

$Unfall$	Anzahl an Autounfällen (in Mio.)
$Alter$	Durchschnittsalter der Bevölkerung in Jahren
$Gurt$	Anteil der Fahrer, die einen Sicherheitsgurt benutzen (gemessen in $1-100\%$)

a) Sie schätzen folgendes lineares Regressionsmodell mit SPSS:

$$Unfall_i = \beta_0 + \beta_1 Alter_i + \beta_2 Gurt_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	1,477	0,368	4,014	0,000
$Alter$	0,068	0,022	3,091	0,003
$Gurt$	-0,025	0,012	-2,083	0,042

a. Abhängige Variable: $Unfall$

Interpretieren Sie den Koeffizienten von $Gurt_i$ sowohl inhaltlich als auch statistisch. (2 Punkte)

- `Steigt der Anteil der Fahrer, die einen Sicherheitsgurt benutzen, um einen Prozentpunkt, so sinkt die Anzahl an Autounfällen c.p. im Durchschnitt um 25.000 Unfälle. [1P]`
- `Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden auf dem 5% Niveau (p-Wert:0,042<0,05). [1P]`

b) Sie schätzen nun folgendes Modell mit SPSS:

$$Unfall_i = \tilde{\beta}_0 + \tilde{\beta}_1 (Alter_i - 40) + \tilde{\beta}_2 (Gurt_i - 80) + \tilde{u}_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	**???**	0,795	**???**	0,008
$Alter - 40$	0,068	0,022	3,091	0,003
$Gurt - 80$	-0,025	0,012	-2,083	0,042

a. Abhängige Variable: $Unfall$

Bestimmen Sie die vorhergesagte Anzahl an Autounfällen bei gegebenen 40 Jahren Durchschnittsalter und 80% Anteil an Fahrern, die einen Sicherheitsgurt benutzen. Berechnen und interpretieren Sie zudem das

99%-Konfidenzintervall für diese Vorhersage. (8 Punkte)

- Punktschätzer für Vorhersagewert:
  $\widehat{Unfall}_i|_{Alter_i=40, Gurt_i=80} = 1,477 + 40 \cdot 0,068 + 80 \cdot (-0,025) = 2,197 = \hat{\theta}$ [2P]
- Standardfehler von Vorhersagewert: $se(\hat{\theta}) = 0,795$ [1P]
- $t_{\frac{\alpha}{2},n-k-1} = t_{0,005,47}$ ist nicht tabelliert, deshalb $t_{0,005,40} = 2,704$ [1P]
- Konfidenzintervall für Vorhersagewert:
  $[\hat{\theta} - t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{\theta}); \hat{\theta} + t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{\theta})]$ = $[2,197 - 2,690 \cdot 0,795; 2,197 + 2,690 \cdot 0,795] = [0,0585; 4,336]$ [3P]
- Interpretation: Für wiederholte Stichproben liegt in 99 % der Fälle der wahre Wert der Vorhersage innerhalb der auf diese Weise berechneten Intervallgrenzen.[1P]

c) Leiten Sie formal her, dass der Koeffizient von $Alter$ aus Teilaufgabe a) dem Koeffizienten von $Alter - 40$ aus Teilaufgabe b) entspricht und genauso der Koeffizient von $Gurt$ aus Teilaufgabe a) dem Koeffizienten von $Gurt - 80$ aus Teilaufgabe b) entspricht. (4 Punkte)

- Regressionsmodell aus a): $Unfall_i = \beta_0 + \beta_1 Alter_i + \beta_2 Gurt_i + u_i$
- $E(Unfall_i|Alter_i = 40, Gurt_i = 80) = \beta_0 + 40\beta_1 + 80\beta_2 = \theta \Leftrightarrow \beta_0 = \theta - 40\beta_1 - 80\beta_2$ [2P]
- Einsetzen in das ursprüngliche Regressionsmodell ergibt:
  $Unfall_i = \theta - 40\beta_1 - 80\beta_2 + \beta_1 Alter_i + \beta_2 Gurt_i + u_i = \theta + \beta_1(Alter_i - 40) + \beta_2(Gurt_i - 80) + u_i$ [2P]
- Die Steigungsparameter sind identisch, lediglich die geschätzte Konstante hat sich verändert.

d) Welche Werte erhält man für die Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_2$ für das Modell aus Teilaufgabe a), wenn die Anzahl an Autounfällen nicht in Millionen, sondern in Tausend gemessen wird? (2 Punkte)

- Umskalierung der abhängigen Variable $Unfall_i$: $1000 \cdot \widehat{Unfall}_i = 1000 \cdot \hat{\beta}_0 + 1000 \cdot \hat{\beta}_1 \cdot Alter_i + 1000 \cdot \hat{\beta}_2 \cdot Gurt_i$
- $\hat{\beta}_{0,neu} = 1000 \cdot \hat{\beta}_0 = 1000 \cdot 1,447 = 1447$ [1P]
- $\hat{\beta}_{2,neu} = 1000 \cdot \hat{\beta}_2 = 1000 \cdot (-0,025) = -25$ [1P]

Aufgabe 2: **[8 Punkte]**

Sie interessieren sich für die Determinanten von Wohnungsmietpreisen. Ihr Datensatz enthält folgende Informationen zu 350 Wohnungen:

$Preis$	Mietpreis der Wohnung in Euro
$Fläche_i$	Wohnfläche in m^2
$Baujahr$	Baujahr der Wohnung
$Stadt$	Indikatorvariable = 1, wenn Wohnung in Stadt; = 0, wenn Wohnung auf Dorf
$Balkon$	Indikatorvariable = 1, wenn Balkon vorhanden; = 0, wenn kein Balkon vorhanden

Sie schätzen folgendes lineares Regressionsmodell mit SPSS:

$$Preis_i = \beta_0 + \beta_1 Fläche_i + \beta_2 Baujahr_i + \beta_3 Stadt_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	-9783,221	4746,832	2,061	0,010
$Fläche$	7,883	1,458	5,405	0,000
$Baujahr$	5,134	1,095	4,689	0,000
$Stadt$	183,729	55,357	3,319	0,001

a. Abhängige Variable: *Preis*

a) Interpretieren Sie den Koeffizienten von $Stadt$ sowohl inhaltlich als auch statistisch. (2 Punkte)

- Eine Wohnung in der Stadt kostet c.p. im Durchschnitt 183,729 Euro mehr als eine Wohnung auf dem Land. [1P]
- Der Koeffizient ist auf dem 1%-Signifikanzniveau statistisch signifikant von Null verschieden. (p-Wert: 0,001<0,01) [1P]

b) Testen Sie auf dem 10%-Signifikanzniveau, ob der Effekt des Baujahres auf den Mietpreis größer als 5 ist. Geben Sie Testverfahren, Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

- Testverfahren: einseitiger t-Test [0,5P]
- Hypothesen: $H_0 : \beta_2 \leq 5, H_1 : \beta_2 > 5$ [1P]
- Teststatistik: $t = \frac{\hat{\beta}_2 - 5}{se(\hat{\beta}_2)} = \frac{5{,}134-5}{1{,}095} = 0{,}122$ [1P]
- Kritischer Wert:$t_{\alpha;n-k-1} = t_{0{,}1;350-3-1} = 1{,}282$ [1P]
- Testentscheidung: Da $t_{empirisch} = 0{,}112 < 1{,}282 = t_{kritisch}$ kann die Nullhypothese auf dem 10% Niveau nicht verworfen werden. [1P]

c) Sie vermuten, dass die Variable *Balkon* eine wichtige ausgelassene Variable des Modells ist. Sie messen eine positive Kovarianz der Variablen $Balkon$ und $Fläche$. Was ist jeweils die Bedingung dafür, dass der Koeffizient von $Fläche$ aus dem ursprünglichen Modell (1) unverzerrt, (2) positiv verzerrt, (3) negativ verzerrt ist? (1,5 Punkte)

- Unverzerrt: Es gibt keinen Einfluss von *Balkon* auf *Preis*. [0,5P]
- Positiv verzerrt: Der Einfluss von *Balkon* auf *Preis* ist positiv. [0,5P]
- Negativ verzerrt: Der Einfluss von *Balkon* auf *Preis* ist negativ. [0,5P]

Aufgabe 3: **[14 Punkte]**

Sie untersuchen die Jahresgehälter von Vorstandsvorsitzenden von Unternehmen. Der Ihnen zur Verfügung stehende Datensatz enthält folgende Informationen zu 209 Firmen:
Sie schätzen folgendes lineares Regressionsmodell mit SPSS.

$$salary_i = \beta_0 + \beta_1 log(sales_i) + \beta_2 invest_i + \beta_3 indus_i + u_i$$

salary	Vorstandsjahresgehalt im Jahr 1990 (in 1000 US-$)
sales	Unternehmensumsatz im Jahr 1990 (in Mio. US-$)
invest	Unternehmensinvestitionen im Jahr 1990 (in Mio. US-$)
indus	Indikatorvariable =1, falls Industrieunternehmen; =0, falls kein Industrieunternehmen.

Hinweis: Das R^2 dieser Schätzung beträgt $0,299$.

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	4,374	0,311	14,080	0,000
$log(sales)$	130,333	12,56	10,377	0,000
invest	0,015	0,004	3,450	0,001
indus	0,283	0,099	2,859	0,005

a. Abhängige Variable: *salary*

a) Interpretieren Sie den geschätzten Koeffizienten für β_1 sowohl inhaltlich als auch statistisch. (2 Punkte)

- Steigt der Firmenumsatz um 1%, so steigt das Jahresgehalt des Vorsitzenden c.p.i.M. um 1303,33$. [1P]
- Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden auf dem 1% Niveau (p-Wert: 0,000<0,01). [1P]

b) Unterstellen Sie, dass der Umsatz einer Firma um 5% zurückgeht. Wie hätten sich die Unternehmensinvestitionen ändern müssen, um diesen Effekt auf das Jahresgehalt des Fimenvorsitzenden auszugleichen? Berechnen und interpretieren Sie die notwendige Änderung. (5 Punkte)

- Ein Rückgang des Umsatzes um 5% reduziert das Jahresgehalt um $\frac{\hat{\beta}_1 \cdot 5}{100} = 6,51$ Einheiten (=6510$). [2P]
- Dieser Effekt wird ausgeglichen durch $\frac{\widehat{\Delta salary}}{\Delta invest} = \hat{\beta}_2 \Leftrightarrow \Delta invest = \frac{6,510}{0,015} = 434$ Einheiten. [2P]
- Interpretation: Um den Rückgang des Umsatzes im Jahresgehalt auszugleichen, hätten die Unternehmensinvestitionen um 434$ Mio. steigen müssen. [1P]

c) Sie vermuten, dass der Effekt von *sales* und *invest* davon abhängt, ob es sich um ein Industrieunternehmen handelt, oder nicht. Stellen Sie das erweiterte Modell auf, mit dessen Hilfe Sie diese Vermutung überprüfen können. Testen Sie zudem auf dem 5% Signifikanzniveau, ob das erweiterte Modell einen signifikant höheren Erklärungsgehalt als das bisherige Modell hat. Geben Sie Null- und Alternativhypothese, Freiheitsgrade, Teststatistik, kritischen Wert und Ihre Testentscheidung an. Hinweis: Das erweiterte Modell hat ein R^2 von $0,321$. Runden Sie auf die **dritte** Nachkommastelle. (7 Punkte)

1. **Erweitertes Modell:** $salary_i = \beta_0 + \beta_1 log(sales_i) + \beta_2 invest_i + \beta_3 indus_i + \beta_4 log(sales_i) \cdot indus_i + \beta_5 invest_i \cdot indus_i + u_i$ [2P]
2. **Nullhypothese:** $H_0 : \beta_4 = \beta_5 = 0$ [0,5P]

3. **Alternativhypothese:** $H_1: \beta_4 \neq 0$ und/oder $\beta_5 \neq 0$ [0,5P]
4. **Freiheitsgrade (aus unrestringiertem Modell):** $n-k-1 = 209-5-1 = 203$ [0,5P] und $q = 2$ [0,5P]
5. **Teststatistik:** $F_{empirisch} = \frac{(R_u^2 - R_R^2)/q}{(1-R_u^2)/(n-k-1)} = \frac{(0{,}321-0{,}299)/2}{(1-0{,}321)/(209-5-1)} = \frac{0{,}011}{0{,}003} = 3{,}667$ [1P]
6. **Kritischer Wert:** $F_{kritisch} = F_{0{,}05;2{,}209} = 3{,}00$ [1P]
7. **Testentscheidung:** Da $F_{empirisch} > F_{kritisch}$ wird die Nullhypothese auf dem 5% Niveau verworfen [1P]. Auf Grundlage des Tests hat das erweiterte Modell einen signifikant höheren Erklärungsgehalt als das ursprüngliche Modell.

Aufgabe 4: **[12 Punkte]**

a) Leiten Sie formal für das Modell $y_i = \beta_1 x_i + u_i$ den KQ-Schätzer $\hat{\beta}_1$ her. (4 Punkte)

- $min \sum \hat{u}_i^2 = \sum (y_i - \hat{\beta}_1 x_i)^2$ [1P]
- Nach $\hat{\beta}_1$ ableiten und gleich Null setzen: $2\sum (y_i - \hat{\beta}_1 x_i)(-x_i) = 0$ [1P]
- Auflösen ergibt $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$ [2P]

b) Interpretieren und erläutern Sie anhand der folgenden Formel, welche Größen die Varianz des Schätzers in welche Richtung beeinflussen:
$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2 \cdot (1-R_j^2)}$. (Hinweis: R_j^2 bezeichnet das Bestimmtheitsmaß einer Hilfsregression, in der x_j auf alle anderen $x_{k,k\neq j}$ regressiert wird.) (3 Punkte)

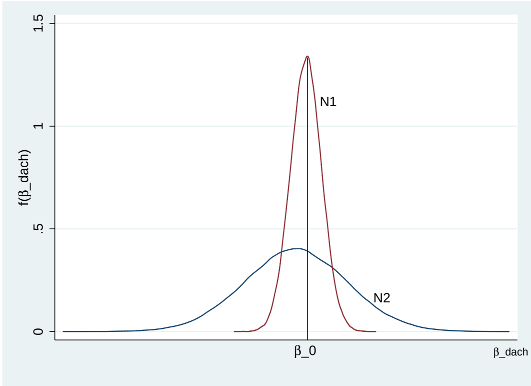
- Eine größere Störtermvarianz ($\hat{\sigma}^2$) [0.5P] erhöht [0.5P] die Varianz.
- Hohe Variation in den erklärenden Variablen ($\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2$) [0.5P] reduziert [0.5P] die Varianz.
- Hohe Multikollinearität ($1 - R_j^2 \rightarrow 0$) [0.5P] erhöht [0.5P] die Varianz.

c) Erläutern Sie verbal, was man unter einem unverzerrten Schätzer versteht. (2 Punkte)

Ein Schätzer ist unverzerrt, wenn der Erwartungswert des Schätzers dem wahren Parameter entspricht. [2P]

d) Stellen Sie die Verteilungsfunktion eines konsistenten Schätzers für β_0 im Fall einer großen Stichprobe (N_1) und einer kleinen Stichprobe (N_2) graphisch dar. Beschriften Sie die Achsen. (3 Punkte)

0,5P je y,x Achsenbeschriftung. 1P je Verteilung.



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.	Nimmt die empirische F-Statistik den Wert 2,5 an, so wird die Nullhypothese
a	**X** bei 5 Nenner- und 20 Zählerfreiheitsgraden am 10% Niveau verworfen.
b	bei 5 Nenner- und 10 Zählerfreiheitsgraden am 10% Niveau verworfen.
c	bei 2 Nenner- und 120 Zählerfreiheitsgraden am 10% Niveau nicht verworfen.
d	bei 2 Nenner- und 20 Zählerfreiheitsgraden am 10% Niveau verworfen.

2.	Eine KQ-Schätzung liefert $\hat{y}_i = 3,2 + 1,7x_i$. Welchen Wert hat das Residuum für die Beobachtung $(y_1, x_1) = (5, 1)$?
a	-2,9.
b	**X** 0,1.
c	1,3.
d	3,7.

3.	Wenn die Störterme einer multiplen Regression normalverteilt sind
a	ist der KQ-Schätzer inkonsistent.
b	**X** sind t- und F-Tests auch in kleinen Stichproben gültig.
c	ist der KQ-Schätzer asymptotisch ineffizient.
d	sind t- und F-Tests nur in großen Stichproben gültig.

4.	Ein hohes R^2 bedeutet, dass
a	das Modell korrekt ist.
b	**X** die erklärte und die gesamte Variation relativ nah beieinander liegen.
c	alle Schätzkoeffizienten statistisch signifikant sind.
d	kausale Effekte zuverlässig ermittelt werden können.

5.	Sie beobachten das Alter für Männer und Frauen in Ost- und Westdeutschland. Sie schätzen folgendes Modell: $alter_i = \beta_0 + \beta_1 mann_i + \beta_2 west_i + \beta_3 mann_i \cdot west_i + u_i$. Was misst der geschätzte Parameter für β_2?
a	Den Mittelwert des Alters für Westdeutschland.
a	Den Mittelwert des Alters für Westdeutsche Frauen.
c	Den Mittelwertunterschied des Alters zwischen West- und Ostdeutschland.
d	**X** Den Mittelwertunterschied des Alters zwischen West- und Ostdeutschland für Frauen.

6.	Man untersucht, ob ein Steigungsparameter gleich -2 ist, indem man
a	den Wert 2 vom Schätzwert subtrahiert und das Ergebnis mit dem kritischen Wert der t-Verteilung vergleicht.
b	den Wert 2 und den Schätzwert addiert und das Ergebnis mit dem kritischen Wert der t-Verteilung vergleicht.
c	**X** überprüft, ob das Konfidenzintervall den Wert -2 einschließt oder nicht.
d	überprüft, ob das Konfidenzintervall den Wert 2 einschließt oder nicht.

7.	Welches Problem tritt bei der Schätzung des folgenden Modells auf: $lohn_i = \beta_0 + \beta_1 mann_i + \beta_2 frau_i + \beta_3 bildung_i + u_i$
a	**X** Das Modell ist perfekt multikollinear.
b	Der KQ-Schätzter ist ineffizient.
c	Der KQ-Schätzer ist verzerrt.
d	Das R^2 ist 1.

8.	Das R^2 einer Regression ist 1, wenn
a	die Summe der Residuen gleich 0 ist.
b	die Fehlerquadratsumme der gesamten Variation entspricht.
c	**X** alle Residuen gleich 0 sind.
d	die Fehlerquadratsumme der erklärten Variation entspricht.

9.	Sie schätzen das folgende Modell für den Stundenlohn mittels KQ: $log(lohn_i) = \beta_0 + \beta_1 frau_i + \beta_2 educ_i + u_i$ und erhalten $\hat{\beta}_1 = -0.357$ (der Lohn ist in Euro gemessen). Wie hoch ist der exakte, auf die erste Nachkommastelle gerundete Lohnunterschied zwischen Männern und Frauen?
a	35,7 Cent.
b	3,6 Euro.
c	35,7%.
d	**X** 30,0%.

10.	Im einfachen linearen Regressionsmodell berechnet sich der Steigungsparameter $\hat{\beta}_1$ als Verhältnis der
a	Stichprobenkovarianz von x und y zur Stichprobenvarianz von y.
b	**X** Stichprobenkovarianz von x und y zur Stichprobenvarianz von x.
c	Stichprobenvarianz von y zur Stichprobenkovarianz von x und y.
d	Stichprobenvarianz von x zur Stichprobenkovarianz von x und y

11.	Welcher Ausdruck beschreibt den marginalen Effekt von exp für das Modell $lohn_i = \beta_0 + \beta_1 frau_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 exp_i \cdot frau_i + u_i$?
a	β_2.
b	$\beta_2 + 2\beta_3 exp_i$.
c	**X** $\beta_2 + 2\beta_3 exp_i + \beta_4 frau_i$.
d	$\beta_0 + \beta_2 + 2\beta_3 exp_i + \beta_4 frau_i$.

12.	Der Standardfehler der Regression (SER) ist ein Schätzer für
a	die auf y bedingte Streuung von u und x.
b	**X** die auf x bedingte Streuung von u und y.
c	die auf u bedingte Streuung von x und y.
d	Keine der Antworten ist korrekt.

13.	Unter den Gauss-Markov Annahmen
a	folgt der KQ-Schätzer einer F-Verteilung.
b	folgt der KQ-Schätzer einer log-Normal Verteilung.
c	**X** folgt der KQ-Schätzer einer Normalverteilung.
d	folgt der KQ-Schätzer einer χ^2- Verteilung.

14.	Wird ein irrelevanter Regressor (d.h. ein Regressor, dessen Koeffizient gleich 0 ist) einem multiplen Regressionsmodell hinzugefügt, so
a	verändern sich die Regressionsergebnisse nicht.
b	**X** verringert sich die Präzision der Schätzung.
c	erhöht sich automatisch das R^2.
d	kann das Problem einer Verzerrung durch ausgelassene Variablen vermieden werden.

15.	Die t-Verteilung
a	hängt nicht von der Zahl der Freiheitsgrade ab.
b	ist symmetrisch um den Erwartungswert 1.
c	**X** konvergiert bei steigenden Freiheitsgraden gegen die Normalverteilung.
d	ergibt quadriert eine χ^2- Verteilung.

16.	Schätzt man ein einfaches lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$, so
a	**X** liegt der Punkt $(\overline{y}, \overline{x})$ auf der geschätzten Regressionsgerade.
b	werden die Abstände der Datenpunkte zur geschätzten Regressionsgerade minimiert.
c	ist die Summe der Residuen ungleich 0.
d	ist das korrigierte R^2 stets positiv.

17.	Wenn eine diskrete Zufallsvariable X die Werte -2, 0 und 2 mit konstanter Wahrscheinlichkeit $\frac{1}{3}$ annimmt, dann gilt
a	**X** $E(X) = 0$.
b	$E(X) = \frac{2}{3}$.
c	$E(X^2) = 0$.
d	$E(X^2) = \frac{1}{9}$.

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

19.	Bei konsistenten Schätzverfahren
a	ist der Schätzer stets unverzerrt.
b	ist der Erwartungswert des Schätzers stets mit dem wahren Wert identisch.
c	steigt die Varianz des Schätzers, je größer die Stichprobe.
d	**X** nähert sich der Wahrscheinlichkeitsgrenzwert des Schätzers mit steigender Stichprobengröße dem wahren Wert an.

20.	Das Konfidenzintervall für einen geschätzten Parameter kann
a	anstelle eines t-Tests ausschließlich für einseitige Tests verwendet werden.
b	anstelle eines t-Tests ausschließlich für zweiseitige Tests verwendet werden.
c	**X** anstelle eines t-Tests für ein- und zweiseitige Tests verwendet werden.
d	nicht anstelle von t-Tests verwendet werden.

21.	Die Eigenschaft der Konsistenz
a	gilt auch in kleinen Stichproben.
b	kann nur für unverzerrte Schätzer nachgewiesen werden.
c	ist für die Qualität eines Schätzers nicht wichtig.
d	**X** ist eine asymptotische Eigenschaft.

22.	Heteroskedastie
a	führt zu unverzerrten Standardfehlern der geschätzten Koeffizienten.
b	**X** führt nicht zu verzerrt geschätzten Koeffizienten.
c	tritt nur bei Zeitreihendaten auf.
d	kann bei Regressionen mit Dummyvariablen nicht auftreten.

23.	Ein Schätzer $\hat{\beta}_1$ für den unbekannten Parameter β_1 im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ ist konsistent, wenn
a	**X** $cov(x,u) = 0$.
b	x normalverteilt ist.
c	$\frac{1}{n} \to \infty$.
d	$plim(\beta_1) = \hat{\beta}_1$.

24.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \beta_4 x_{3i} + \beta_5 x_{4i} + u_i$. Man testet $H_0 : \beta_2 < 2$ vs. $H_1 : \beta_2 \geq 2$. Bei einem Signifikanzniveau von 5% führt eine Teststatistik von $1,690$
a	bei $n = 10$ zur Ablehung von H_0.
b	bei $n = 20$ zur Ablehung von H_0.
c	bei $n = 30$ zur Ablehung von H_0.
d	**X** bei $n = 40$ zur Ablehung von H_0.

25.	Sie schätzen folgende Konsumfunktion mit KQ: $konsum_i = \beta_0 + \beta_1 log(einkommen_i) + u_i$. Hinweis: *konsum* und *einkommen* sind beide in Euro gemessen. $\hat{\beta}_1$ ist 900. Welche Aussage ist richtig?
a	Wenn das Einkommen um 1000 Euro ansteigt, steigt der Konsum um 900 Euro.
b	Wenn das Einkommen um 1 Euro ansteigt, steigt der Konsum um 0,9%.
c	**X** Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 9 Euro.
d	Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 0,9%.

26.	Im linearen Wahrscheinlichkeitsmodell
a	**X** gibt die Prognose $\hat{y}$ die Wahrscheinlichkeit für das Ereignis $y = 1$ an.
b	wird die abhängige Variable in Prozentpunkten gemessen.
c	werden nur binär kodierte Variablen als erklärende Variablen verwendet.
d	können kausale Effekte nicht identifiziert werden.

27.	Sie führen einen Chow-Test auf Geschlechterunterschiede unter Verwendung des folgenden unrestringierten Modells für 66 Beobachtungen durch: $lohn_i = \beta_0 + \beta_1 frau_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 exp_i \cdot frau_i + \beta_5 exp_i^2 \cdot frau_i + u_i$. Was ist der kritische Wert am 1% Niveau?
a	1,76.
b	3,78.
c	**X** 4,13.
d	7,08.

28.	Um die Präzision einer Schätzung zu erhöhen, sollte man
a	standardisierte Koeffizienten berechnen.
b	**X** keine irrelevanten erklärenden Variablen verwenden.
c	auf geringe Variation in den erklärenden Variablen achten.
d	ein möglichst hohes Signifikanzniveau wählen.

29.	Kausale Effekte
a	beschreiben den Effekt einer Größe X auf Y ohne andere Faktoren konstant zu halten.
b	können im multiplen linearen Regressionsmodell nicht geschätzt werden.
c	können nur mit experimentellen Daten geschätzt werden.
d	**X** lassen sich mit Befragungsdaten schätzen.

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

31.	Typ I-Fehlerwahrscheinlichkeiten
a	beschreiben die Wahrscheinlichkeit, die Nullhypothese anzunehmen.
b	beschreiben die Wahrscheinlichkeit, die Alternativhypothese zu verwerfen.
c	werden durch t-Werte beschrieben.
d	**X** sind für t-Tests relevant.

32.	Für das Modell $lohn_i = \beta_0 + \beta_1 alter_i + \beta_2 alter_i^2 + u_i$ ergibt eine KQ-Schätzung $\hat{\beta}_1 = 0,28$ und $\hat{\beta}_2 = -0,004$. In welchem Alter wird der durchschnittliche Lohn maximiert?
a	32.
b	**X** 35.
c	38.
d	40.

33.	Die empirische F-Statistik
a	kann nicht dafür genutzt werden, die Signifikanz eines einzelnen Parameters zu testen.
b	entspricht asymptotisch der t-Statistik.
c	nimmt nur Werte im Interval (0,10) an.
d	**X** kann mittels *SSR* berechnet werden.

34.	Die F-Verteilung
a	**X** wird durch zwei Parameter (Freiheitsgradmaße) charakterisiert.
b	wird durch einen Parameter (Freiheitsgradmaß) charakterisiert.
c	ist eine symmetrische Verteilung.
d	nimmt auch negative Werte an.

35.	Für ein lineares Wahrscheinlichkeitsmodell gilt, dass
a	der KQ-Schätzer unverzerrt ist.
b	die Störterme homoskedastisch sind.
c	**X** vorhergesagte Wahrscheinlichkeiten auch außerhalb des Intervalls (0,1) liegen können.
d	der KQ-Schätzer BLUE (best-linear-unbiased estimator) ist.

36.	In einer quadratischen Spezifikation $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ ergibt sich für $x > 0$ eine u-förmige Beziehung zwischen x und y, wenn
a	$\beta_1 > 0$ und $\beta_2 < 0$.
b	**X** $\beta_1 < 0$ und $\beta_2 > 0$.
c	$\beta_1 > 0$ und $\beta_2 > 0$.
d	$\beta_1 < 0$ und $\beta_2 = 0$.

37.	In dem Modell $y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$ mit zwei 0/1 kodierten Dummyvariablen D_{1i} und D_{2i} gilt:
a	**X** $\hat{\beta}_1$ gibt den Mittelwert von y_i für alle Beobachtungen mit $D_{1i} = 1$ und $D_{2i} = 0$ an.
b	β_1 gibt an, um wieviel sich der Erwartungswert von y_i ändert, wenn D_{2i} von 0 auf 1 steigt.
c	$D_{1i} + D_{2i} = 3$.
d	kann β_2 nur Werte zwischen 0 und 1 annehmen.

38.	Im Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ kommt es bei Weglassen von x_{2i} zu einer negativen Verzerrung von β_1, wenn (Hilfsregression: $x_{1i} = \delta_0 + \delta_2 x_{2i} + \varepsilon_i$)
a	**X** $\beta_2 > 0$ und $\delta_2 < 0$.
b	$\beta_2 > 0$ und $\delta_2 > 0$.
c	$\beta_2 < 0$ und $\delta_2 < 0$.
d	$\beta_2 > 0$ und $\delta_2 = 0$.

39.	Perfekte Multikollinearität
a	lässt sich vermeiden, wenn man den Stichprobenumfang erhöht.
b	tritt nur in Modellen mit binären Variablen auf.
c	führt zu kollinearen Parametern.
d	**X** lässt sich durch das Weglassen von erklärenden Variablen vermeiden.

40.	Im log-log Modell $log(y_i) = \beta_0 + \beta_1 log(x_i) + u$
a	hat das R^2 keine sinnvolle Interpretation.
b	kann x auf dem Interval $(-\infty, +\infty)$ definiert sein.
c	sind die Störterme heteroskedastisch.
d	**X** wird β_1 als Elastizität interpretiert.