Bachelorprüfung

Fach: Praxis der empirischen Wirtschaftsforschung

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

Name, Vorname		
Matrikelnr.		
E-Mail		
Studiengang		
Semester		
Datum		
Raum, Platznr.		
Im SS 2012 freiwillige Hausarbeit abgegeben	**❑ Ja**	**❑ Nein**
Unterschrift		

<u>Vorbemerkungen:</u>

Anzahl der Aufgaben: Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen.

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: **[11 Punkte]**

Sie interessieren sich für die Determinanten des Spendeverhaltens. Ihnen liegen Daten des Sozio-oekonomischen Panels (SOEP) mit 16.739 Personen vor. Folgende Informationen sind gegeben:

lnspende	Logarithmierter Spendenbetrag in Euro pro Jahr
mann	=1 wenn Person männlich ist, =0 wenn weiblich
alter	Alter in Jahren
lnlohn	Logarithmierter Monatslohn in Euro.

Sie schätzen die folgende Gleichung mit SPSS:

$$\mathrm{lnspende_i} = \beta_0 + \beta_1 \mathrm{mann_i} + \beta_2 \mathrm{alter_i} + \beta_3 \mathrm{lnlohn_i} + \varepsilon_i$$

Sie erhalten folgendes Schätzergebnis und unterstellen die Gültigkeit der Gauss-Markov Annahmen:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,300[a]	,0898	**??**	2,2662

Modell	Nicht standardisierte Koeffizienten		T	Sig.
	RegressionskoeffizientB	Standardfehler		
(Konstante)	-0,440	0,062	-7,080	,0000
mann	-0,081	0,035	-2,300	0,022
alter	0,039	0,001	37,610	0,000
lnlohn	0,471	0,019	24,850	0,000

a) Interpretieren Sie den geschätzten Koeffizienten für β_2 inhaltlich und statistisch. (2 Punkte)

Wenn das Alter um 1 Jahr steigt, so steigt der Spendenbetrag pro Jahr c.p. im Mittel um ca. 3,9%.

Der Koeffizient ist am 1%-Niveau statistisch signifikant. Berechnen und interpretieren Sie, wie sich der Spendenbetrag ändert, wenn eine Person eine Gehaltserhöhung von 2,5% bekommt. (2,5 Punkte)

$\Delta\%\mathrm{spende} = \hat{\beta}_3 \cdot \%\Delta\mathrm{lohn} = 0{,}471 \cdot 2{,}5 = 1{,}18$

Wenn eine Person eine Gehaltserhöhung von 2,5% bekommt, so steigt der jährliche Spendenbetrag c.p. im Mittel um 1,18%.

b) Berechnen Sie das 99% Konfidenzintervall für den geschätzten Parameter $\hat{\beta}_1$ (mann) und interpretieren sie es. (3,5 Punkte)

$99\%\ \mathrm{Konfidenzintervall}_{\hat{\beta}_1} = \left[\hat{\beta}_1 \pm c \cdot \mathrm{se}\left(\hat{\beta}_1\right)\right]$

$c = t_{\frac{0{,}01}{2};\infty} = 2{,}576$

$[-0{,}081 \pm 2{,}576 \cdot 0{,}035] = [-0{,}171; 0{,}009]$

Bei wiederholter Stichprobenziehung liegt in 99% der Fälle der wahre Parameter β_1 innerhalb der auf diese Weise bestimmten Konfidenzintervalle.

c) Interpretieren Sie das Bestimmtheitsmaß. (1 Punkt)

8,98% der Variation in der abhängigen Variable wird durch das Modell erklärt.

d) Berechnen Sie das angepasste Bestimmtheitsmaß $\overline{R}^2$. (2 Punkte)

$$\overline{R^2} = 1 - (1 - R^2)\frac{n-1}{n-k-1} = 1 - (1 - {,}0898) \cdot \frac{16739 - 1}{16739 - 3 - 1} = 0{,}0896$$

Das angepasste Bestimmtheitsmaß beträgt 0,0896.

Aufgabe 2: **[13 Punkte]**

Sie interessieren sich für die Determinanten der Deutschkenntnisse von Zuwanderern. Sie untersuchen einen Querschnittsdatensatz aus dem Jahr 2010 mit 1.652 Zuwanderern. Es liegen Ihnen folgende Informationen vor:

speak	Sprachkompetenz in Deutsch, gemessen durch einen Test in Punkten (0-100)
female	=1 wenn Person weiblich ist, =0 wenn männlich
agemig	Alter in Jahren zum Zeitpunkt der Zuwanderung
resid	Aufenthaltsdauer in Deutschland gemessen in Jahren/10
resid2	resid quadriert ($resid^2$)
age	Alter in Jahren zum Zeitpunkt der Befragung.

Sie stellen folgendes Modell auf:

$$\text{speak}_i = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{agemig}_i + \beta_3 \text{resid}_i + \beta_4 \text{resid2}_i + \varepsilon_i$$

SPSS liefert Ihnen folgende Koeffizienten:

$$\widehat{\text{speak}}_i = 11{,}26 + 0{,}174\ \text{female}_i - 0{,}033\ \text{agemig}_i + 0{,}312\ \text{resid}_i - 0{,}051\ \text{resid2}_i$$

a) Nennen Sie eine weggelassene Variable, die den Koeffizienten $\hat{\beta}_1$ verzerren könnte und begründen Sie, warum in Ihrem Beispiel eine Verzerrung bestehen kann. Nennen Sie die Richtung der Verzerrung, geben Sie an ob β_1 über- oder unterschätzt wird und begründen Sie Ihre Antwort. (5 Punkte)

   Beispiel: Bildung

   Wahres Modell: $\text{speak}_i = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{agemig}_i + \beta_3 \text{resid}_i + \beta_4 \text{resid2}_i + \beta_5 \text{educ}_i + \varepsilon_i$

   Könnte $\hat{\beta}_1$ verzerren, weil Bildung mit der Sprachkompetenz und mit dem Geschlecht korreliert sein könnte (z.B. wenn weibliche Migranten bessere Bildung haben als männliche, kann auch umgekehrt sein).

   Die Verzerrung wird bestimmt durch:

   $\text{Bias}(\hat{\beta}_1) = \beta_5 \frac{\text{Cov}(\text{female}_i, \text{educ}_i)}{\text{Var}(\text{educ}_i)}$

   Die Verzerrung ist positiv wenn

   $\text{Cov}(\text{educ}_i, \text{female}_i) > 0$ und $\beta_5 > 0$, d.h. β_1 würde überschätzt.

   (Auch möglich: Die Verzerrung ist negativ, wenn

   $\text{Cov}(\text{educ}_i, \text{female}_i) < 0$ und $\beta_5 > 0$ (oder umgekehrt), d.h. β_1 würde unterschätzt.

b) Berechnen Sie den marginalen Effekt der Aufenthaltsdauer in Deutschland auf die Sprachkompetenz für Personen, die seit 20 Jahren in Deutschland leben. (2 Punkte)

   $$\left.\frac{\partial \text{speak}}{\partial \text{resid}}\right|_{\text{resid}=2} = 0{,}312 - 2 \cdot 0{,}051 \cdot 2 = 0{,}108$$

   Für Personen, die seit 20 Jahren in Deutschland leben steigt die Sprachkompetenz c.p. im Mittel um 0,108 Punkte, mit weiteren 10 Jahren in Deutschland.

c) Die Aufenthaltsdauer in Deutschland wird nicht wie bisher in 10 Jahren, sondern in Jahren gemessen.

   i. Zeigen Sie die Konsequenzen der Umskalierung für die Koeffizienten $\hat{\beta}_3$ und $\hat{\beta}_4$. (2 Punkte)

   $$\widehat{\text{speak}} = \hat{\beta}_0 + \hat{\beta}_1 \text{female} + \hat{\beta}_2 \text{agemig} + \hat{\beta}_3 \frac{(\text{resid} \cdot 10)}{10} + \hat{\beta}_4 \frac{(\text{resid} \cdot 10)^2}{100}.$$

   $$\widehat{\text{speak}} = \hat{\beta}_0 + \hat{\beta}_1 \text{female} + \hat{\beta}_2 \text{agemig} + \underbrace{\frac{\hat{\beta}_3}{10}}_{\text{neues } \widetilde{\beta_3}} (\text{resid} \cdot 10) + \underbrace{\frac{\hat{\beta}_4}{100}}_{\text{neues } \widetilde{\beta_4}} (\text{resid} \cdot 10)^2.$$

   ii. Berechnen Sie die neuen Koeffizienten. (1 Punkt)

$\widetilde{\beta_3} = \frac{0{,}311}{10} = 0{,}031$ und $\widetilde{\beta_4} = \frac{-0{,}051}{100} = -0{,}00051$

d) Sie vermuten, dass das Alter einer Person eine wichtige Determinante der Sprachkompetenz ist. Sie nehmen zusätzlich zu den Variablen agemig und resid die Variable age ins Modell:

$$\text{speak}_\text{i} = \beta_0 + \beta_1\text{female}_\text{i} + \beta_2\text{agemig}_\text{i} + \beta_3\text{resid}_\text{i} + \beta_4\text{resid2}_\text{i} + \beta_5\text{age}_\text{i} + \varepsilon_\text{i}$$

Benennen und begründen Sie das Problem, das durch die Aufnahme von age in die Regression entstanden ist. Welche Auswirkungen hat dies auf die Schätzung? (3 Punkte)

Perfekte Multikollinearität, da age $= (\text{agemig} + 10 \cdot \text{resid})$.
Die Annahme MLR3 ist verletzt. Eine KQ-Schätzung nicht durchführbar.

Aufgabe 3: **[14 Punkte]**

Sie möchten untersuchen, wie sich der Zigarettenkonsum auf die körperliche Gesundheit auswirkt. Ihnen liegen Daten des Sozio-oekonomischen Panels (SOEP) mit 17.183 Rauchern und Nichtrauchern vor. Folgende Informationen sind gegeben:

pcs	Skala der physischen Gesundheit, von 0 (sehr schlecht) bis 100 (sehr gut)
lnhhinc	Logarithmiertes Haushaltseinkommen
educ	Anzahl der Schuljahre
cigs	Anzahl der gerauchten Zigaretten pro Tag

Sie schätzen die folgende Gleichung mit SPSS:

$$pcs_\text{i} = \beta_0 + \beta_1 lnhhinc_\text{i} + \beta_2 educ_\text{i} + \beta_3 cigs_\text{i} + \varepsilon_\text{i}$$

Sie erhalten folgendes Schätzergebnis und unterstellen die Gültigkeit der Gauss-Markov Annahmen:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	47,406	0,97	48,895	0,000
	lnhhinc	1,344	0,124	10,825	0,000
	educ	0,515	0,027	19,403	0,000
	cigs	-0,027	0,008	-3,497	0,000

a) Wie verändert sich die erwartete Gesundheit bei einer Reduktion des Zigarettenkonsums um 5 Zigaretten pro Tag? Berechnen und interpretieren Sie das Ergebnis. (2 Punkte)

$\Delta\text{pcs} = \hat{\beta}_3 \cdot \Delta\text{cigs} =$ -0,027·(-5) = 0,135

Wenn der Zigarettenkonsum um 5 Zigaretten gesenkt wird, so steigt die Gesundheit auf der Skala c.p. im Mittel um 0,135 Punkte.

b) Interpretieren Sie den geschätzten Koeffizienten für β_1 inhaltlich und statistisch. (2 Punkte)

Wenn das Haushaltseinkommen um 1% ansteigt, so steigt das Gesundheitsniveau auf der Skala c.p. im Mittel um 1,344/100 = 0,01344

Der Koeffizient ist am 1%-Niveau statistisch signifikant.

c) Sie vermuten, dass es einen nicht-linearen Zusammenhang zwischen der Anzahl der gerauchten Zigaretten und der Gesundheit gibt. Sie entfernen die lineare Variable (cigs) aus dem Modell und nehmen stattdessen Dummy-Variablen zum Rauchkonsum auf (cigs1_10 = 1 bis 10 Zigaretten; cigs11_20 = 11 bis 20 Zigaretten; cigs20plus = 20 und mehr Zigaretten). Sie stellen folgendes Modell auf:

$$pcs_\text{i} = \beta_0 + \beta_1 lnhhinc_\text{i} + \beta_2 educ_\text{i} + \beta_3 cigs1\_10_\text{i} + \beta_4 cigs11\_20_\text{i} + \beta_5 cigs20plus_\text{i} + \varepsilon_\text{i}$$

Sie erhalten folgende Schätzergebnisse:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	47,451	0,976	48,62	0
	lnhhinc	1,342	0,124	10,8	0
	educ	0,517	0,027	19,443	0
	cigs1_10	-0,346	0,215	-1,609	0,108
	cigs1_120	-0,532	0,192	-2,771	0,006
	cigs20plus	-0,842	0,328	-2,567	0,01

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,551[a]	,3183	,318	8,45382

i) Welchen Gesundheitseffekt hat der Anstieg des Zigarettenkonsums von 8 auf 22 Zigaretten? Zeigen Sie den Rechenweg. Berechnen und interpretieren Sie das Ergebnis. (4 Punkte)

1) Vorhergesagter Wert bei 8 Zigaretten:
$\widehat{pcs} = \hat{\beta}_0 + \hat{\beta}_1 lnhhinc + \hat{\beta}_2 educ + \hat{\beta}_3 cigs1\_10 + \hat{\beta}_4 cigs11\_20 + \hat{\beta}_5 cigs20plus = \hat{\beta}_0 + \hat{\beta}_1 lnhhinc + \hat{\beta}_2 educ + \hat{\beta}_3$

2) Vorhergesagter Wert bei 22 Zigaretten:
$\widehat{pcs} = \hat{\beta}_0 + \hat{\beta}_1 lnhhinc + \hat{\beta}_2 educ + \hat{\beta}_3 cigs1\_10 + \hat{\beta}_4 cigs11\_20 + \hat{\beta}_5 cigs20plus = \hat{\beta}_0 + \hat{\beta}_1 lnhhinc + \hat{\beta}_2 educ + \hat{\beta}_5$

3) $\Delta\widehat{pcs} = \hat{\beta}_5 - \hat{\beta}_3 = -0{,}842 - (-0{,}346) = -0{,}496$

4) Beim Anstieg des Zigarettenkonsums von 8 auf 22 Zigaretten verschlechtert sich die Gesundheit c.p. im Mittel um 0,496 Punkte.

ii) Testen Sie am 1%-Niveau, ob die Koeffizienten der Zigaretten Dummy-Variablen gemeinsam statistisch signifikant sind. Geben Sie Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und ihre Testentscheidung an. (Hinweis: Im restringierten Modell beträgt die residuale Quadratsumme (SSR) 1196370,66. Die Summe der gesamten Variation der abhängigen Variable (SST) beträgt 1752725,38). (6 Punkte)

R^2_{restr}: $1 - \frac{1196370{,}66}{1752725{,}38} = 0{,}3174$

$H_0: \beta_3 = \beta_4 = \beta_5 = 0$

H_1: mindestens ein $\beta_j \neq 0$ mit j = 3, 4, 5

Teststatistik

$$F = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-k-1)} = \frac{(0{,}3183-0{,}3174)/3}{(1-0{,}3183)/(17183-5-1)} = \frac{0{,}0003}{0{,}0000397} = 7{,}56$$

Zählerfreiheitsgrade: 3

Nennerfreiheitsgrade: n-k-1=17183-5-1=17177

Kritischer Wert: $F_c = F_{0,01;3;17177} = 3{,}78$

Testentscheidung: Da $F > F_c$ wird die Nullhypothese verworfen. Die Koeffizienten der Zigaretten Dummy-Variablen sind gemeinsam statistisch signifikant von Null verschieden.

Aufgabe 4: **[12 Punkte]**

Sie interessieren sich für die Determinanten der wöchentlichen Arbeitsstunden. Sie untersuchen eine Zufallsstichprobe mit 31.824 Beobachtungen von amerikanischen Frauen. Es liegen Ihnen folgende Informationen vor:

hours	Wöchentliche Arbeitsstunden
nonlinc	Nichtarbeitseinkommen in \$1000 (pro Jahr)
age	Alter in Jahren
age2	Alter quadriert (age^2)
hisp	=1 wenn hispanisch, =0 wenn nicht hispanisch.

Sie stellen folgendes Modell auf:

$$hours_i = \beta_0 + \beta_1 nonlinc_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 hisp_i + \varepsilon_i$$

SPSS liefert Ihnen den folgenden Output:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	-9,404	6,442	-1,46	0,144
	nonlinc	-0,058	0,005	-10,663	0,000
	age	2,05	0,448	4,574	0,000
	age²	??	??	??	??
	hispanic	-6,158	0,231	-26,658	0,000

a) Wie verändert sich die erwartete Arbeitszeit bei einer Erhöhung des Nichtarbeitseinkommens um \$7500? Berechnen und interpretieren Sie kurz das Ergebnis. (2 Punkte)

$\Delta hours = \hat{\beta}_1 \cdot \Delta nonlinc =$ -0,058·7,5 = -0,435

Wenn sich das Nichtarbeitseinkommen um 7500\$ erhöht, so sinkt die Anzahl der wöchentlichen Arbeitsstunden c.p. im Mittel um ca. 0,435 Stunden.

b) Welchen Wert nimmt der geschätzte Koeffizient $\hat{\beta}_3$ an, wenn Frauen gemäß diesem Modell im Alter von 36,9 Jahren c.p. im Mittel ihre höchste wöchentliche Arbeitszeit realisieren? (2 Punkte)

$\frac{\Delta hours}{\Delta age} = \hat{\beta}_2 + 2\hat{\beta}_3 \cdot age$

$\hat{\beta}_2 + 2\hat{\beta}_3 \cdot age = 0$

$2{,}05 + 2\hat{\beta}_3 \cdot 36{,}9 = 0$

$\hat{\beta}_3 = -0{,}0277$

c) Testen Sie am 5%-Signifikanzniveau, ob Hispanics signifikant mindestens mehr als 5 Wochenarbeitsstunden weniger arbeiten als nicht-Hispanics. Geben Sie Testvefahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

Einseitiger T-test

$H_0: \beta_4 \geq -5$

$H_1: \beta_4 < -5$

Teststatistik: $t = \frac{\hat{\beta}_4 - (-5)}{se(\hat{\beta}_4)} = \frac{-6{,}158+5}{0{,}231} = -5{,}022$

Freiheitsgrade: n-k-1=31824-4-1=31819

Kritischer Wert: $(-)\, t_{\alpha;n-k-1} = (-)t_{0{,}05;31824-4-1} = -1{,}645$

Entscheidung: Die Nullhypothese kann am 5%-Niveau verworfen werden.

d) Gehen Sie nun von folgendem Modell aus:

$$\text{hours}_i = \beta_0 + \beta_1 \text{nonlinc}_i + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \varepsilon_i$$

i) Stellen Sie die Regressionsgleichung für ein Modell auf, mit dem Sie die Vermutung überprüfen können, dass sich alle Koeffizienten des Modells für Hispanics und nicht-Hispanics voneinander unterscheiden. (2 Punkte)

Schätzung eines vollständig interagierten Modells (unrestringiertes Modell)

$\text{hours}_i = \beta_0 + \beta_1 \text{nonlinc}_i + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 \text{hisp}_i + \beta_5 \text{nonlinc}_i \cdot \text{hisp}_i + \beta_6 \text{age}_i \cdot \text{hisp}_i + \beta_7 \text{age}_i^2 \cdot \text{hisp}_i + \tilde{\varepsilon}_i$

ii) Wie können Sie Ihre Vermutung testen? Nennen Sie Testverfahren, Null- und Alternativhypothese. (1,5 Punkte)

F-Test auf gemeinsame Signifikanz

Hypothesen:

$H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

H_1: mindestens ein $\beta_j \neq 0$ mit j=4, 5, 6, 7

Aufgabe 5: **[40 Punkte]**

Welche Antwort ist richtig? Kreuzen Sie nur **eine Antwort** pro Aufgabe an. Falls mehrere Aussagen korrekt sind, kreuzen Sie **nur** die entsprechende **Antwortkombination** an. Für jede richtige Antwort gibt es 1 Punkt. Für falsche Antworten werden keine Punkte abgezogen.

1.	Eine Bernoulli verteilte Zufallsvariable	
a	☑	nimmt Werte von 0 oder 1 an.
b	❑	nimmt jeden einzelnen Wert mit der Wahrscheinlichkeit 0 an.
c	❑	ist eine stetige Zufallsvariable.
d	❑	kann durch die Eintrittswahrscheinlichkeit θ von X = 1 wie folgt vollständig beschrieben werden: P(X=1) = θ und P(X=0) = θ-1.
e	❑	b und c.

2.	Wenn eine diskrete Zufallsvariable X die Werte -1, 0 und 1 mit Wahrscheinlichkeit ¼, ½ und ¼ annimmt, dann gilt	
a	❑	$E(X) = 0$.
b	❑	$E(X^2) = 0$.
c	❑	$E(X^2) = [E(X)]^2$.
d	❑	$E(X^2) \neq [E(X)]^2$.
e	☑	a und d.

3.	Eine standardisierte Zufallsvariable	
a	❑	hat den Erwartungswert von 1 und eine Varianz von 0.
b	❑	hat den Erwartungswert von 1 und eine Varianz von 1.
c	❑	hat den Erwartungswert von 1 und eine Standardabweichung von 0.
d	☑	hat den Erwartungswert von 0 und eine Standardabweichung von 1.
e	❑	a und c.

4.	Die Normalverteilung $N(\mu, \sigma^2)$	
a	❑	ist symmetrisch um den Erwartungswert 0.
b	❑	hat einen Erwartungswert μ, der gleich dem Median ist.

c	❑	mit $\mu = 0$ und $\sigma^2 = 1$ heißt Standardnormalverteilung.
d	❑	a und c.
e	☑	b und c.

5.	Das angepasste Bestimmtheitsmaß $\bar{R}^2$	
a	❑	kann nicht negativ werden.
b	❑	kann genauso interpretiert werden wie das Bestimmtheitsmaß R^2.
c	☑	kann verwendet werden, um genestete Modelle miteinander zu vergleichen.
d	❑	wird durch eine Umskalierung der abhängigen Variable beeinflusst.
e	❑	a, c und d.

6.	Beim Umskalieren der abhängigen Variable im Modell $y = \beta_0 + \beta_1 x + u$ ändern sich	
a	❑	die geschätzten Koeffizienten.
b	❑	die Konfidenzintervalle für beide Koeffizienten.
c	❑	die Signifikanzniveaus nicht.
d	❑	das Bestimmtheitsmaß R^2 und das angepasste R^2 nicht.
e	☑	alle der genannten Antworten.

7.	Unter Homoskedastie	
a	❑	liegt ein Endogenitätsproblem vor.
b	☑	ist die Varianz des Fehlerterms u für alle Beobachtungen identisch.
c	❑	sind die Parameterschätzer verzerrt.
d	❑	ist ein Signifikanztest nicht gültig.
e	❑	a und b.

8.	Ein nicht-linearer Zusammenhang zwischen zwei Variablen	
a	❑	kann in einem einfachen Regressionsmodell abgebildet werden.
b	❑	kann in einem multiplen Regressionsmodell abgebildet werden.
c	❑	wird mit dem Korrelationskoeffizienten gemessen.
d	❑	wird mit dem F-Test gemessen.
e	☑	a und b.

9.	Ein exogener Regressor	
a	☑	ist mit dem Störterm unkorreliert.
b	❑	verursacht omitted variable bias.
c	❑	verletzt die Annahme $E[u\|x]=0$.
d	❑	führt zur Verzerrung.
e	❑	b, c und d.

10.	Durch Aufnahme einer irrelevanten Dummyvariable	
a	❑	entsteht die dummy variable trap.
b	☑	kann das Bestimmtheitsmaß R^2 nicht fallen.
c	❑	werden die Parameter effizient geschätzt.
d	❑	b und c.
e	❑	keine der Antworten.

11.	Führt die zusätzliche Berücksichtigung einer erklärenden Variable im Modell zu nicht-perfekter Multikollinearität, so	
a	❑	ist eine KQ-Schätzung nicht durchführbar.
b	❑	ist das Bestimmtheitsmaß $R^2 =1$.
c	❑	führt eine KQ-Schätzung zu verzerrten Parametern.
d	☑	kann die Präzision einer KQ-Schätzung fallen.
e	❑	a und b.

12.	Die Modelle $y = \beta_0 + \beta_1\sqrt{x_1} + u$ und $y = \beta_0 + \beta_1\sqrt{x_1} + \beta_2 x_2 + \tilde{u}$	
a	❑	sind nicht-linear in Parametern.
b	❑	sind multiple lineare Modelle.
c	☑	sind genestet.
d	❑	b und c.
e	❑	alle der genannten Antworten.

13.	Im Modell $\text{Lohn} = \beta_0 + \beta_1\text{Frau} + \beta_2\text{Alter} + \beta_3(\text{Frau} \cdot \text{Alter}) + u$ bei dem Frau ein Indikator für das Geschlecht ist und Alter stetig das Alter in Jahren misst	
a	❑	misst β_0 den durchschnittlichen Lohn von alten Männern.
b	❑	werden unterschiedliche Achsenabschnittsparameter für Männer und Frauen bestimmt.
c	❑	werden unterschiedliche Steigungsparameter des Alters für Männer und Frauen bestimmt.
d	☑	b und c.
e	❑	alle der genannten Antworten.

14.	Im Modell $\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 \log(x_2) + u$	
a	☑	kann β_1 als Semielastizität interpretiert werden.
b	❑	kann β_1 als Elastizität interpretiert werden.
c	❑	ändert sich bei einer einprozentigen Änderung von x_2 y um $100\beta_2$ Prozent.
d	❑	ändert sich bei einer einprozentigen Änderung von x_2 y um β_2 Prozentpunkte.
e	❑	a und c.

15.	Wenn das Modell $y = \beta_1 x_1 + \beta_2 x_2 + u$ mit der KQ-Methode geschätzt wird,	
a	❑	kann man das R^2 als Maß der Schätzgüte verwenden.
b	☑	ist $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2$ der vorhergesagte Wert von y an den Stellen $x_1 = 1$ und $x_2 = 1$.
c	❑	ist der Mittelwert der Residuen gleich Null.
d	❑	a und b.
e	❑	keine der Antworten.

16.	Ergibt sich nach einer KQ-Schätzung für die erste Beobachtung ein Residuum von $\hat{u}_1 = -1$,	
a	❑	so wird y_1 unterschätzt.
b	☑	so wird y_1 überschätzt.
c	❑	so liegt y_1 auf der Regressionsgerade.
d	❑	so ist die Annahme $E(u\|x) = E(u) = 0$ verletzt.
e	❑	a und d.

17.	Bei konsistenten Schätzverfahren	
a	❑	ist der Schätzer stets unverzerrt.
b	❑	ist der Erwartungswert des Schätzers stets mit dem wahren Wert identisch.
c	❑	liegt der Wahrscheinlichkeitsgrenzwert des Schätzers umso näher am wahren Wert, je größer die Stichprobe.
d	❑	sinkt die Varianz des Schätzers, je größer die Stichprobe.
e	☑	c und d.

18.	Gilt $E(u\|x) = 0$ im einfachen Regressionsmodell dann	
a	❑	$\text{Cov}(x, u) = 0$.
b	❑	ist x eine exogene Variable.
c	❑	kann der KQ-Schätzer als kausal interpretiert werden.
d	❑	a und c.
e	☑	alle der genannten Antworten.

19.	Die BLUE-Eigenschaft des KQ-Schätzers	
a	❑	impliziert seine Unverzerrtheit.
b	❑	gilt nicht,wenn die Residuen nicht normalverteilt sind.
c	❑	gilt nicht wenn perfekte Kollinearität vorliegt.
d	❑	gilt nicht, wenn der Störterm heteroskedastisch ist.
e	☑	a, c und d.

20.	Die Stichprobenvarianz	
a	❑	beschreibt die Summe der quadrierten Abweichungen einer Variable um ihren Mittelwert.
b	☑	ist ein Dispersionsmaß.
c	❑	kann negative Werte annehmen.
d	❑	ist ein erwartungstreuer Schätzer für die Varianz des Störterms.
e	❑	keine der Antworten.

21.	Wenn der p-Wert für den Signifikanztest eines Parameters den Wert 0,04 annimmt, dann	
a	☑	wird H_0 am 5% Niveau verworfen.

b	❑	wird H_0 am 1% Niveau verworfen.
c	❑	beträgt der Typ II Fehler 4%.
d	❑	beträgt der Typ I 2%.
e	❑	keine der Antworten.

22.	Die Annahme $x \sim N(\mu, \sigma^2)$ impliziert	
a	❑	$Var(x) = \sigma$.
b	❑	$E(x) = 0$.
c	❑	$E(\mu) = x$.
d	❑	$E(x) = \sigma$.
e	☑	keine der Antworten.

23.	Bei zwei unverzerrten Schätzern $\hat{\beta}_1$ und $\tilde{\beta}_1$ für einen unbekannten Bevölkerungsparameter β_1 ist $\tilde{\beta}_1$ effizient, wenn	
a	❑	$E(\tilde{\beta}_1) < E(\hat{\beta}_1)$.
b	☑	$\sqrt{Var(\tilde{\beta}_1)} < \sqrt{Var(\hat{\beta}_1)}$.
c	❑	$Var(\tilde{\beta}_1) > Var(\hat{\beta}_1)$.
d	❑	a und b.
e	❑	keine der Antworten.

24.	Der zentrale Grenzwertsatz besagt, dass für eine Zufallsvariable Y mit Mittelwert μ und Varianz σ^2, gilt:	
a	❑	$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ folgt asymptotisch der F-Verteilung.
b	☑	$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ folgt asymptotisch der Standardnormalverteilung.
c	❑	$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ folgt asymptotisch der Bernoulli-Verteilung.
d	❑	$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ folgt asymptotisch der Chi-Quadrat-Verteilung.
e	❑	keine der Antworten.

25.	Die Varianz des Störterms in der Grundgesamtheit	
a	❑	steigt mit steigender Stichprobengröße.
b	☑	kann nach einer KQ-Schätzung anhand der unerklärten Quadratsumme der abhängigen Variable approximiert werden.
c	❑	ist eine Funktion der erklärenden Variablen.
d	❑	kann anhand der erklärten Quadratsumme der unabhängigen Variable approximiert werden.
e	❑	a und b.

26.	Wenn in einem einfachen linearen Modell die abhängige Variable y eine Dummyvariable ist, dann	
a	❑	sind die Parameter des Modells nicht mehr unverzerrt.
b	❑	variiert die Varianz des Störterms über die Beobachtungseinheiten.
c	❑	$\Delta P(y = 1 \mid x) = \beta_j \Delta x_i$.
d	❑	a und b.
e	☑	b und c.

27.	Ein p-Wert	
a	❑	ist bei einseitigen Tests doppelt so groß wie bei zweiseitigen Tests.
b	☑	beschreibt das Signifikanzniveau eines Tests, bei dem die berechnete Teststatistik gleich dem kritischen Wert ist.
c	❑	ist für t- aber nicht für F-Tests relevant.
d	❑	a und b.
e	❑	keine der Antworten.

28.	Im Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ kommt es bei Weglassen von x_2 zu einer negativen Verzerrung von β_1, wenn: (Hilfsregression: $x_1 = \delta_0 + \delta_2 x_2 + \varepsilon$)	
a	❑	$\beta_2 < 0, \delta_2 > 0$.
b	❑	$\beta_2 < 0, \delta_2 = 0$.
c	❑	$\beta_2 < 0, \delta_2 < 0$.
d	❑	$\beta_2 > 0, \delta_2 < 0$.
e	☑	a und d.

29.	Ein Chow-Test	
a	❑	kann durch Verwendung von geeigneten Interaktionstermen durchgeführt werden.
b	❑	prüft, ob Regressionskoeffizienten nach Gruppen unterschiedlich ausfallen.
c	❑	kann durch Schätzung getrennter Regressionen für zwei Gruppen durchgeführt werden.
d	❑	a und c.
e	☑	alle der Antworten.

30.	Man erhält den KQ-Schätzer im einfachen linearen Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ durch die Minimierung folgender Zielfunktion:	
a	❑	$\sum_{i=1}^{n}(\hat{u}_i)^2$.
b	❑	$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$.
c	❑	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$.
d	❑	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
e	☑	a, b und d.

31.	In einer quadratischen Spezifikation $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ ergibt sich ein umgekehrt u-förmiger Verlauf der Regressionsgeraden, wenn	
a	☑	$\beta_1 > 0, \beta_2 < 0$.
b	❑	$\beta_1 > 0, \beta_2 > 0$.
c	❑	$\beta_1 < 0, \beta_2 > 0$.
d	❑	$\beta_1 = 0, \beta_2 < 0$.
e	❑	keine der Antworten.

32.	Im linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + \beta_2 {x_i}^2 + u_i$ (mit $\beta_1 > 0$)	
a	❑	steigt der marginale Effekt von x auf y mit steigendem x, wenn $\beta_2 > 0$.
b	❑	variiert der marginale Effekt von x auf y mit x, wenn $\beta_2 \neq 0$.
c	❑	ist der marginale Effekt von x auf y konstant, wenn $\beta_2 = 0$.
d	❑	variiert der marginale Effekt von x auf y mit x, wenn $\beta_2 < 0$.
e	☑	alle der genannten Antworten.

33.	Für eine unverzerrte Schätzung des wahren Bevölkerungsparameters mittels KQ ist nicht erforderlich,	
a	❑	dass der Erwartungswert der Konstante null ist.
b	❑	dass die Stichprobe gegen unendlich geht.
c	❑	dass die Störterme normalverteilt sind.
d	❑	dass der Erwartungswert des Störterms unabhängig von den erklärenden Variablen ist.
e	☑	a, b und c.

34.	Beim Umskalieren der abhängigen Variable mit einem konstanten multiplikativen Faktor im Modell $y = \beta_0 + \beta_1 x + u$ ändert sich	
a	❑	der Steigungsparameter.
b	❑	der Standardfehler der Konstanten.
c	❑	SSR (die Quadratsumme der Residuen).
d	❑	a und b.
e	☑	alle Antworten sind zutreffend.

35.	Bei der Regressionsfunktion $y = \beta_0 + \beta_1 cx + \beta_2 (cx)^2 + u$ (*c*=konstanter Skalar) kann der marginale Effekt der Variablen *x* mit steigendem *x*	
a	❑	fallen.
b	❑	steigen.
c	❑	konstant bleiben.
d	☑	a, b, und c.
e	❑	keine der Antworten.

36.	Werden irrelevante Variablen ins lineare Regressionsmodell aufgenommen, so	
a	❑	werden verzerrte Parameterschätzer ausgewiesen.
b	❑	verringert sich die Effizienz der Parameterschätzer.
c	❑	sinkt das angepasste Bestimmtheitsmaß.
d	❑	erhöht sich die Effizienz der Parameterschätzer.

e	☑	b und c.

37.	Die Präzision der Schätzung des Steigungsparameters β_1 im einfachen Regressionsmodell	
a	❑	steigt, wenn die Störtermvarianz zunimmt.
b	☑	steigt, wenn die Variation in x zunimmt.
c	❑	variiert mit dem Bevölkerungswert der Regressionskonstante.
d	❑	fällt, wenn der Stichprobenumfang steigt.
e	❑	b und d.

38.	Gepoolte Querschnittsdaten unterscheiden sich von Zeitreihendaten dadurch, dass	
a	❑	gepoolte Querschnittsdaten eine gegebene Beobachtungseinheit mehrfach betrachten.
b	❑	gepoolte Querschnittsdaten mehr Beobachtungen haben.
c	☑	gepoolte Querschnittsdaten einzelne Beobachtungseinheiten nicht wiederholt betrachten.
d	❑	a und c.
e	❑	keine der Antworten.

39.	Die Eigenschaft der Konsistenz	
a	❑	kann unter schwächeren als den Gauss-Markov-Annahmen nachgewiesen werden.
b	❑	setzt voraus, dass ein Schätzer normalverteilt ist.
c	❑	ist eine asymptotische Eigenschaft.
d	❑	a und b.
e	☑	a und c.

40.	Ein Schätzer $\hat{\beta}$ für den unbekannten Parameter β ist konsistent wenn	
a	❑	$n \to \infty$.
b	❑	$E(\hat{\beta}) = \beta$
c	❑	$plim(\hat{\beta}) \to \hat{\beta}$ für $n \to \infty$.
d	❑	b und c.
e	☑	keine der Antworten.