Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Prof. Regina T. Riphahn, Ph.D.

Bachelorprüfung WS 2012/13 - 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: **[10 Punkte]**

Sie untersuchen Kriminalitätsraten in den USA und wollen diese erklären. Ihnen steht dazu ein Datensatz von 47 US-Staaten mit den folgenden Informationen zur Verfügung:

Crime	Kriminalitätsrate gemessen als Anzahl von Vergehen pro 1 Mio. Einwohner
Age	Anteil der Männer im Alter zwischen 14 und 24 an der Bevölkerung (in %)
Unemp	Jugend-Arbeitslosenquote (in %)
Exp	Ausgaben für Bildung und Soziales als Anteil am Buttoinlandsprodukt (in %)

Sie schätzen das folgende Modell:

$$Crime_i = \beta_0 + \beta_1 Age_i + \beta_2 Unemp_i + \beta_3 Exp_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten		T	Sig.
	RegressionskoeffizientB	Standardfehler		
(Konstante)	-167,002	70,008	-2,39	0,022
Age	1,085	,376	2,88	0,006
Unemp	1,259	2,259	,557	0,592
Exp	-1,323	1,240	-1,07	0,269

a. Abhängige Variable: Crime

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

> Wenn der Anteil der Männer im Alter zwischen 14 und 24 Jahren an der Bevölkerung um einen Prozentpunkt steigt, so erhöht sich die Kriminalitätsrate c.p. im Mittel um 1,085 Vergehen pro 1 Mio. Einwohner.
>
> Der Koeffizient ist statistisch signifikant am 1% Niveau.

b) Die Regierung prognostiziert für das Jahr 2013 einen Anstieg der Jugendarbeitslosigkeit um 5% im Vergleich zum Vorjahr. Im Jahr 2012 lag die Jugendarbeitslosenquote bei 10%. Welche Auswirkungen hat diese Veränderung auf die Kriminalitätsrate? Berechnen Sie den Effekt auf $crime_i$ und interpretieren Sie das Ergebnis inhaltlich. (3.5 Punkte)

> - Berechnung der Änderung: $10 \times 0,05 = 0,5$ Prozentpunkte
> - Auswirkung: $\Delta \widehat{Crime}_i = \hat{\beta}_2 \times 0,5 = 1,259 \times 0,5 = 0,629$
> - Interpretation: Wenn sich die Jugendarbeitslosigkeit um 5 Prozent (bzw. um 0,5 Prozentpunkte) erhöht, so erhöht sich die Kriminalitätsrate c.p. im Mittel um 0.629 Vergehen pro 1 Mio. Einwohner.

c) Um wie viel müsste der Anteil der Ausgaben für Bildung und Soziales am Bruttoinlandsprodukt steigen, damit die negative Entwicklung aus b) ausgeglichen wird? Berechnen und interpretieren Sie inhaltlich die nötige Steigerung des Ausgabenanteils. (*Hinweis:* Verwenden Sie für die Änderung in *Crime* die fiktive Änderung $\Delta \widehat{Crime}_i = 5$, falls die Lösung in b) nicht berechnet werden konnte.) (4.5 Punkte)

- $\Delta \hat{Crime}_i = \hat{\beta}_3 \times \Delta Exp \rightarrow \Delta Exp = \frac{\Delta \hat{Crime}_i}{\hat{\beta}_3}$
- mit $\Delta \hat{Crime}_i = -0,629$ wird der Effekt aus b) ausgeglichen.
- $\Delta Exp = \frac{-0,629}{-1,323} = 0,475$
- Interpretation: Um die negative Entwicklung in der Kriminalitätsrate auszugleichen, müsste der Anteil der Ausgaben für Bildung und Soziales am BIP um 0,475 Prozentpunkte steigen.

Hilfsgröße:

- mit $\Delta \hat{Crime}_i = -5$ wird der Effekt aus b) ausgeglichen.
- $\Delta Exp = \frac{-5}{-1,323} = 3,779$
- Interpretation: Um die negative Entwicklung in der Kriminalitätsrate auszugleichen, müsste der Anteil der Ausgaben für Bildung und Soziales am BIP um 3,779 Prozentpunkte steigen.

Aufgabe 2: **[15 Punkte]**

Sie interessieren sich für Kinofilmbewertungen. Für Ihre Untersuchung haben Sie zufällig 100 Filme ausgewählt. Für jeden Film stehen Ihnen folgende Informationen zur Verfügung:

Rating	Rating des Films auf einer Skala stetig von 1 bis 4 kodiert, wobei 1=schlecht und 4=gut
Year	Erscheinungsjahr (4-stellig)
Length	Länge des Films in Minuten
Length2	Länge des Films quadriert / 100 = $\frac{Length^2}{100}$
Descr	Anzahl der Zeilen Text, die zur Beschreibung des Films notwendig sind

Sie schätzen das folgende Modell mit SPSS:

$$Rating_i = \beta_0 + \beta_1 Year_i + \beta_2 Length_i + \beta_3 Length2_i + u_i$$

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	11,205	3	3,735	9,607	,000[b]
	Nicht standardisierte Residuen	37,323	96	?		
	Gesamt	48,528	99			

a. Abhängige Variable: Rating

b. Einflußvariablen : (Konstante), Length2, Year, Length

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	32,247	7,773	4,148	,000
	Year	-,016	,004	-3,895	,000
	Length	,023	,026	,884	,379
	Length2	-,001	,013	-,096	,924

a. Abhängige Variable: Rating

a) Berechnen und interpretieren Sie das Bestimmtheitsmaß R^2. (2 Punkte)

- $R^2 = \frac{SSE}{SST} = \frac{11{,}205}{48{,}528} = 0{,}231$
- `23,1% der Variation in den Filmratings werden durch das Modell erklärt.`

b) Berechnen Sie die geschätzte Störtermvarianz $\hat{\sigma}^2$. (1 Punkt)

$\hat{\sigma}^2 = \frac{SSR}{N-k-1} = \frac{37{,}323}{96} = 0,389$

c) Berechnen und interpretieren Sie inhaltlich den marginalen Effekt einer zusätzlichen Minute Filmlänge bei einer Filmlänge von 2 Stunden. (3 Punkte)

- $\Delta \hat{Rating}_i = \hat{\beta}_2 + \hat{\beta}_3 \frac{Length_i}{100} \times 2$
- $\Delta \hat{Rating}_i = 0,023 - 0,001 \times \frac{120}{100} \times 2 = 0,023 - 0,0024 = 0,021$
- Interpretation: Bei einer Filmlänge von 2 Stunden erhöht eine zusätzliche Minute Filmlänge das Rating um c.p. im Mittel 0,021 Skalenpunkte.

d) Der Spielfilm *Der Hobbit* kam im Dezember 2012 in die Kinos. Der Film hat eine Gesamtlänge von 2 Stunden und 49 Minuten. Sagen Sie auf Basis Ihres Modell das Rating des Films voraus. (2 Punkte)

- Dauer in Minuten: $60 \times 2 + 49 = 169$
- $\hat{Rating}_i = 32,247 - 0,016 \times 2012 + 0,023 \times 169 - 0,001 \frac{169^2}{100} = 3,656$
- Das erwartete Rating beträgt 3,656.

e) Ihr Kommilitone behauptet, dass entscheidende Erklärungsgrößen in Ihrem Modell fehlen. Er denkt, dass die Komplexität eines Filmes dessen Erfolg entscheidend beeinflusst und schlägt das folgende Modell vor:

$$Rating_i = \beta_0 + \beta_1 Year_i + \beta_2 Length_i + \beta_3 Length2_i + \beta_4 Descr_i + u_i$$

i. Unter welchen Bedingungen führt das Weglassen der Variable $Descr_i$ zu einer Verzerrung des Schätzers $\hat{\beta}_1$? (2.5 Punkte)

Der Schätzer $\hat{\beta}_1$ wird durch das Weglassen der Variable $Descr_i$ verzerrt, wenn

- $Descr_i$ und die Variable $Year_i$ korreliert sind ($Cov(Descr_i, Year_i) \neq 0$) und
- $Descr_i$ einen Effekt auf das Filmrating hat ($\beta_4 \neq 0$)

ii. Der geschätzte Wert für β_4 beträgt 0,08 mit einem Standardfehler von 0,02. Testen Sie am 5%-Niveau, ob der Koeffizient signifikant von Null verschieden ist. Geben Sie Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Entscheidung an. (4,5 Punkte)

- Hypothesen: $H_0 : \beta_4 = 0$, $H_1 : \beta_4 \neq 0$
- Teststatistik: $t = \frac{0{,}08}{0{,}02} = 4$
- Kritischer Wert: $c = t_{\frac{\alpha}{2}, n-k-1} = t_{0,025,100-4-1} = t_{0,025,95} =$ nicht tabelliert, nächst kleinerer Wert $= t_{0,025,90} = 1,987$
- $FG = n - k - 1 = 95$
- Die Nullhypothese kann am 5% Niveau abgelehnt werden.

Aufgabe 3: **[13,5 Punkte]**

Sie wollen untersuchen, wie sich die psychische Gesundheit einer erwerbstätigen Person auf deren Lohn auswirkt. Ihnen liegen Daten des sozio-ökonomischen Panels mit 15 696 Beobachtungen vor:

ln(wage)	Stundenlohn in Euro, logarithmiert
educ	Anzahl der Schuljahre
mental	Index der psychischen Gesundheit, gemessen auf einer Skala von 0 (schlecht) bis 100 (sehr gut)
vollzeit	=1 wenn Person in Vollzeit erwerbstätig ist; =0 sonst.

Sie schätzen das folgende Modell mit SPSS:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 mental_i + \beta_3 vollzeit_i + u_i$$

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	881,715	3	293,905	1210,916	,000[b]
	Nicht standardisierte Residuen	3808,652	15692	,243		
	Gesamt	4690,366	15695			

a. Abhängige Variable: lnwage

b. Einflußvariablen : (Konstante), vollzeit, mental, educ

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	6,572	,027	240,631	,000
	educ	,056	,002	35,826	,000
	mental	,006	,000	13,709	,000
	vollzeit	,309	,008	36,933	,000

a. Abhängige Variable: lnwage

a) Interpretieren Sie den den geschätzten Koeffizienten für *mental* inhaltlich. (1 Punkt)

- `Wenn der Index der psychischen Gesundheit um einen Punkt ansteigt, so steigt der Lohn c.p.i.M. um ca. 0,6% an.`

b) Der Stundenlohn $wage_i$ wird nun statt in Euro in 1000 Euro gemessen. Wie ändern sich,

i. die Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_3$? (3 Punkte)

- Neues Modell: Da $ln(\frac{wage_i}{1000}) = \tilde{\beta}_0^{neu} + \tilde{\beta}_1^{neu} educ_i + \tilde{\beta}_2^{neu} mental_i + \tilde{\beta}_3^{neu} vollzeit_i + u_i$
- ergibt sich $ln(wage_i) - ln(1000) = \tilde{\beta}_0 + \tilde{\beta}_1 educ_i + \tilde{\beta}_2 mental_i + \tilde{\beta}_3 vollzeit_i + u_i$
- und somit ist $\tilde{\beta}_0^{neu} = \hat{\beta}_0^{alt} + ln(1000)$.
- $\hat{\beta}_3^{alt} = \tilde{\beta}_3^{neu}$. Keine Änderung

ii. der p-Wert von $\hat{\beta}_3$? (1 Punkt)

Der p-Wert von $\hat{\beta}_3$ bleibt gleich, da t-Statistik unverändert.

c) Berechnen Sie ein 90% Konfidenzintervall für den Paramter von educ ($\hat{\beta}_1$) und interpretieren Sie es. Zeigen Sie Ihren Rechenweg und runden Sie auf die dritte Nachkommastelle. (3,5 Punkte)

- Konfidenzintervall von $\hat{\beta}_1$ lautet $[\hat{\beta}_1 \pm c \cdot se(\hat{\beta}_1)]$
- $c = t_{\frac{0,1}{2};\infty} = 1,645$
- $[0,056 \pm 1,645 \cdot 0,002] = [0,053; 0,059]$
- Bei wiederholter Stichprobenziehung liegt der wahre Parameter in 90% der Fälle innerhalb der auf dieser Weise bestimmten Konfidenzintervalle.

d) Sie nehmen zusätzlich noch die Variablen *female* und *age* auf und schätzen folgendes Modell:

$$ln(wage)_i = \beta_0 + \beta_1 educ_i + \beta_2 mental_i + \beta_3 work_i + \beta_4 female_i + \beta_5 age_i + u_i$$

Die Residuenquadratsumme (SSR) des neuen Modells beträgt 3804.26. Testen Sie am 10%-Niveau, ob die Koeffizienten der neuen Variablen gemeinsam statistisch signifikant sind. Geben Sie Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und ihre Testentscheidung an. (5 Punkte)

- Hypothese: H0: $\beta_4 = \beta_5 = 0$; H1: mind. ein $\beta_j \neq 0$ mit j=4,5.
- Teststatistik: $F_{emp} = \frac{(S_r - S_{ur})/q}{S_{ur}/(n-k-1)} = \frac{(3808,65 - 3804,26)/2)}{(3804,26/(15696-5-1)} = 9,05$
- Zählerfreiheitsgrade: 2 (Restriktionen)
- Nennerfreiheitsgrade: n-k-1= 15696-5-1=15690
- kritischer Wert: $F_c = F_{0,1;2;15690} = 2,30$
- Entscheidung: Da $F_{emp} > F_c$ wird die Nullhypothese verworfen. Die Koeffizienten der Variablen *female* und *age* sind gemeinsam statistisch signifikant von Null verschieden.

Aufgabe 4: **[11,5 Punkte]**

Sie möchten die Determinanten des Körpergewichts untersuchen und verfügen über einen Datensatz, der Informationen zu 73 456 Individuen und einer Reihe von erklärenden Größen enthält:

gewicht	Körpergewicht, gemessen in Kilogramm
smoker	Dummy=1 wenn Raucher; =0 wenn nicht.
keinsport	Dummy=1 wenn Person keinen Sport macht; =0 wenn nicht.
ln(hhinc)	logarithmiertes Haushaltseinkommen in Euro

Sie unterstellen die Gültigkeit der Gauss-Markov Annahmen und schätzen das folgende Modell mit SPSS:

$$gewicht_i = \beta_0 + \beta_1 smoker_i + \beta_2 keinsport_i + \beta_3 ln(hhinc_i) + u_i$$

ANOVA[a]

Modell		Quadratsumme	df	F	Sig.
1	Regression	**?**	3	30,064	,000[b]
	Nicht standardisierte Residuen	17682,387	73452		
	Gesamt	17704,099	73455		

a. Abhängige Variable: gewicht

b. Einflußvariablen : (Konstante), ln_hhinc, smoker, keinsport

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		Regressionskoeffizient B	Standardfehler		
1	(Konstante)	70,544	**?**	85,405	,000
	smoker	-,409	,125	-3,266	,001
	keinsport	1,306	,194	6,721	,000
	ln_hhinc	,662	,107	6,213	,000

a. Abhängige Variable: gewicht

a) Berechnen Sie in der Tabelle die beiden mit „?“ bezeichneten fehlenden Werte. (2 Punkte).

- $SE(\hat{\beta}_0) = \frac{70,54}{85,405} = 0,826$.
- $SSR = SST - SSE = 17704,099 - 17682,387 = 21,712$.

b) Sie nehmen zusätzlich eine irrelevante Variable in ihr Modell auf, die mit *smoker* positiv korreliert ist. Was sind die Konsequenzen für die Schätzung von β_1? (2 Punkte)

- Da die ausgelassene Variable irrelevant ist, d.h. keinen Einfluss auf die abhängige Variable hat, bleibt $\hat{\beta}_1$ unverzerrt.
- Der Standardfehler von $\hat{\beta}_1$ wird ansteigen, die Effizienz der Schätzung fällt.

c) Wie verändert sich das Körpergewicht, wenn das Haushaltseinkommen um 3,5% ansteigt? (2.5 Punkte)

- $\Delta \widehat{gewicht}_i = \hat{\beta}_3 \times \Delta ln(hhinc) = \frac{3,5 \times 0,662}{100} = 0,023$
- Das Körpergewicht steigt c.p.i.M. um ca. 23 Gramm an, wenn das Haushaltseinkommen um 3,5% ansteigt.

d) Sie möchten überprüfen, ob sich der Effekt der sportlichen Aktivität für Raucher und Nichtraucher statistisch signifikant unterscheidet.

i. Geben Sie eine Regressionsgleichung für ein Modell an, die Ihnen das gewünschte Ergebnis liefert. (1 Punkt)

Einfügen eines Interaktionsterms von Raucher und sportlicher Aktivität:

$$gewicht_i = \beta_0 + \beta_1 smoker_i + \beta_2 keinsport_i + \beta_3 ln(hhinc)_i + \beta_4 smoker_i \times keinsport_i + u_i$$

ii. Leiten Sie den marginalen Effekt der sportlichen Aktivität für das neue Modell her. (1 Punkt)

ME: $\frac{\delta gewicht}{\delta keinsport} = \beta_2 + \beta_4 smoker$

e) Sie vermuten, dass sich alle Parameter des Modells für Männer und Frauen unterscheiden. Stellen Sie ein Modell auf, mit dem Sie diese Vermutung testen können. (3 Punkte)

Schätzung eines vollständig interagierten Modells:

$$gewicht_i = \beta_0 + \beta_1 smoker_i + \beta_2 keinsport_i + \beta_3 ln(hhinc)_i +$$

$$\beta_4 female_i + \beta_5 smoker_i \times female_i + \beta_6 keinsport_i \times female_i + \beta_7 ln(hhinc)_i \times female_i + u_i$$

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.	Das Bestimmtheitsmaß R^2 im einfachen linearen Regressionsmodell
a	**X** ist größer als das korrigierte Bestimmtheitsmaß.
b	berücksichtigt die zur Schätzung benötigten Freiheitsgrade.
c	gibt das Verhältnis von unerklärter Variation zur Gesamtvariation in der Konstanten an.
d	kann negative Werte annehmen.
e	a und c.

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

3.	Ein Omitted Variable Bias entsteht, wenn
a	irrelevante Variablen in ein Modell aufgenommen werden.
b	die Schätzung ineffizient ist.
c	die Variation in der abhängigen Variable sehr gering ist.
d	perfekte Multikollinearität besteht.
e	**X** keine der genannten Antworten.

4.	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	$\sum_{i=1}^{n} (\bar{u}_i)^2$
b	**X** $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
c	$\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2$
d	$\sum_{i=1}^{n} (\bar{y}_i - y_i)^2$
e	a und b.

5.	Kausale Effekte
a	können im einfachen linearen Regressionsmodell nicht geschätzt werden.
b	sind als Basis für Politikempfehlungen grundsätzlich ungeeignet.
c	lassen sich mit Befragungsdaten schätzen.
d	können sich von Korrelationen unterscheiden.
e	**X** c und d.

6.	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.
e	a und c

7.	Wenn Heteroskedastie vorliegt,
a	kann der KQ-Schätzer nicht mehr berechnet werden.
b	ist der KQ-Schätzer BLUE.
c	**X** ist der KQ-Schätzer ineffizient aber erwartungstreu wenn $E[u_i\|x_i] = E[u_i] = 0$.
d	ist der KQ-Schätzer ineffizient und verzerrt.
e	b und d.

8.	Ein relevanter endogener Regressor
a	beeinträchtigt die Konsistenz einer Schätzung.
b	verletzt die Annahme $E[u\|x] = 0$.
c	korreliert mit dem Störterm.
d	führt zur Verzerrung.
e	**X** alle der Antworten.

9.	Das Bestimmtheitsmaß R^2
a	wird bei einer KQ-Schätzung minimiert.
b	**X** kann durch die erklärte Variation und die gesamte Variation in der abhängigen Variable berechnet werden.
c	kann nicht verwendet werden, um F-Tests durchzuführen.
d	wird genauso interpretiert wie das angepasste Bestimmtheitsmaß $\overline{R}^2$.
e	keine der genannten Antworten.

10.	Stetige Zufallsvariablen
a	nehmen mit einer Wahrscheinlichkeit von 1 eine bestimmte Realisation an.
b	können nicht der Normalverteilung folgen.
c	können nur die Werte 0 oder 1 annehmen.
d	sind immer symmetrisch um ihren Erwartungswert.
e	**X** keine der genannten Antworten

11.	Werden KQ-Vorhersagen für die abhängige Variable y im Modell $y = \beta_0 + \beta_1 x + u$ generiert,
a	entspricht der Mittelwert der Vorhersagen $\hat{y}$ immer dem Mittelwert der abhängigen Variable $\bar{y}$.
b	ist die Stichprobenkovarianz zwischen jeder unabhängigen Variable und den KQ-Residuen Null.
c	ist die Stichprobenkovarianz zwischen Vorhersagen $\hat{y}$ und den KQ-Residuen Null.
d	kann für diese kein Konfidenzintervall bestimmt werden.
e	**X** a, b und c

12.	Wenn Homoskedastie vorliegt,
a	**X** können die Parameter unverzerrt geschätzt werden.
b	ist das angepasste R^2 nicht mehr gültig.
c	sind die Standardfehler der Parameter normalverteilt.
d	ist die Varianz des Fehlerterms u für alle Beobachtungen unterschiedlich.
e	keine der Antworten.

13.	Sie betrachten Erwartungswert E und Varianz Var einer Zufallsvariablen X. Es gilt:
a	$E[X^2] = 2E[X]$
b	$Var[a \cdot X] = a^2 Var[X]$ mit a konstant
c	$E[a + X] = a + E[X]$ mit a konstant.
d	$Var[a + X] = a \cdot Var[X]$ mit a konstant.
e	**X** b und c.

14.	Im Modell $log(y) = \beta_0 + \beta_1 log(x) + u$
a	kann x auch eine Dummyvariable sein.
b	**X** steigt y um β_1 Prozent an, wenn x um 1 Prozent erhöht wird.
c	ist β_1 der marginale Effekt von log(x) auf y.
d	steigt y um β_1 Prozentpunkte an, wenn x um 1 Prozent erhöht wird.
e	keine der Antworten.

15.	Der geschätzte Koeffizient $\hat{\beta}_1$ des Modells $y = \beta_0 + \beta_1 x_1 + u$ ist nach unten verzerrt, falls eine ausgelassene Variable x_2
a	positiv mit x_1 korreliert ist und y positiv beeinflusst.
b	negativ mit x_1 korreliert ist und y negativ beeinflusst.
c	nicht mit x_1 korreliert ist und y negativ beeinflusst.
d	**X** negativ mit x_1 korreliert ist und y positiv beeinflusst.
e	b und d.

16.	Im Modell $y = \beta_0 + \beta_1 log(x) + u$
a	wird β_1 als Elastizität interpretiert.
b	**X** gibt β_1 den marginalen Effekt von $log(x)$ auf y an.
c	ist der marginale Effekt von x auf y unabhängig von x.
d	gibt $\beta_1 \cdot 100$ den marginalen Effekt von x auf y an.
e	c und d.

17.	Geht in die Spezifikation eines Modells eine Dummy-Variable als erklärende Variable ein,
a	fällt die Schätzgüte des Modells (R^2).
b	**X** misst diese einen Niveauunterschied in der abhängigen Variable für zwei Gruppen.
c	darf die abhängige Variable nicht logarithmiert sein.
d	sinkt die Effizienz der Schätzung.
e	b und d.

18.	Der marginale Effekt von x_1 im Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2^2 + u_i$ lautet
a	$\beta_1 + \beta_3$
b	β_1
c	$\beta_1 + 2\beta_3 x_2$
d	**X** $\beta_1 + \beta_3 x_2^2$
e	keine der Antworten.

19.	Im log-linearen Modell $log(y) = \beta_0 + \beta_1 x + u$
a	wird β_1 als Elastizität interpretiert.
b	hängt der marginale Effekt der Variable x von deren Ausprägung ab.
c	**X** ist die abhängige Variable $log(y)$ eine lineare Funktion von x_i.
d	ist y_i eine konvexe Funktion von x_i.
e	a und c.

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

21.	Wenn in dem Regressionsmodell $log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ die erklärende Variable x_2 umskaliert wird, ändern sich
a	das R^2.
b	alle geschätzten Parameter.
c	die Signifikanzniveaus nicht.
d	der geschätzte Parameter $\hat{\beta}_2$.
e	**X** c und d.

22.	Wenn im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ der KQ-Schätzer $\hat{\beta}_1 = 0,5$ mit dem Standardfehler $se(\hat{\beta}_1) = 0,1$ geschätzt wird, so
a	steigt y im Mittel c.p. um 50%, wenn x um eine Einheit steigt.
b	**X** beträgt die empirische t-Statistik für $\hat{\beta}_1$ bei einem Signifikanztest $t = 5$.
c	beträgt die empirische t-Statistik für $\hat{\beta}_1$ bei einem Signifikanztest $t = 0,2$.
d	steigt y im Mittel c.p. um 0,5%, wenn x um 1% steigt.
e	keine der genannten Antworten.

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

24.	Dummy-Variablen
a	können die Werte 0 und 2 annehmen.
b	können mit anderen Variablen interagiert werden.
c	erhöhen die Fehlertermvarianz.
d	führen zu Heteroskedastie.
e	**X** a und b.

25.	Die Modelle $y = \beta_0 + \beta_1 \sqrt{x_1} + u$ und $y = \beta_0 + \beta_1 \sqrt{x_1} + \beta_2 x_2^2 + \beta_3 x_2^3 + \tilde{u}$ sind
a	nicht genestet.
b	**X** linear in Parametern.
c	multiple lineare Modelle.
d	lassen sich nur für stetige erklärende Variablen schätzen.
e	a und b.

26.	Im linearen Wahrscheinlichkeitsmodell
a	wird die abhängige Variable als Prozentsatz gemessen.
b	werden nur binär kodierte erklärende Variablen verwendet.
c	hat der Störterme für alle Werte der erklärenden Variablen die gleiche Varianz.
d	**X** gibt die Prognose $\hat{y}$ die Wahrscheinlichkeit des Ereignisses y=1 an.
e	c und d.

27.	Wenn das Modell $y = \beta_1 x_1 + \beta_2 x_2 + u$ mit der KQ-Methode geschätzt wird,
a	sind die Fehlerterme heteroskedastisch.
b	ist die Annahme $E[u] = 0$ stets erfüllt.
c	**X** ist $\hat{y} = 2\hat{\beta}_1 + \hat{\beta}_2$ der vorhergesagte Wert von y an den Stellen $x_1 = 2$ und $x_2 = 1$
d	kann man das R^2 als Maß der Schätzgüte verwenden.
e	c und d.

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

29.	Ein Chow-Test
a	für zwei Gruppen kann nicht auf Basis der Schätzung eines einzigen Regressionsmodells durchgeführt werden.
b	kann durch Verwendung von geeigneten Interaktionstermen durchgeführt werden.
c	prüft, ob Regressionskoeffizienten nach Gruppen unterschiedlich ausfallen.
d	ist nur für Modelle gültig, in denen die abhängige Variable auf einem stetigen Interval definiert ist.
e	**X** b und c.

30.	Für das Modell $lohn = \beta_0 + \beta_1 alter_1 + \beta_2 alter^2 + u$ ergibt sich nach einer KQ-Schätzung für $\hat{\beta}_1 = 0,16$ und für $\hat{\beta}_2 = -0,002$. In welchem Alter wird der durchschnittliche Lohn maximiert?
a	32 .
b	**X** 40 .
c	48 .
d	56 .
e	keine der Antworten.

31.	Die Kovarianz zwischen zwei Zufallsvariablen
a	kann nicht negativ sein.
b	liegt zwischen -1 und 1.
c	ist ein Maß nicht-linearer Zusammenhänge.
d	nimmt für unabhängige Zufallsvariablen den Wert 1 an.
e	**X** keine der Antworten.

32.	Die t-Verteilung
a	hat mit der Beobachtungszahl variierende Freiheitsgrade.
b	ist symmetrisch um den Erwartungswert 0.
c	konvergiert bei wachsenden Freiheitsgraden gegen die Normalverteilung.
d	konvergiert bei wachsenden Freiheitsgraden gegen die F-Verteilung.
e	**X** a, b und c

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

34.	In einem einseitigen t-Test auf dem 5% Signifikanzniveau mit $H_1 : \beta_1 < 0$
a	können keine Typ II Fehler gemacht werden.
b	liegt der Ablehnungsbereich auf der rechten Seite der t-Verteilung.
c	wird die H_0 verworfen, wenn $p > 0,05$.
d	wird die H_0 verworfen, wenn der empirische t-Wert größer als 0 ist.
e	**X** keine der Antworten.

35.	Die Stichprobenvarianz
a	ist die Summe der quadrierten Abweichungen einer Variable von ihrem Mittelwert.
b	ist kein Dispersionsmaß.
c	ist ein erwartungstreuer Schätzer für das R^2.
d	nimmt nie den Wert Null an.
e	**X** keine der genannten Antworten

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

37.	Für eine unverzerrte Schätzung des wahren Bevölkerungsparameters mittels KQ ist nicht erforderlich,
a	dass der Erwartungswert der Konstante null ist.
b	dass die Störterme normalverteilt sind.
c	dass der auf x bedingte Erwartungswert des Störterms null ist.
d	dass die Stichprobengröße gegen unendlich konvergiert.
e	**X** a, b und d

38.	Wenn das Bestimmtheitsmaß R^2 den Wert 0,55 annimmt,
a	dann wird 55% der Variation in den erklärenden Variablen erklärt.
b	dann ist das angepasste Bestimmtheitsmaß $\overline{R}^2 < 0,55$
c	dann wird 55% der Variation der abhängigen Variable erklärt.
d	dann ist die KQ-Schätzung effizient.
e	**X** b und c

39.	Unter den Gauss-Markov Annahmen
a	gilt die BLUE-Eigenschaft des KQ-Schätzers.
b	ist der KQ-Schätzer der beste, lineare, unverzerrte Schätzer.
c	folgt der KQ-Schätzer der Normalverteilung.
d	ist der KQ-Schätzer konsistent.
e	**X** a, b und d.

40.	Typ I-Fehlerwahrscheinlichkeiten
a	**X** werden durch p-Werte beschrieben.
b	sind für t- aber nicht für F-Tests relevant.
c	unterscheiden sich je nachdem ob ein- oder zweiseitig getestet wird.
d	beschreiben die Wahrscheinlichkeit, die Nullhypothese zu verwerfen.
e	c und d.