Bachelorprüfung SS 2016

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen. **Es wird nur der Lösungsbogen eingesammelt.** Angaben auf dem Aufgabenzettel werden nicht gewertet.
Bewertung:	Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
Erlaubte Hilfsmittel:	• Formelsammlung (ist der Klausur beigefügt) • Tabellen der statistischen Verteilungen (sind der Klausur beigefügt) • Taschenrechner • Fremdwörterbuch
Wichtige Hinweise:	• Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert. • Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1: **[16,5 Punkte]**

Sie interessieren sich für die Determinanten der Teilnahme an Public Viewing Veranstaltungen bei der Fußball-Europameisterschaft 2016. Ihr Datensatz enthält für 890 Individuen folgende Informationen:

$Public\_Viewing_i$	Anzahl der Teilnahmen an Public Viewing Veranstaltungen der EM 2016 von Person *i*
$Alter_i$	Alter von Person *i* in Jahren
WM_i	Anteil der WM-Spiele 2014, bei welchen Person *i* an Public Viewing Veranstaltungen teilnahm, gemessen in Prozent (0-100)
FB_1_i	Binäre Variable: =1, wenn Person *i* höchstens 100 Facebook-Freunde hat, =0 sonst
FB_2_i	Binäre Variable: =1, wenn Person *i* mindestens 101 und höchstens 300 Facebook-Freunde hat, =0 sonst
FB_3_i	Binäre Variable: =1, wenn Person *i* mindestens 301 Facebook-Freunde hat, =0 sonst

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Public\_Viewing_i = \beta_0 + \beta_1 Alter_i + \beta_2 WM_i + \beta_3 FB\_2_i + \beta_4 FB\_3_i + u_i$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	3,676	1,488	2,47	0,014
$Alter$	-0,054	0,006	-9,261	0,000
WM	0,081	0,009	8,58	0,000
FB_2	0,388	0,041	9,463	0,000
FB_3	1,014	1,060	0,957	?????

a. Abhängige Variable: *Public_Viewing*

Hinweis: Das R^2 des Modells beträgt 0,401.

a) Interpretieren Sie den geschätzten Koeffizienten von FB_3 inhaltlich. Ist der geschätzte Koeffizient signifikant? Begründen Sie. (2 Punkte)

b) Sie möchten testen, ob die Variablen FB_2 und FB_3 gemeinsam signifikant zum Erklärungsgehalt des Modells beitragen. Sie schätzen das Modell erneut ohne die beiden Variablen und erhalten ein R^2 von 0,375. Führen Sie einen entsprechenden Test am 5% Signifikanzniveau durch. Geben Sie Null- und Alternativhypothese, Freiheitsgrade, Teststatistik, kritischen Wert und Testentscheidung an. (5 Punkte)

c) 2014 gab es 64 WM-Spiele. Anton hat bei der WM damals 14 Spiele bei einer Public Viewing Verantaltung gesehen, Susanne hat nur 3 Spiele bei einer Public Viewing Veranstaltung verfolgt. Sowohl Anton als auch Susanne sind 35 Jahre alt und haben jeweils 200 Facebook-Freunde. Nutzen Sie obigen Output um zu berechnen, wie viele EM-Spiele 2016 Anton im Vergleich zu Susanne bei einer Public Viewing Veranstaltung mehr/weniger gesehen hat. (2,5 Punkte)

d) Bestimmen Sie das 99%-Konfidenzintervall für β_1 und interpretieren Sie Ihr Ergebnis. Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (3 Punkte)

e) Welchen Wert würde $\hat{\beta}_1$ annehmen, wenn das Alter der Individuen nicht in Jahren, sondern in Quartalen gemessen worden wäre? *Hinweis:* Ein Quartal entspricht 3 Monaten. (1,5 Punkte)

f) Berechnen Sie für eine 27-jährige Person, welche 192 Facebook-Freunde hat und bei der WM 2014 ein Viertel aller Spiele bei einer Public Viewing Veranstaltung gesehen hat, die vorhergesagte Anzahl der Teilnahmen an Public Viewing Veranstaltungen während der EM 2016. (2,5 Punkte)

Aufgabe 2: **[15,5 Punkte]**

Sie interessieren sich für die Determinanten der PEWI-Endnoten im Sommersemester 2016. Es liegt ein Datensatz mit Informationen für 600 Studierende vor:

$Note_i$	Endnote im Fach PEWI, gemessen von 1=sehr gut bis 5=mangelhaft, von Student i
$Lernzeit_i$	Wöchentlicher Lernaufwand für PEWI in Stunden von Student i
$Vorlesung_i$	Anzahl im Semester besuchter Vorlesungen von Student i
$Buch_i$	Anzahl der von Student i gelesenen Lehrbuchkapitel
$Hausarbeit_i$	Binäre Variable: =1, wenn Student i die Hausarbeit geschrieben hat, =0 sonst

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Note_i = \beta_0 + \beta_1 log(Lernzeit_i) + \beta_2 Hausarbeit_i + \beta_3 Buch_i + \beta_4 Vorlesung_i + \beta_5 (Vorlesung_i \cdot Buch_i) + u_i$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$Konstante$	4,390	1,475	2,976	0,003
$log(Lernzeit)$	-12,800	4,100	-3,122	0,002
$Hausarbeit$	-0,417	0,093	-4,484	0,000
$Buch$	-0,089	0,032	-2,781	0,005
$Vorlesung$	-0,094	0,039	-2,413	0,016
$Vorlesung \cdot Buch$	-0,012	0,006	-2,063	0,040

a. Abhängige Variable: $Note$

a) Interpretieren Sie den geschätzten Koeffizienten $\hat{\beta}_1$ inhaltlich. (1 Punkt)

b) Eine KQ-Schätzung des obigen Modells ergibt $R^2 = 0,375$.

i) Interpretieren Sie diesen Wert. (1 Punkt)

ii) Berechnen Sie das angepasste Bestimmtheitsmaß des obigen Modells. Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (2 Punkte)

c) Bestimmen Sie die erwartete Notenänderung für eine zusätzlich besuchte Vorlesung für Studierende, die 4 Kapitel im Lehrbuch gelesen haben und interpretieren Sie das Ergebnis. (3 Punkte)

d) Sie vermuten, dass zusätzlich gelesene Kapitel im Lehrbuch zu einer stärkeren Verbesserung der Endnote führen, wenn mehr Vorlesungen besucht werden. Hat sich Ihre Vermutung bestätigt? Begründen Sie Ihre Antwort. (1,5 Punkte)

e) Sie vermuten, dass eine geschriebene Hausarbeit die Endnote um mehr als 0,3 senkt. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das

5%-Signifikanzniveau an. Wird Ihre Vermutung bestätigt? Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (4,5 Punkte)

f) Sie vermuten, dass es in den Regressionsparametern des vorliegenden Modells Unterschiede zwischen Frauen und Männern gibt. Geben Sie an, wie Sie mit einer einzigen Schätzung diese Vermutung testen können. Notieren Sie die Schätzgleichung und geben Sie die Hypothesen an. (2,5 Punkte)

Aufgabe 3: **[18 Punkte]**

Sie führen eine klinische Studie durch, bei der Sie untersuchen, wie sich zwei verschiedene Diätprogramme auf den Gewichtsverlust der Probanden auswirken. Sie rekrutieren insgesamt 10 Personen für die Studie, von denen Sie zufällig 5 für Programm A (Gruppe A) und 5 für Programm B (Gruppe B) auswählen. Sie vergleichen den Gewichtsverlust nach 1 Monat.

Ihnen liegen folgende Informationen nach einem Monat vor:

A_i	Binäre Variable: =1, wenn Person i in Gruppe A war, =0 wenn Gruppe B
$Verlust_i$	Gewichtsverlust (in Kilogramm) von Person i nach 1 Monat

A	0	1	1	0	1	0	0	1	0	1
$Verlust$	3	6	5	4	7	3	3	8	5	6

Sie schätzen das folgende lineare Regressionsmodell:

$$Verlust_i = \beta_0 + \beta_1 A_i + u_i$$

Hinweis: $\sum_{i=1}^{10}(\widehat{Verlust}_i - \overline{Verlust})^2 = 19,6$.

a) Berechnen und interpretieren Sie inhaltlich die geschätzten Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$. (4 Punkte)

b) Berechnen Sie die t-Statistik für den geschätzten Koeffizienten $\hat{\beta}_1$. Nehmen Sie an, dass der Standardfehler der Regression bei 1,024 liegt. Runden Sie auf die dritte Nachkommastelle.

Für den Fall, dass Sie bei 3a) keine Lösung errechnet haben, verwenden Sie $\hat{\beta}_1 = 4$ für Ihre Rechnung. (3 Punkte)

c) Berechnen Sie das R^2 der Schätzung. (3 Punkte)

d) Welche Annahme muss erfüllt sein, damit Sie den kausalen Effekt des Diätprogramms auf den Gewichtsverlust schätzen können? Erläutern Sie die Annahme am Beispiel und diskutieren Sie, ob die Annahme in dem vorliegenden Beispiel erfüllt ist. (2 Punkte)

e) Ein Kollege behauptet, dass sich das Alter einer Person auf den Gewichtsverlust auswirkt. Erklären Sie verbal, unter welchen Umständen die Auslassung der Variable Alter weiterhin zu einer unverzerrten Schätzung der Parameter der Schätzung aus 3a) führt. (2 Punkte)

f) Zeigen Sie für das einfache Modell $y_i = \beta_0 + \beta_1 x_i + u_i$, dass der Mittelwert der vorhergesagten Werte der abhängigen Variable dem Mittelwert der beobachteten Werte der abhängigen Variable entspricht. (4 Punkte)

Aufgabe 4 - 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.	Welche Momentbedingung(en) benötigen Sie, um den KQ-Schätzer herleiten zu können?
a	$E[u] \neq 0$.
b	$E[u] = 0$, $E[y] = 0$.
c	$E[u] = 0$, $E[x_j] = 0$ (für alle j=1,2,...,k)
d	$E[u] = 0$, $E[x_j u] = 0$ (für alle j=1,2,...,k)

2.	Unter den Gauss-Markov Annahmen sind die Störterme eines linearen Regressionsmodells
a	log-normalverteilt.
b	normalverteilt.
c	χ^2 verteilt.
d	homoskedastisch.

3.	Sie schätzen das lineare Modell $y_i = \beta_0 + \beta_1 x_i + u_i$. Ein Kollege behauptet, dass $cov(x,u) \neq 0$. Was ist die Konsequenz für eine KQ-Schätzung?
a	die Störterme sind heteroskedastisch.
b	das R-Quadrat wird negativ.
c	der KQ-Schätzer kann nicht bestimmt werden.
d	die Parameter werden verzerrt geschätzt.

4.	Sie möchten für das lineare Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ testen, ob $cov(x,u) \neq 0$. Zu diesem Zweck regressieren Sie die Residuen auf die erklärende Variable: $\hat{u}_i = \gamma_0 + \gamma_1 x_i + \varepsilon_i$. Welche Aussage trifft zu?
a	γ_0 ist signifikant von 0 verschieden.
b	Der Koeffizient γ_1 ist 0.
c	Das R^2 der Schätzung ist größer als 0.
d	Es liegt keine Variation in $\hat{u}$ vor.

5.	Sie möchten den kausalen Effekt von einer Variable x auf die Variable y schätzen. In welcher Situation besteht das Problem von Over-Controlling?
a	Wenn das Modell zu viele irrelevante Variablen enthält.
b	Wenn der funktionale Zusammenhang zwischen x und y fehlspezifiert ist.
c	Wenn das Modell zusätzliche Variablen enthält, die von x bestimmt werden.
d	Wenn das Modell weitere relevante, mit x unkorrelierte, erklärende Variablen enthält.

6.	Sind zwei Zufallsvariablen X und Y unabhängig, so entspricht die gemeinsame Wahrscheinlichkeitsfunktion von X und Y
a	der Differenz der Randverteilung von X und der Randverteilung von Y.
b	dem Produkt der Randverteilung von X und der Randverteilung von Y.
c	dem Quotienten aus der Randverteilung von X und der Randverteilung von Y.
d	der Summe der Randverteilung von X und der Randverteilung von Y.

7.	Im linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$
a	ist β_1 positiv, wenn $corr(x,y) > 0$.
b	ist β_1 positiv, wenn $Var(x) > 0$.
c	ist β_1 positiv, wenn $corr(\beta_0, \beta_1) > 0$.
d	ist β_1 positiv, wenn $\sum(x_i - \bar{x})^2 > 0$.

8.	Die t-Verteilung
a	ist symmetrisch mit Erwartungswert 1.
b	konvergiert bei steigenden Freiheitsgraden gegen die F-Verteilung.
c	ergibt quadriert eine χ^2- Verteilung.
d	variiert mit der Zahl der Freiheitsgrade.

9.	Für ein lineares Wahrscheinlichkeitsmodell gilt, dass
a	die abhängige Variable stetig verteilt ist.
b	die Störterme heteroskedastisch sind.
c	vorhergesagte Wahrscheinlichkeiten innerhalb des Intervalls (0,1) liegen.
d	der KQ-Schätzer nur linear erklärende Variablen berücksichtigt.

10.	Im linearen Wahrscheinlichkeitsmodell
a	hat das R^2 die gleiche inhaltliche Interpretation wie im linearen Regressionsmodell.
b	gibt die Prognose $1-\hat{y}$ die Wahrscheinlichkeit für das Ereignis $y=1$ an.
c	werden binär kodierte Variablen nicht als erklärende Variablen verwendet.
d	können Interaktionseffekte nicht geschätzt werden.

11.	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	das Modell mit Konstante geschätzt wird.
b	u normalverteilt ist.
c	der Wahrscheinlichkeitsgrenzwert von $\hat{\beta}_1$ dem wahren Wert β_1 entspricht.
d	die Stichprobe unendlich groß wird.

12.	Die Eigenschaft der Konsistenz
a	ist nicht für kleine Stichproben definiert.
b	kann nur für unverzerrte Schätzer nachgewiesen werden.
c	muss gelten, um die geschätzten Koeffizienten inhaltlich interpretieren zu können.
d	kann nicht für lineare Wahrscheinlichkeitsmodelle nachgewiesen werden.

13.	Eine KQ-Schätzung liefert $\hat{y}_i = -5 - 2,5x_i$. Welchen Wert hat das Residuum für die Beobachtung $(y_1, x_1) = (10,4)$?
a	-5.
b	5.
c	15.
d	25.

14.	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	H_0 verworfen wird, obwohl sie zutrifft.
d	H_0 nicht verworfen wird, obwohl sie nicht zutrifft.

15.	Das Auslassen der Variable z_i aus dem Modell $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$ führt zur positiv verzerrten Schätzung von β_1, wenn gilt:
a	$cov(x,z) < 0$ und $\beta_2 > 0$.
b	$cov(x,z) > 0$ und $\beta_2 < 0$.
c	$cov(x,z) < 0$ und $\beta_2 < 0$.
d	$cov(x,z) = 0$, und $\beta_2 < 0$.

16.	Gegeben ist die Schätzgleichung: $\widehat{Lohn}_i = 400 + 200\widehat{Bildung}_i + 300\widehat{Erfahrung}_i - 5\widehat{Erfahrung}_i^2$. Bei welcher Berufserfahrung wird der erwartete Lohn maximiert?
a	15.
b	20.
c	30.
d	60.

17.	F-Tests
a	können verwendet werden, um Hypothesen bezüglich einer Restriktion zu testen.
b	können nicht verwendet werden, um Hypothesen bezüglich mehrerer Restriktionen zu testen.
c	können bei mehr als 10 Zählerfreiheitsgraden nicht durchgeführt werden.
d	können bei mehr als 120 Nennerfreiheitsgraden nicht durchgeführt werden.

18.	Beim F-Test
a	gilt bei Hypothesentests bezüglich einzelner Parameter $F^2_{1,n-k-1} = t_{n-k-1}$.
b	ist die Teststatistik nie negativ.
c	führen Hypothesentests bei zweiseitigen Alternativen zu anderen Ergebnissen als der t-Test.
d	wird die Nullhypothese verworfen, wenn F < c.

19.	Mit einem F-Test vergleichen Sie zwei Modelle (M1 und M2) mit der gleichen abhängigen Variable. M1 enthält neben der Konstante ausschließlich die erklärenden Variablen Arbeitserfahrung und Bildung. M2 berücksichtigt nur die Konstante. Welche Aussage trifft zu?
a	M1 ist das restringierte Modell.
b	M1 enthält zwei Parameter.
c	Die Anzahl der Zählerfreiheitsgrade des F-Tests ist $q = 2$.
d	Das restringierte Modell berücksichtigt einen Steigungsparameter.

20.	Wenn die Störterme einer multiplen Regression nicht normalverteilt sind, dann
a	können keine Hypothesentests durchgeführt werden.
b	sind die Verteilungsannahmen für die Teststatistiken der t- und F-Tests richtig.
c	ist der KQ-Schätzer inkonsistent.
d	können t- und F-Tests für kleine Stichproben das falsche Ergebnis liefern.

21.	Wird das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ ohne Regressionskonstante geschätzt, und gilt dass $\beta_0 \neq 0$, dann
a	nimmt $\widehat{y_i}$ für $x_{1i} = 0, x_{2i} = 0$ den Wert eins an.
b	ist gewährleistet, dass der Mittelwert der Residuen 0 ist.
c	wird β_1 verzerrt geschätzt.
d	wird β_0 implizit gleich 1 gesetzt.

22.	Um die Präzision einer Schätzung zu erhöhen, sollte man
a	die Stichprobengröße verringern.
b	erklärende Variablen mit hoher Variation verwenden.
c	das Modell ohne Konstante schätzen.
d	untereinander stark korrelierte erklärenden Variablen ins Modell aufnehmen.

23.	Sie führen nacheinander einen rechtsseitigen und einen linksseitigen Signifikanztest als t-Test durch. Wie unterscheiden sich die Werte der t-Statistik, wenn beide Tests für das gleiche Modell, die gleiche Stichprobe, den gleichen Koeffizienten, die gleiche Restriktion und das gleiche Signifikanzniveau durchgeführt werden?
a	Der Wert beim linksseitigen Tests ist größer.
b	Der Wert ist in beiden Tests gleich groß.
c	Die Antwort hängt von dem Vorzeichen des Koeffizienten ab.
d	Der Wert beim rechtsseitigen Tests ist größer.

24.	Ein Konfidenzintervall wird
a	breiter, wenn man das Signifikanzniveau von 5% auf 10% erhöht.
b	mit steigendem Standardfehler $se(\beta_j)$ breiter.
c	mit steigendem Schätzwert $\hat{\beta}_j$ breiter.
d	mit steigender Stichprobengröße breiter.

25.	Der Standardfehler von $\hat{\beta}_j$
a	ist keine Zufallsvariable.
b	ist ein Schätzer für die Varianz von $\hat{\beta}_j$.
c	steigt bei einer präziseren Schätzung.
d	steigt mit steigender Residuenquadratsumme (SSR).

26.	Leitet man den Kleinstquadrate-Schätzer für das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ her, ergeben sich höchstens
a	eine Bedingung erster Ordnung.
b	zwei Bedingungen erster Ordnung.
c	drei Bedingungen erster Ordnung.
d	vier Bedingungen erster Ordnung.

27.	Gegeben ist folgendes Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$. Die aufgestellten Hypothesen lauten $H_0 : \beta_3 = -2$ vs. $H_1 : \beta_3 \neq -2$. Auf einem Signifikanzniveau von 1% führt eine Teststatistik von $-2,78$
a	bei $n = 24$ zur Ablehnung von H_0.
b	bei $n = 25$ zur Ablehnung von H_0.
c	bei $n = 26$ zur Ablehnung von H_0.
d	bei $n = 27$ nicht zur Ablehnung von H_0.

28.	In welchem Fall weisen die geschätzten Koeffizienten des Modells $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ auf einen konvexen Zusammenhang zwischen $\hat{y}_i$ und x_i hin?
a	$\hat{\beta}_0 > 0$ und $\hat{\beta}_2 < 0$.
b	$\hat{\beta}_1 > 0$ und $\hat{\beta}_2 < 0$.
c	$\hat{\beta}_1 < 0$ und $\hat{\beta}_2 > 0$.
d	$\hat{\beta}_1 < 0$ und $\hat{\beta}_2 < 0$.

29.	Wenn Sie in einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ die Variable x_i durch $\frac{1}{5} \cdot x_i$ ersetzen,
a	verringert sich der Wert der geschätzten Konstante um den Faktor 5.
b	verfünffacht sich der Wert der geschätzten Konstante sowie des geschätzten Steigungsparameters.
c	verringert sich der Wert des geschätzten Steigungsparameters um den Faktor 5.
d	verfünffacht sich der Wert des geschätzten Steigungsparameters.

30.	Was ändert sich bei Umskalierung der abhängigen Variable nicht?
a	Konfidenzintervalle der Parameter.
b	R^2 der Schätzung.
c	Residuen der Beobachtungen.
d	Standardfehler der Parameter.

31.	Die Residuenquadratsumme SSR im einfachen linearen Regressionsmodell
a	sinkt bei Aufnahme eines relevanten Regressors.
b	berechnet sich als Summe der Gesamtvariation (SST) und der durch das Modell erklärten Variation (SSE).
c	berücksichtigt die zur Berechnung benötigten Freiheitsgrade.
d	bleibt bei Umskalierung der abhängigen Variable konstant.

32.	Für das Modell $lohn_i = \beta_0 + \beta_1 alter_i + \beta_2 alter_i^2 + u_i$ ergibt eine KQ-Schätzung $\hat{\beta}_1 = 0,7$ und $\hat{\beta}_2 = 0,04$. Der Stundenlohn $lohn_i$ ist in Euro gemessen. Der marginale Effekt des Alters auf den Stundenlohn ist für eine 40-jährige Person im Vergleich zu einer 24-jährigen Person nach dieser Schätzung im Mittel
a	0,64 Euro pro Stunde höher.
b	1,28 Euro pro Stunde höher.
c	1,34 Euro pro Stunde höher.
d	1,98 Euro pro Stunde höher.

33.	Für das Modell $log(lohn_i) = \beta_0 + \beta_1 frau_i + u_i$ ergibt eine KQ-Schätzung $\hat{\beta}_1 = -0,08$. Die Variable $lohn_i$ beschreibt den Stundenlohn in Euro und $frau_i$ ist eine Indikatorvariable (=1, wenn Person weiblich und $frau_i$=0, wenn Person männlich). Welche Interpretation ist korrekt?
a	Frauen verdienen im Mittel 80 Cent pro Stunde weniger als Männer.
b	Frauen verdienen im Mittel 8 Prozentpunkte weniger als Männer.
c	Männer verdienen im Mittel 8 Prozent mehr als Frauen.
d	Männer verdienen im Mittel 8 Euro pro Stunde mehr als Frauen.

34.	Sie schätzen folgendes Modell mit KQ: $log(y_i) = \beta_0 + \beta_1 log(x_i) + u_i$. $\hat{\beta}_0$ beträgt 0,3 und $\hat{\beta}_1$ beträgt -0,7. Welche Aussage ist richtig?
a	Für eine Beobachtung mit $x_i = 0$ beträgt $\hat{y}_i = 0,3$.
b	Für eine Beobachtung mit $x_i = 0$ beträgt $\hat{y}_i = 30\%$.
c	Sinkt x um 3%, so steigt y im Mittel um 2,1%.
d	Sinkt x um 2%, so steigt y im Mittel um 140%.

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

36.	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	benutzt eine t-verteilte Teststatistik.
d	ist nur für Modelle gültig, in denen die abhängige Variable auf einem stetigen Intervall definiert ist.

37.	Sie beobachten Abiturnoten für Jungen und Mädchen in Ost- und Westdeutschland. Sie schätzen folgendes Modell: $note_i = \beta_0 + \beta_1 junge_i + \beta_2 west_i + \beta_3 junge_i \cdot west_i + u_i$. Was misst der geschätzte Parameter für β_2?
a	Den Mittelwert der Abiturnoten für Westdeutschland.
a	Den Mittelwert der Abiturnoten für Westdeutsche Mädchen.
c	Den Mittelwertunterschied der Abiturnoten zwischen West- und Ostdeutschland.
d	Den Mittelwertunterschied der Abiturnoten zwischen West- und Ostdeutschland für Mädchen.

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

39.	Sie regressieren den Stundenlohn für Personen mit Schulabschluss auf die Indikatorvariablen Abitur 1/0 und Realschulabschluss 1/0, wobei Hauptschulabschluss 1/0 als Referenzkategorie dient. Welches Problem tritt bei der Schätzung auf, wenn in Ihrer Stichprobe keine Personen mit Hauptschulabschluss vorhanden sind?
a	Die Beobachtungszahl N sinkt.
b	Perfekte Multikollinearität.
c	Das R^2 ist kleiner als 0.
d	Problem relevanter ausgelassener Variablen (omitted variable bias).

40.	Gegeben sind die erklärenden Variablen X_1, X_2 und X_3. Welcher der folgenden Ausdrücke stellt einen Interaktionsterm dar?
a	$X_1 \cdot X_2 \cdot X_3$.
b	$X_2 + X_3$.
c	$X_3 - X_1$.
d	$X_1^{X_2}$.

Formelsammlung – Praxis der empirischen Wirtschaftsforschung

Kapitel 1:

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_i(y_i-\bar{y})$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})y_i$$

$$=\sum_{i=1}^{n}x_iy_i-n\bar{x}\cdot\bar{y}$$

$$E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i)$$

$$E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)$$

$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y) + 2ab\,Cov(X,Y)$

Für identisch und unabhängig verteilte Zufallsvariablen Y_i:

$$\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_i\right)^2$$

Kapitel 2:

$y_i = \beta_0 + \beta_1 x_i + u_i$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\sum_{i=1}^{n}\hat{u}_i=0 \qquad \sum_{i=1}^{n}x_i\hat{u}_i=0$$

$$SST\equiv\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$SSE\equiv\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$SSR\equiv\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$$

$$R^2=\frac{SSE}{SST}=1-\frac{SSR}{SST}\ ,\ 0\le R^2\le 1$$

$$E\left(\hat{\beta}_0\right)=\beta_0 \qquad E\left(\hat{\beta}_1\right)=\beta_1$$

$$Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sigma^2}{SST_x}$$

$$Var(\hat{\beta}_0)=\frac{\sigma^2\frac{1}{n}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{\sigma}^2=\frac{1}{(n-2)}\sum_{i=1}^{n}\hat{u}_i^2=\frac{SSR}{(n-2)}$$

Regression durch den Ursprung:

$$\tilde{\beta}_1=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$$

Kapitel 3:

$$R^2=\frac{SSE}{SST}=\frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2\right)}$$

Wenn $y=\beta_0+\beta_1x_1+\beta_2x_2+u$

und $\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$

dann $\tilde{\beta}_1=\hat{\beta}_1+\hat{\beta}_2\tilde{\delta}_1$ mit $x_2=\tilde{\delta}_0+\tilde{\delta}_1x_1$

Allgemein für j = 1, 2, …, k:

$$Var\left(\hat{\beta}_j\right)=\frac{\sigma^2}{SST_j\left(1-R_j^2\right)}$$

$$SST_j=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-k-1}=\frac{SSR}{n-k-1}$$

$$se(\hat{\beta}_j)=\frac{\hat{\sigma}}{\left[SST_j(1-R_j^2)\right]^{\frac{1}{2}}}$$

MLR.1: Modell der Grundgesamtheit
$y = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \ldots + \beta_k x_k + u$

MLR.2: Zufallsstichprobe der Größe n folgt dem Bevölkerungsmodell.

MLR.3: Keine unabhängige Variable ist konstant. Keine perfekte Kollinearität.

MLR.4: $E\left(u\,|\,x_1,x_2,\ldots,x_k\right)=0$

MLR.5: $Var\left(u\,|\,x_1,x_2,\ldots,x_k\right)=\sigma^2$

MLR.6: u ist von x_1, x_2, …, x_k unabhängig und $u \sim \text{Normal}(0, \sigma^2)$.

Kapitel 4:

$$(\hat{\beta}_j - \beta_j) \big/ se(\hat{\beta}_j) \sim t_{n-k-1}$$

$$-t_{\frac{\alpha}{2},n-k-1} \le \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \le t_{\frac{\alpha}{2},n-k-1}$$

$$\hat{\beta}_j - c \cdot se(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + c \cdot se(\hat{\beta}_j)$$

$$F \equiv \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)}$$

$$F = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)}$$

Kapitel 5:

$$\lim_{n\to\infty} P(\left|\hat{\beta}_1 - \beta_1\right| > \varepsilon) \to 0$$

$$\operatorname{plim}\left(\hat{\beta}_1\right) = \beta_1$$

Kapitel 6:

Standardisierung:

$$\frac{y_i - \bar{y}}{\hat{\sigma}_y} = \hat{\beta}_1 \left(\frac{\hat{\sigma}_1}{\hat{\sigma}_y}\right) \left(\frac{x_{i1} - \bar{x}_1}{\hat{\sigma}_1}\right) + \hat{\beta}_2 \left(\frac{\hat{\sigma}_2}{\hat{\sigma}_y}\right) \left(\frac{x_{i2} - \bar{x}_2}{\hat{\sigma}_2}\right)$$
$$+ \ldots + \hat{\beta}_k \left(\frac{\hat{\sigma}_k}{\hat{\sigma}_y}\right) \left(\frac{x_{ik} - \bar{x}_k}{\hat{\sigma}_k}\right) + \frac{\hat{u}_i}{\hat{\sigma}_y}$$

Semielastizität:
$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_j \Delta x_j) - 1]$$

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$$

$$\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

$$P[\hat{y}^0 - t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0) \le y^0 \le \hat{y}^0 + t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0)] = 1\text{-}\alpha$$

$$\widehat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \ldots + \hat{\beta}_k x_k$$
$$E(y|\mathbf{x}) = \exp(\sigma^2/2) \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Kapitel 7:

Regression nach Gruppen
- Modell gepoolt: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$

Chow-Test (mit $SSR_P = SSR_{\text{gepooltes Modell}}$):

$$F = \frac{(SSR_P - (SSR_1 + SSR_2))/(k+1)}{(SSR_1 + SSR_2)/\left(n - 2(k+1)\right)}$$

TABLE G.2

Critical Values of the t Distribution

	Significance Level				
1-Tailed:	.10	.05	.025	.01	.005
2-Tailed:	.20	.10	.05	.02	.01
Degrees of Freedom 1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
90	1.291	1.662	1.987	2.368	2.632
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 df is 2.485. The 5% critical value for a two-tailed test with large (> 120) df is 1.96.

Source: This table was generated using the Stata® function invttail.

TABLE G.3b

5% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator df (∞) is 2.37.

Source: This table was generated using the Stata® function invFtail.