Bachelorprüfung WS 2019/2020

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: **[17 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:

$ln(wage_i)$	Stundenlohn in Euro, logarithmiert
$educ_i$	Anzahl der Ausbildungsjahre
$mental_i$	Index der psychischen Gesundheit, gemessen auf einer Skala von 0 (schlecht) bis 100 (sehr gut)
$vollzeit_i$	=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]

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Regression	881,715	3	293,905	1210,916	0,000[b]
Nicht standardisierte Residuen	3808,652	15692	0,243		
Gesamt	4690,366	15695			

a. Abhängige Variable: ln(wage)

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

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	6,572	0,027	240,631	0,000
educ	0,056	0,002	35,826	0,000
mental	0,006	0,000	13,709	0,000
vollzeit	0,309	0,008	36,933	0,000

a. Abhängige Variable: ln(wage)

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

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

b) Die Variable $wage_i$ wird nun statt in Euro in 1000 Euro gemessen, alles andere bleibt gleich. Wie ändern sich,

   i. die Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_2$? Zeigen Sie Ihren Rechenweg. (4 Punkte)

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

c) Berechnen Sie das 99%-Konfidenzintervall für den Paramter von *vollzeit* ($\hat{\beta}_3$) und interpretieren Sie es. Zeigen Sie Ihren Rechenweg und runden Sie auf die dritte Nachkommastelle. (3,5 Punkte)

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

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

Die Residuenquadratsumme (SSR) des neuen Modells beträgt 3804,260. Testen Sie am 1%-Niveau, ob die Koeffizienten der neuen Variablen gemeinsam statistisch signifikant sind. Geben Sie Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und ihre Testentscheidung an. (4,5 Punkte)

e) Als Nächstes wird die Variable *educ* quadriert und als *educ*2 mit dem Parameter β_{educ2} zusätzlich in das Model aufgenommen. Nehmen Sie an, dass der Schätzer von $\beta_1 > 0$ bleibt. Skizzieren Sie drei möglichen Verläufe des Grafen, der die Beziehung zwischen $ln(wage)$ und *educ* darstellt, wenn a) $\widehat{\beta_{educ2}} > 0$, b) $\widehat{\beta_{educ2}} < 0$ oder c) $\widehat{\beta_{educ2}} = 0$. (2 Punkte)

Aufgabe 2: **[17 Punkte]**

Ihnen liegen Daten zu 159 Superhelden vor, die jemals Teil der Avengers (einer Superheldengruppe aus Comics & Filmen) waren. Ihr Datensatz enthält folgende Informationen:

$death_i$	=1, wenn Superheld/in *i* mindestens einmal gestorben ist; =0 sonst
$female_i$	=1, wenn Superheld/in *i* eine Frau ist; =0 wenn Superheld/in *i* ein Mann ist
$avengersince_i$	Anzahl der Jahre, die Superheld/in *i* bereits Teil der Avengers ist

Sie unterstellen folgendes Regressionsmodell und schätzen dieses mit SPSS:

$$death_i = \beta_0 + \beta_1 female_i + \beta_2 avengersince_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	0,270(a)	0,073	???	0,481

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	0,277	0,065	4,290	0,000
female	-0,027	0,082	-0,331	0,743
avengersince	0,008	0,002	????	????

a. Abhängige Variable: death

Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle.

a) Interpretieren Sie $\hat{\beta}_1$ inhaltlich und statistisch. (2 Punkte)

b) Berechnen Sie das korrigierte R^2 des Modells. (1,5 Punkte)

c) Sie möchten testen, ob der Koeffizient der Variable *avengersince* signifikant ist. Führen Sie einen entsprechenden Test durch. Geben Sie Testverfahren, Null- und Alternativhypothese, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das 1%-Signifikanzniveau an. (4,5 Punkte)

d) Tony Stark ist seit 11 Jahren Mitglied der Avengers und männlich. Wie hoch ist seine vorhergesagte Wahrscheinlichkeit, mindestens einmal gestorben zu sein? (2 Punkte)

e) Zeichnen Sie die Regressionsgerade über die Variable *avengersince* je für Männer und Frauen in ein Koordinatensystem ein. Achten Sie auf die korrekte Achsenbeschriftung. (5 Punkte)

f) Nennen Sie zwei Schwächen des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

Aufgabe 3: **[16 Punkte]**

Sie wollen die Determinanten des Spendenverhaltens untersuchen. Ihnen liegen Daten des sozio-ökonomischen Panels mit 2.486 Beobachtungen vor:

$ln(spende_i)$	Spendenbetrag in Euro, logarithmiert
$female_i$	=1, wenn Person eine Frau ist; =0 wenn Person ein Mann ist
age_i	Alter in Jahren
$age2_i$	quadriertes Alter
$income_i$	monatliches Einkommen in Euro

Sie schätzen das folgende Modell:

$$ln(spende_i) = \beta_0 + \beta_1 female_i + \beta_2 age_i + \beta_3 age2_i + \beta_4 income_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	???	0,236	0,235	0,387

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	1,538	0,050	30,720	0,000
female	0,222	0,017	???	???
age	0,062	0,003	11,010	0,000
age2	-0,001	0,000	-8,220	0,000
income	0,063	0,003	???	???

a. Abhängige Variable: ln(spende)

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

a) Testen Sie die Signifikanz des Modells auf dem 5%-Niveau. Geben Sie die Nullhypothese, die Alternativhypothese, die Teststatistik, den kritischen Wert und die Testentscheidung an. (4,5 Punkte)

b) Berechnen und interpretieren Sie den präzisen Effekt von *female* auf das Spendenverhalten. Ist der Effekt statistisch signifikant? Begründen Sie ihre Antwort. (3 Punkte)

c) Bei welchen Alter erreicht der Spendenbetrag sein Maximum? (3 Punkte)

d) Um den Unterschied zwischen Männern und Frauen genauer zu untersuchen, schlägt ihr Kommilitone vor, zusätzlich eine Dummy-Variable für Männer ($male = 1$, wenn Mann; 0 sonst) in das Modell aufzunehmen. Welches Problem entsteht, wenn Sie das umsetzen und wie kann man es lösen? Erläutern Sie ihre Antwort. (2 Punkte)

e) Beschreiben Sie allgemein, welche Bedingungen zutreffen müssen, damit ein Problem ausgelassener Variablen vorliegt. Welche Folgen hätte dies für den geschätzten Koeffizienten? Diskutieren Sie, ob es sich bei $\hat{\beta}_4$ um den kausalen Effekt des Einkommens auf die Spendensumme handelt. (3,5 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.	Schätzt man ein einfaches lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$, so
a	ist das korrigierte R^2 stets positiv.
b	liegt der Punkt $(\bar{y}, \bar{x})$ auf der geschätzten Regressionsgerade.
c	ist die Summe der Residuen ungleich 0.
d	werden die Abstände der Datenpunkte zur geschätzten Regressionsgerade minimiert.

2.	In einem einfachen linearen Regressionsmodell gibt der Steigungsparameter die Semi-Elastizität an, wenn
a	sowohl abhängige als auch unabhängige Variable logarithmiert sind.
b	lediglich die unabhängige Variable logarithmiert ist.
c	lediglich die abhängige Variable logarithmiert ist.
d	weder abhängige, noch unabhängige Variable logarithmiert sind.

3.	Sie haben Daten über Löhne von Männern und Frauen aus Ost- und Westdeutschland. Unterstellen Sie, dass *Frau*, *Mann*, *Ost* und *West* Indikatorvariablen sind. Bei welcher Modellspezifikation kommt es zum Dummy-Variable-Trap?
a	$Lohn_i = \beta_0 + \beta_1(Mann_i \cdot Ost_i) + \beta_2(Frau_i \cdot Ost_i) + \beta_3(Mann_i \cdot West_i) + u_i$
b	$Lohn_i = \beta_0 + \beta_1 Mann_i + \beta_2(Mann_i \cdot Ost_i) + \beta_3(Frau_i \cdot Ost_i) + u_i$
c	$Lohn_i = \beta_1 Mann_i + \beta_2 Frau_i + \beta_3(Mann_i \cdot Ost_i) + \beta_4(Mann_i \cdot West_i) + u_i$
d	Keine der Antworten ist korrekt.

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

5.	Wenn Sie in einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ die Variable x_i durch $\frac{10}{5} \cdot x_i$ ersetzen,
a	verzehnfacht sich der Wert der geschätzten Konstante.
b	verringert sich der Wert des geschätzten Steigungsparameters um den Faktor 2.
c	verdoppelt sich der Wert des geschätzten Steigungsparameters.
d	verringert sich der Wert des geschätzten Steigungsparameters um den Faktor 5.

6.	Im Modell $log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 log(x_{2i}) + u_i$
a	wird β_2 als Elastizität interpretiert.
b	wird β_1 als Elastizität interpretiert.
c	wird β_2 als Semi-Elastizität interpretiert.
d	werden sowohl β_1, als auch β_2 als Semi-Elastizitäten interpretiert.

7.	Eine KQ-Schätzung liefert $\hat{y}_i = 2,8 - 1,5x_{1i} + 2,1x_{2i}$. Welchen Wert hat das Residuum für die Beobachtung $(y_1; x_{11}; x_{21}) = (14; 0; 3)$?
a	0,1.
b	4,9.
c	-2,7.
d	-1,8.

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

9.	In einer quadratischen Spezifikation $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i} + \beta_4 x_{3i} + u_i$ ergibt sich eine u-förmige Beziehung zwischen x_1 und y, wenn
a	$\beta_1 > 0$ und $\beta_2 < 0$.
b	$\beta_1 < 0$ und $\beta_2 = 0$.
c	$\beta_1 > 0$ und $\beta_2 > 0$.
d	$\beta_1 < 0$ und $\beta_2 > 0$.

10.	Die F-Verteilung
a	konvergiert gegen Null.
b	ist nicht symmetrisch.
c	kann negative Werte annehmen.
d	entspricht asymptotisch der t-Verteilung.

11.	Sie beobachten den Stundenlohn für Männer und Frauen in Nord- und Südeuropa. Sie schätzen folgendes Modell: $Lohn_i = \beta_0 + \beta_1 Frau_i + \beta_2 Nord_i + \beta_3 Frau_i \cdot Nord_i + u_i$. Was misst der geschätzte Parameter für β_3?
a	Den Mittelwert des Stundenlohns für Männer aus Nordeuropa.
b	Den Mittelwertunterschied des Stundenlohns zwischen Männern und Frauen aus Südeuropa.
c	Den Mittelwert des Stundenlohns für Frauen.
d	Den Mittelwertunterschied des Stundenlohns zwischen Frauen aus Südeuropa und Frauen aus Nordeuropa.

12.	Der marginale Effekt von x_1 im Modell $y_i = \beta_0 + \beta_1 log(x_{1i}) + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i$ lautet
a	$\beta_1 \frac{1}{x_1} + \beta_3 x_2$.
b	$\beta_1 x_1 + \beta_3 x_2$.
c	$\beta_1 + \beta_3 x_2$.
d	$\beta_1 \frac{1}{x_1} + \beta_3 \frac{1}{x_1} x_2$.

13.	Eine Typ-1-Fehlerwahrscheinlichkeit beschreibt die Wahrscheinlichkeit,
a	eine zutreffende Nullhypothese zu verwerfen.
b	eine zutreffende Nullhypothese nicht zu verwerfen.
c	eine falsche Nullhypothese zu verwerfen.
d	eine falsche Nullhypothese nicht zu verwerfen.

14.	Wie viele Dummyvariablen müssen in das Modell aufgenommen werden, um die kategoriale Variable Schulabschluss (1=Kein Schulabschluss, 2=Hauptschulabschluss, 3=Realschulabschluss, 4=Abitur) in einem Modell mit Konstante vollständig abzubilden?
a	1.
b	2.
c	3.
d	4.

15.	Sie führen nacheinander einen rechtsseitigen und einen beidseitigen t-Test durch. Wie unterscheiden sich die kritischen Werte, wenn beide Tests für das gleiche Modell, die gleiche Stichprobe und das gleiche Signifikanzniveau durchgeführt werden?
a	Der kritische Wert des einseitigen Tests ist größer.
b	Der Absolutbetrag der kritischen Werte des beidseitigen Tests ist größer.
c	Der kritische Wert ist in beiden Tests gleich groß.
d	Die Antwort hängt von dem Vorzeichen des Koeffizienten ab.

16.	Die Aufnahme irrelevanter Variablen in ein Regressionsmodell
a	erhöht die Konsistenz der Schätzung.
b	verringert die Konfidenzintervalle um den Parameterschätzer.
c	verringert den Wert des angepassten Bestimmtheitsmaßes.
d	erhöht die Effizienz der Schätzung.

17.	Das R^2 beschreibt
a	den Anteil der unerklärten Variation an der erklärten Variation.
b	den Anteil der erklärten Variation an der unerklärten Variation.
c	den Anteil der erklärenden Variation an der unerklärten Variation.
d	den Anteil der erklärten Variation an der Gesamtvariation.

18.	Je größer das Signifikanzniveau α,
a	desto enger das Konfidenzintervall.
b	desto breiter das Konfidenzintervall.
c	desto höher das Konfidenzintervall.
d	desto flacher das Konfidenzintervall.

19.	Bei konsistenten Schätzverfahren
a	steigt die Varianz der Schätzung, je größer die Stichprobe.
b	weicht der Schätzer vom wahren Wert stärker ab, je größer die Stichprobe.
c	sinkt die Varianz des Schätzers, je größer die Stichprobe.
d	erhält man immer den wahren Wert des Schätzers.

20.	Wird die Nullhypothese eines Strukturbruchtests abgelehnt, so
a	liegt Autokorrelation vor.
b	liegt perfekte Multikollinearität vor.
c	liegen signifikante Unterschiede in den Steigungsparametern verschiedener Gruppen vor.
d	liegt Konsistenz vor.

21.	Sie regressieren in einem einfachen linearen Modell den Lohn pro Monat auf eine Konstante und die Arbeitszeit pro Monat $Lohn_i = \beta_0 + \beta_1 Arbeitszeit_i + \varepsilon$. Welche Schätzergebnisse würden sich ändern, wenn sowohl der Lohn als auch die Arbeitszeit pro Tag gemessen wird?
a	$\hat{\beta}_0$.
b	$\hat{\beta}_1$.
c	die Schätzergebnisse bleiben unverändert.
d	$\hat{\beta}_0$ und $\hat{\beta}_1$.

22.	Unter den Annahmen MLR.1-5 ist KQ
a	unter allen konsistenten und inferenten Schätzverfahren das einfachste Schätzverfahren.
b	unter allen linearen und unverzerrten Schätzverfahren das effizienteste Schätzverfahren.
c	unter allen deterministischen und stetigen Schätzverfahren das effizienteste Schätzverfahren.
d	unter allen variablen und konsistenten Schätzverfahren das praktischste Schätzverfahren.

23.	Ein nicht-linearer Zusammenhang zwischen zwei Variablen
a	führt zu Heteroskedastie.
b	führt zu einem Strukturbruch.
c	kann mittels KQ geschätzt werden.
d	kann nicht konsistent geschätzt werden.

24.	Wenn die Störterme einer multiplen Regression asymptotisch normalverteilt sind, dann
a	ist der KQ-Schätzer inkonsistent.
b	gilt das Gauss-Markov Theorem.
c	liegt Heteroskedastie vor.
d	sind F-Tests für große Stichproben approximativ gültig.

25.	Das Problem perfekter Multikollinearität
a	führt zu ineffizienten Schätzergebnissen.
b	kann durch Vergrößerung der Stichprobe gelöst werden.
c	kann nicht gelöst werden.
d	keine der Aussagen ist richtig.

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

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

28.	Sie möchten den kausalen Effekt von einer Variable x auf die Variable y schätzen. In welcher Situation besteht das Problem von Overcontrolling?
a	Wenn das Modell nichtlinear ist.
b	Wenn das Modell zusätzliche Variablen enthält, die von x bestimmt werden.
c	Wenn das Modell weitere relevante, mit x unkorrelierte, erklärende Variablen enthält.
d	Wenn das Modell zu viele irrelevante Variablen enthält.

29.	In einem linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ ist
a	β_1 positiv, wenn $Var(y) > 0$.
b	β_1 positiv, wenn $Corr(\beta_0, \beta_1) > 0$.
c	β_1 positiv, wenn $Cov(y,x) < 0$.
d	β_1 positiv, wenn $Corr(y,x) > 0$.

30.	Sie stellen folgendes Modell auf: $wage_i = \beta_0 + \beta_1 female_i + \beta_2 age_i + \beta_3 {age^2}_i + \varepsilon_i$. Welche Aussage ist korrekt?
a	Das Modell kann in der vorliegenden Form nicht geschätzt werden.
b	Der marginale Effekt von *age* auf *wage* ist $2\beta_3 \cdot age + \beta_2$.
c	β_0 bildet den Mittelwert von *wage* in der Stichprobe für Frauen ab.
d	Die Variable *age* korreliert nicht mit *female*, weil sie auch in quadratischer Form vorliegt.

31.	Ein Signifikanztest für den Koeffizientenschätzer $\hat{\beta}$ ergibt einen p-Wert von 0,005. Welche Aussage ist korrekt?
a	Das größte Signifikanzniveau, an dem $H_0 : \beta = 0$ abgelehnt werden kann, ist 0,10.
b	Die Nullhypothese $H_0 : \beta = 0$ kann sowohl am 5%-Signifikanzniveau, als auch am 1%-Signifikanzniveau abgelehnt werden.
c	Bei dem geschätzten Koeffizienten handelt es sich um einen kausalen Effekt.
d	$\hat{\beta}$ ist positiv.

32.	Gegeben sei folgendes Modell: $zufriedenheit_i = \beta_0 + \beta_1 kind_i + u_i$. Sie gehen davon aus, dass es eine (unbeobachtete) Variable $verheiratet_i$ gibt, die positiv mit *kind* und positiv mit *zufriedenheit* korreliert. Welche Aussage ist dann richtig? Im vorliegenden Modell
a	wird β_1 unverzerrt geschätzt.
b	wird β_1 überschätzt.
c	kann β_1 nicht geschätzt werden.
d	wird β_1 unterschätzt.

33.	Ein R^2 von 0,23 bedeutet, dass
a	kein Problem ausgelassener Variablen vorliegt.
b	23% der Variation in der abhängigen Variable durch die erklärenden Variablen erklärt werden.
c	die abhängige Variable keine binäre Variable ist.
d	Homoskedastie vorliegt.

34.	Sie schätzen das Modell $log(einkommen_i) = \beta_0 + \beta_1 female_i + \beta_2 arbeitszeit_i + u_i$. Welcher Koeffizient gibt eine Elastizität an?
a	$\hat{\beta}_0$.
b	$\hat{\beta}_1$.
c	$\hat{\beta}_2$.
d	Keiner der Koeffizienten.

35.	Ein Konfidenzintervall wird enger
a	mit größerer Konstante $\hat{\beta}_0$.
b	mit sinkendem Standardfehler $se(\beta_j)$.
c	mit steigendem Schätzwert $\hat{\beta}_j$.
d	mit sinkender Stichprobengröße.

36.	Was bleibt gleich bei einer Umskalierung einer erklärenden Variable?
a	Das Konfidenzintervall des entsprechenden Schätzkoeffizienten.
b	Das R^2 der Schätzung.
c	Der Schätzkoeffizient der Variable.
d	Der Standardfehler des Schätzkoeffizienten.

37.	Logarithmieren der abhängigen Variable
a	ergibt fehlende Werte für Beobachtungen mit $y = 0$.
b	erhöht die Streuung der abhängigen Variable.
c	steigert den Einfluss der Ausreißerbeobachtungen auf die Schätzergebnisse.
d	erhöht die Streuung der erklärenden Variablen.

38.	Ein konsistenter Schätzer
a	führt gleichzeitig auch zu Homoskedastie.
b	kann nicht unverzerrt sein.
c	erfüllt notwendigerweise die Annahmen MLR.1-MLR.4.
d	erfüllt $plim\ \hat{\beta} = \beta$.

39.	Gilt $E[u\|x] = 0$ für eine Gleichung $y = \beta_0 + \beta_1 x + u$, dann
a	liegt kein Problem ausgelassener Variablen vor.
b	ist der KQ-Schätzer verzerrt.
c	liegt Homoskedastie vor.
d	muss die Konstante den Wert 1 annehmen.

40.	Sie schätzen das Modell $erwerbstaetig_i = \beta_0 + \beta_1 frau_i + \beta_2 alter_i + u_i$, wobei *erwerbstaetig* den Wert 1 annimmt, wenn Person *i* erwerbstätig ist und den Wert 0, wenn Person *i* nicht erwerbstätig ist. Welche Aussage ist richtig?
a	Es liegt zwangsläufig Homoskedastie vor.
b	Das Modell kann in der vorliegenden Form nicht geschätzt werden.
c	β_2 kann nicht sinnvoll interpretiert werde.
d	β_1 ist der Unterschied in der Erwerbstätigenquote zwischen Männern und Frauen.

Formelsammlung – Praxis der empirischen Wirtschaftsforschung

Kapitel 1:

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

Regression durch den Ursprung:

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

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 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

und $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$

dann $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\,\tilde{\delta}_1$ mit $x_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$

Allgemein für j = 1, 2, …, k:

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

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

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

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

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: $E\left(u \mid x_1, x_2, ..., x_k\right) = 0$

MLR.5: $\mathrm{Var}\left(u \mid x_1, x_2, ..., x_k\right) = \sigma^2$

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:

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

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

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.

TABLE G.3c

1% 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	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.

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