Bachelorprüfung SS 2017 - 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.** 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: **[14 Punkte]**

Sie interessieren sich für die Wohnortwahl von sechs neuen Studierenden am Fachbereich Wirtschaftswissenschaften der FAU. Folgende Daten stehen Ihnen zur Verfügung:

$Nuernberg_i$	=1 wenn Person i in Nürnberg wohnt, sonst =0
$PartnerNuernberg_i$	=1 wenn Partner von Person i in Nürnberg lebt, sonst =0
$Distanz_i$	Distanz zwischen Nürnberg und Heimatort von Person i in km

Name	$Nuernberg_i$	$PartnerNuernberg_i$	$Distanz_i$
Agnes	0	0	18
Boris	1	1	100
Clara	1	1	12
Dirk	1	1	41
Emilia	0	0	9
Franz	0	0	6

a) Nennen Sie zwei Nachteile des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

- Die vorhergesagten Werte für die abhängige Variable können Werte außerhalb (0,1) annehmen.
- Die Varianz der Störterme ist nicht konstant, es liegt Heteroskedastie vor. alternativ: Die Schätzung ist nicht effizient.
- Die Störterme sind nicht normalverteilt, daher sind t- und F-Tests nicht gültig.

b) Sie möchten folgendes Modell schätzen:

$$Nuernberg_i = \beta_0 + \beta_1 Distanz_i + u_i$$

Berechnen Sie $\hat{\beta}_0$ und $\hat{\beta}_1$. (4 Punkte)

Hinweise: $\overline{Nuernberg} = 0,5 \qquad \overline{Distanz} = 31 \qquad \sum_{i=1}^{n}(x_i - \bar{x})^2 = 6500$

Runden Sie alle Zwischenschritte auf die <u>dritte</u> Nachkommastelle.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \text{ und } \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

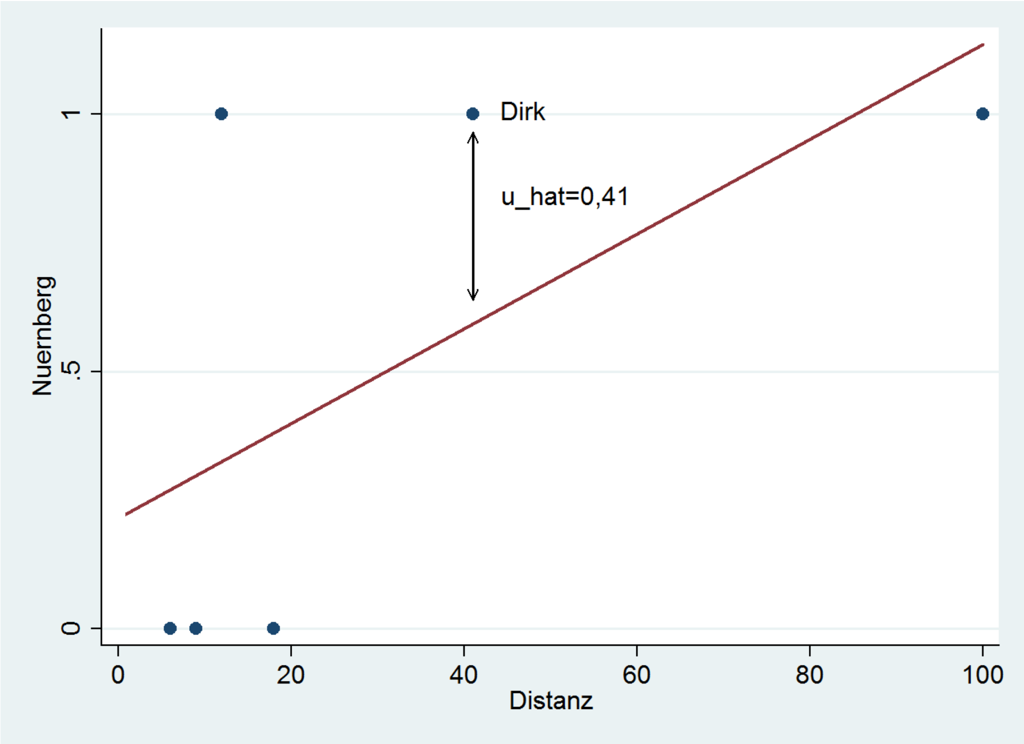
- $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y} = 18 \cdot 0 + 100 \cdot 1 + 12 \cdot 1 + 41 \cdot 1 + 9 \cdot 0 + 6 \cdot 0 - 6 \cdot 0,5 \cdot 31 = 60$
- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{60}{6500} = 0,009$
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0,5 - 0,009 \cdot 31 = 0,221$

c) Zeichnen Sie die sechs Datenpunkte und die Regressionsgerade der Regression aus Teilaufgabe b) in ein Koordinatensystem. Berechnen Sie das Residuum für Dirk und kennzeichnen Sie es in Ihrer Zeichnung. (4

Punkte)

Falls Sie in Teilaufgabe b) zu keinem Ergebnis gekommen sind, verwenden Sie folgende Schätzgleichung:
$\widehat{Nuernberg}_i = 0,1 + 0,02 \cdot Distanz_i$

- `Achsenbeschriftung und Skalierung korrekt`
- `Achsenabschnitt korrekt`
- `Steigung korrekt`
- `Datenpunkte korrekt eingetragen`
- `Residuum Dirk:` $\hat{u} = y - \hat{y} = 1 - (0,221 + 0,009 \cdot 41) = 0,41$
- `Residuum Dirk, falls kein Ergebnis in c):` $1 - (0,1 + 0,02 \cdot 41) = 0,08$
- `Residuum Dirk korrekt eingezeichnet`



d) Ein Kommilitone behauptet, dass die Nichtberücksichtigung der monatlichen finanziellen Mittel der Studierenden bei der Wahl des Wohnortes zu einer verzerrten Schätzung von β_1 führt. Nennen und erklären Sie die Bedingungen, unter denen er Recht hätte. (2 Punkte)

`Zwei Bedingungen müssen erfüllt sein, damit es in Folge einer ausgelassenen Variablen zu einer verzerrten Schätzung kommt:`

- `Die ausgelassene Variable muss relevant sein, i.e. die finanziellen Mittel müssen einen Einfluss auf die Wahl des Wohnortes haben.`
- `Die ausgelassene Variable (finanzielle Mittel) muss mit der erklärenden Variable (Distanz zwischen Nürnberg und Heimatort) korreliert sein.`

e) Gehen Sie nun von folgendem Modell aus:

$$Nuernberg_i = \beta_0 + \beta_1 PartnerNuernberg_i + u_i$$

Bestimmen Sie $\hat{\beta}_0$ und $\hat{\beta}_1$. (2 Punkte)
Hinweis: Für die Beantwortung dieser Teilaufgabe ist eine Berechnung wie in Teilaufgabe b) nicht erforderlich.

- $\hat{\beta}_0$ ist der Mittelwert der Variable ***Nuernberg***, wenn die Variable ***PartnerNuernberg*** den Wert Null annimmt. Daher gilt $\hat{\beta}_0 = 0$
- Da $\hat{\beta}_1 = \frac{\hat{y}-\hat{\beta}_0}{\bar{x}}$, nimmt $\hat{\beta}_1$ den Wert 1 an.
- Die abhängige und die erklärende Variable nehmen für alle Beobachtungen identische Werte an. Die gesamte Variation in der Variable ***Nuernberg*** wird durch die Variable ***PartnerNuernberg*** erklärt.

Aufgabe 2: **[10 Punkte]**

Im Auftrag einer Versicherung analysieren Sie die Einbruchszahlen in 200 Landkreisen in Deutschland für das Jahr 2016.

$Einbrueche_i$	Anzahl Einbrüche in Landkreis i je 1000 Wohneinheiten
$Grenznaehe_i$	=1 wenn Landkreis i an eine Staatsgrenze Deutschlands angrenzt, sonst =0
$log(EK_i)$	logarithmiertes Pro-Kopf-Jahreseinkommen in Landkreis i
$Alarm_i$	Anzahl Alarmanlagen je 1000 Wohneinheiten in Landkreis i

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Einbrueche_i = \beta_0 + \beta_1 Grenznaehe_i + \beta_2 log(EK_i) + \beta_3 Alarm_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten RegressionskoeffizientB	Standardfehler	T	Signifikanz
Konstante	206,604	156,018	1,32	0,187
Grenznaehe	66,573	28,519	2,33	0,021
log(EK)	6,796	12,196	0,56	0,578
Alarm	-2,891	0,136	-21,32	0,001

a. Abhängige Variable:Einbrueche

a) Interpretieren Sie $\hat{\beta}_2$ inhaltlich und statistisch. (2 Punkte)

- Steigt das Einkommen um 1%, so steigt die Zahl der Einbrüche c.p.i.M. um $\frac{\hat{\beta}_2}{100} = 0{,}068$ je 1000 Wohneinheiten.
- Der Effekt ist statistisch an keinem konventionellen Niveau signifikant. Alternativ: Der Effekt ist am 10%-Signifikanzniveau nicht signifikant.

b) Eine Sicherheitsfirma wirbt damit, dass eine zusätzliche Alarmanlage pro Tausend Wohneinheiten die Zahl der Einbrüche um mehr als 2,7 verringert. Überprüfen Sie diese Behauptung mittels eines statistischen Tests

am 10%- sowie am 5%-Signifikanzniveau. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung sowohl für das 10%-Signifikanzniveau als auch für das 5%-Signifikanzniveau an.
Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (6 Punkte)

- Testverfahren: einseitiger (linksseitiger) t-Test
- Hypothesen: $H_0 : \beta_3 \geq -2{,}7$, $H_1 : \beta_3 < -2{,}7$
- Teststatistik: $t = \frac{\hat{\beta}_3 - (-2{,}7)}{se(\hat{\beta}_3)} = \frac{-2{,}891 - (-2{,}7)}{0{,}136} = -1{,}404$
- Freiheitsgrade: $n - k - 1 = 200 - 3 - 1 = 196$
- Kritischer Wert c 5%-Signifikanzniveau: $c = t_{\alpha;n-k-1} = t_{0,05;196} = 1{,}645$
- Testentscheidung: Da $t_{empirisch} = -1{,}404 > -1{,}645 = -c$ kann die Nullhypothese auf dem 5%-Niveau nicht verworfen werden . Der Effekt ist nicht signifikant kleiner als -2,7. .
- Kritischer Wert c 10%-Signifikanzniveau: $c = t_{\alpha;n-k-1} = t_{0,1;196} = 1{,}282$
- Testentscheidung: Da $t_{empirisch} = -1{,}404 < -1{,}282 = -c$ kann die Nullhypothese auf dem 10%-Niveau verworfen werden . Der Effekt ist signifikant kleiner als -2,7. .

c) Wie hoch ist die Zahl der erwarteten Einbrüche in einem nicht an die Staatsgrenze angrenzenden Landkreis mit einem durchschnittlichen Pro-Kopf-Jahreseinkommen von 42.000 Euro und 50 Alarmanlagen je 1000 Wohneinheiten? (2 Punkte)

- $\widehat{Einbrueche} = 206{,}604 + 0 \cdot 66{,}573 + log(42.000) \cdot 6{,}796 - 50 \cdot 2{,}891 = 134{,}400$

Aufgabe 3: **[8,5 Punkte]**

Sie interessieren sich für die Determinanten der Hotelbewertungen von Nutzern einer Buchungswebseite. Sie haben folgende Informationen für 150 Hotels aus dem Jahr 2016:

$rating_i$	Durchschnittliche Bewertung in Punkten (1-10) von Hotel i
$stars_i$	Anzahl der Hotelsterne (1-5) von Hotel i
$location_i$	Entfernung des Hotels i vom Stadtzentrum in km
$price_i$	Preis pro Übernachtung im Hotel i

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$rating_i = \beta_0 + \beta_1 stars_i + \beta_2 location_i + \beta_3 stars_i \cdot location_i + \beta_4 log(price_i) + u_i$$

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	193,648	4	48,412	3,28	,038[b]
	Residuen	???	???	???		
	Gesamt	645,493	149			

a. Abhängige Variable: rating
b. Einflussvariablen: (Konstante), stars, location, stars · location, log(price)

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	7,890	1,235	6,389	0,000
stars	0,850	???	???	???
location	-0,120	0,029	-4,133	0,000
stars · location	-0,020	0,004	-4,544	0,000
log(price)	-0,900	0,297	-3,030	0,002

a. Abhängige Variable: rating

Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle.

a) Ein Kommilitone behauptet, dass die durchschnittliche Bewertung bei einem Anstieg der Hotelsterne für ein Hotel im Stadtzentrum stärker zunimmt als für ein Hotel am Stadtrand. Teilen Sie die Meinung Ihres Kommilitonen? Begründen Sie Ihre Antwort. (2 Punkte)

- $\frac{\Delta \widehat{rating}}{\Delta stars} = \hat{\beta}_1 + \hat{\beta}_3 \cdot location_i$
- $\hat{\beta}_3 = -0,020 < 0$
- Interpretation: Da $\hat{\beta}_3 < 0$, verbessert sich die durchschnittliche Bewertung eines Hotels mit einem zusätzlichen Hotelstern weniger, wenn das Hotels weiter vom Stadtzentrum entfernt ist. Die Vermutung ist somit bestätigt.

b) Berechnen Sie den Standardfehler des Koeffizienten $\hat{\beta}_1$. (4 Punkte)
Hinweis:
(i) Das Bestimmtheitsmaß der Regression von $stars_i$ auf die restlichen erklärenden Variablen beträgt 0,61.
(ii) Die Varianz der Variable $stars_i$ beträgt 0,240.

- $SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{SST_1 \cdot (1-R_1^2)}}$
- $\hat{\sigma}^2 = SER^2 = (\sqrt{\frac{SSR}{n-k-1}})^2 = \frac{SST-SSE}{n-k-1} = \frac{645,493-193,648}{150-4-1} = 3,116$
- $SST_1 = \sum_{i=1}^{N}(x_i - \bar{x})^2 = N \cdot Var(stars_i) = 150 \cdot 0,240 = 36$
- $1 - R_1^2 = 1 - 0,61 = 0,39$
- $SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{SST_1 \cdot (1-R_1^2)}} = \sqrt{\frac{3,116}{36 \cdot 0,39}} = 0,471.$

c) Wie ändern sich die geschätzten Koeffizienten der obigen Schätzung, wenn die Entfernung eines Hotels vom Stadtzentrum nicht in Kilometer, sondern in Meter gemessen worden wäre. (2,5 Punkte)
Hinweise: 1km = 1000m
Runden Sie auf die fünfte Nachkommastelle.

- Umskalierung der unabhängigen Variable $location$:
  - Entfernung in Kilometer: $location$

- Entfernung in Meter: $\widetilde{location} = 1000 \cdot location$

$$\widehat{rating}_i = \hat{\beta}_0 + \hat{\beta}_1 stars_i + \hat{\beta}_2 location_i + \hat{\beta}_3 stars_i \cdot location_i + \hat{\beta}_4 log(price_i)$$

$$\Leftrightarrow \widehat{rating}_i = \hat{\beta}_0 + \hat{\beta}_1 stars_i + \underbrace{\frac{\hat{\beta}_2}{1000}}_{\tilde{\beta}_2} \underbrace{1000 \cdot location_i}_{\widetilde{location}} + \underbrace{\frac{\hat{\beta}_3}{1000}}_{\tilde{\beta}_3} stars_i \cdot \underbrace{1000 \cdot location_i}_{\widetilde{location}} + \hat{\beta}_4 log(price_i)$$

$\Rightarrow \tilde{\beta}_2 = \frac{\hat{\beta}_2}{1000} = \frac{-0,120}{1000} = -0,00012$

$\Rightarrow \tilde{\beta}_3 = \frac{\hat{\beta}_3}{1000} = \frac{-0,020}{1000} = -0,00002$

$\Rightarrow$ Andere Koeffizienten sind nicht betroffen.

Aufgabe 4: **[17,5 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen Bildung und Einkommen. Sie haben folgende Informationen für 1563 Personen aus dem Jahr 2013:

$wage_i$	Monatliches Bruttogehalt von Person i in EURO
$educyrs_i$	Bildung von Person i in Jahren
IQ_i	IQ von Person i (in Punkten eines Tests)
$female_i$	=1 wenn Person i eine Frau ist; sonst =0
age_i	Alter von Person i in Jahren

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$\text{Modell I: } log(wage_i) = \beta_0 + \beta_1 educyrs_i + \beta_2 female_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,254(a)	,0647	???	,444

a. Einflussvariablen: (Konstante), educyrs, female

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	4.890	0,082	59,63	0,000
educyrs	0,021	0,003	3,89	0,000
female	-0,140	0,022	-6,25	0,000

a. Abhängige Variable: log(wage)

a) Interpretieren Sie $\hat{\beta}_2$ inhaltlich und statistisch. (2 Punkte)

- Inhaltliche Interpretation: Frauen verdienen monatlich ceteris paribus im Mittel 14,0% ($\hat{\beta}_2 \cdot 100\%$) (exakt: $exp(-0,14) - 1 = -0,131 = -13,1\%$ weniger als Männer.
- Statistische Interpretation: Der Koeffizient ist auf dem 1%-Niveau statistisch signifikant von Null verschieden (p-Wert: 0,000<0,010).

b) Interpretieren Sie das R^2 des Modells. (1 Punkt)

- Das Modell erklärt 6,47% der Variation im logarithmierten Gehalt.

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

- $\bar{R^2} = 1 - \frac{SSR/(n-k-1)}{SST/n-1)} = 1-(1-R^2)\frac{n-1}{n-k-1}$
- $\bar{R^2} = 1-(1-0,0647)\frac{1563-1}{1563-2-1}$
  $= 0,0635$

d) Berechnen und interpretieren Sie das 95%-Konfidenzintervall von $\hat{\beta}_2$. (3 Punkte)

- Das Konfidenzintervall wird für $\alpha = 0,05$ berechnet: $t_{\frac{0,05}{2};1563-2-1} = 1,960$
- $[\hat{\beta}_2 - 1,960 \cdot se(\hat{\beta}_2); \hat{\beta}_2 + 1,960 \cdot se(\hat{\beta}_2)] = [-0,140-1,960\cdot 0,022; -0,140+1,960\cdot 0,022]$
  $= [-0,184; -0,096]$
- Für wiederholte Stichproben liegt in 95% der Fälle der wahre Wert des Koeffizienten innerhalb der auf diese Weise berechneten Intervallgrenzen.

e) Handelt es sich bei dem geschätzten Koeffizienten der Variable *educyrs* um den kausalen Effekt der Bildung auf das Gehalt? Begründen Sie Ihre Antwort. (1,5 Punkte)

- $\hat{\beta}_1$ beschreibt den kausalen Effekt der Bildung auf das Gehalt unter der Annahme, dass alle anderen Einflussfaktoren konstant gehalten werden.
- In diesem Fall gibt es andere Variablen, die vermutlich sowohl mit Bildung, als auch mit dem Lohn korrelieren können, die in dieser Schätzung nicht berücksichtigt werden (z.B. IQ, Motivation,...).

-> Die hier geschätzte Korrelation zwischen Bildung und dem logarithmierten Bruttogehalt ist vermutlich kein kausaler Effekt der Bildung auf das Gehalt.

Sie schätzen zusätzlich ein zweites Regressionsmodell:

$$\text{Modell II: } log(wage_i) = \beta_0 + \beta_1 educyrs_i + \beta_2 female_i + \beta_3 IQ_i + \beta_4 age_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,281(a)	,0789	,0765	,489

a. Einflussvariablen: (Konstante), educyrs, female, IQ, age

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	4,421	0,094	47,03	0,000
educyrs	0,016	0,003	5,56	0,000
female	-0,141	0,022	-6,32	0,000
IQ	0,005	0,001	4,10	0,000
age	0,003	0,001	2,62	0,009

a. Abhängige Variable: log(wage)

f) Wie verändert sich $\hat{\beta}_1$ im Vergleich zur ersten Schätzung, in der nicht für *IQ* kontrolliert wird? Treffen Sie eine Aussage zur Richtung der Verzerrung von $\hat{\beta}_1$ in Modell 1. Welche Aussage lässt sich dadurch über die Kovarianz von *educyrs* und *IQ* treffen, wenn Sie annehmen, dass $Cov(educyrs, age) = 0$ und $Cov(female, IQ) = 0$? (2,5 Punkte)

- $\hat{\beta}_1$ sinkt im Vergleich zur ersten Schätzung -> β_1 wurde in Modell 1 überschätzt.
- Da $\hat{\beta}_1$ in Modell I überschätzt wird und $\hat{\beta}_3$ in Modell II >0 ist: $Cov(educyrs, IQ) > 0$.

g) Sie möchten testen, ob Alter und IQ gemeinsam einen signifikanten Einfluss auf das logarithmierte Monatsgehalt haben. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das 1%-Signifikanzniveau an. (6 Punkte)

- Testverfahren: F-Test
- H_0: $\beta_3 = \beta_4 = 0$, H_1: mindestens ein $\beta_j \neq 0$ mit $j = 3,4$
- Teststatistik: $F_{emp} = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)} = \frac{(0{,}0789-0{,}0647)/2}{(1-0{,}0789)/(1563-4-1)} = 12{,}009$
- Freiheitsgrade: $n-k-1 = 1563-4-1 = 1558$
- Kritischer Wert: $F_{kritisch} = F_{0{,}01;2;1558} = 4{,}61$
- Vergleich der Teststatistik mit kritischem Wert: $F_{emp} > F_{kritisch}$
- Testentscheidung: Die Nullhypothese kann auf dem 1%-Niveau verworfen werden. IQ und Alter tragen gemeinsam signifikant zur Erklärung des logarithmierten Bruttogehalts bei.

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.	Welche Aussage ist korrekt? Im linearen Wahrscheinlichkeitsmodell
a	kann die abhängige Variable y Werte außerhalb [0,1] annehmen.
b	**X** können die Koeffizienten stetiger erklärender Variablen als marginale Effekte interpretiert werden.
c	sind die Koeffizienten im Falle binärer erklärender Variablen effizient geschätzt.
d	nimmt $\hat{y}$ nur die Werte 0 oder 1 an.

2.	Sie stellen folgendes Modell auf: $income_i = \beta_0 + \beta_1 male_i + \beta_2 female_i + \beta_3 exper_i + \beta_4 exper^2 + \varepsilon_i$. Welche Aussage ist korrekt?
a	**X** Das Modell kann in der vorliegenden Form nicht geschätzt werden.
b	Der marginale Effekt von *exper* auf *income* ist $2\beta_4 \cdot exper$.
c	β_0 ist der Effekt für eine weibliche Person ohne Berufserfahrung.
d	Die Variable *exper* korreliert nicht mit *male*, wenn sie mit *female* korreliert.

3.	Die Aufnahme einer irrelevanten erklärenden Variable x_2 in das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$
a	führt zu einer Verringerung des Bestimmheitsmaßes R^2.
b	führt zu einer verzerrten Schätzung von β_1, wenn x_2 mit x_1 korreliert ist.
c	hat keine Auswirkung auf die Effizienz der Schätzung.
d	**X** reduziert im Falle der Korrelation mit x_1 die Effizienz der Schätzung.

4.	Eine Verletzung der Annahme MLR.5 (Homoskedastie)
a	führt zu einer Überschätzung von β_1, wenn die Varianz des Störterms mit x_1 zunimmt.
b	führt zu einer erwartungstreuen Schätzung der Koeffizienten und einer verzerrten Schätzung der marginalen Effekte.
c	führt zu korrekten Werten der t-Statistik der Koeffizientenschätzer.
d	**X** führt zu fehlerhaften Konfidenzintervallen.

5.	Ein Signifikanztest für den Koeffizientenschätzer $\hat{\beta}$ ergibt einen p-Wert gleich 0,04. Welche Aussage ist korrekt?
a	Das größte Signifikanzniveau, an dem $H_0 : \beta = 0$ abgelehnt werden kann, ist 0,04.
b	**X** Die Nullhypothese $H_0 : \beta = 0$ kann am 5%-Signifikanzniveau, nicht aber am 1%-Signifikanzniveau abgelehnt werden.
c	4% der Variation in den Beobachtungen werden durch β erklärt.
d	Die Wahrscheinlichkeit, die Nullhypothese $H_0 : \beta = 0$ fälschlicherweise abzulehnen, beträgt 96%.

6.	Welche Aussage bezüglich eines 95%-Konfidenzintervalls bei Signifikanztests ist korrekt?
a	Bei wiederholten Stichproben liegt der Schätzer $\hat{\beta}$ in 5% der Fälle nicht in dem ermittelten Konfidenzintervall.
b	Liegt der wahre Wert β außerhalb des Konfidenzintervalls, so beträgt die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise nicht abzulehnen 5%.
c	Wenn $\hat{\beta}$ in dem Konfidenzintervall liegt, so liegt der wahre Wert β bei wiederholten Stichproben zu 100% in dem Konfidenzintervall.
d	**X** Schließt ein Konfidenzintervall Null nicht mit ein, so ist $\hat{\beta}$ statistisch signifikant.

7.	Die Schätzung des Modells: $Einkommen_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$ ergibt $\hat{\beta}_1 = 6,1$; $\hat{\beta}_2 = 8,5$; $\hat{\beta}_3 = -0,2$. Bei welchem Wert von *exper* wird das Einkommen maximiert?
a	**X** 21,25
b	8,10
c	14,20
d	42,50

8.	Wann liegt ein Problem ausgelassener Variablen vor?
a	**X** Wenn eine erklärende Variable mit dem Störterm korreliert.
b	Wenn das R^2 der Regression sehr klein ist ($< 0,05$).
c	Wenn zwei berücksichtigte erklärende Variablen miteinander korrelieren.
d	Wenn die Konstante nicht signifikant von Null verschieden ist.

9.	Welche Aussage ist richtig?
a	Das angepasste R^2 belohnt die Aufnahme zusätzlicher erklärender Variablen.
b	Beim Vorliegen von Heteroskedastie kann kein R^2 berechnet werden.
c	**X** R^2 kann niemals durch die Aufnahme zusätzlicher erklärender Variablen sinken.
d	R^2 kann Werte kleiner als 0 annehmen.

10.	Gegeben sei folgendes Modell: $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 female_i + u_i$. Welches der folgenden Modelle ist in diesem Modell genestet?
a	$log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 female_i + u_i$
b	$wage_i = \beta_0 + \beta_2 female_i + \beta_3 exper_i + u_i$
c	**X** $wage_i = \beta_0 + \beta_1 educ_i + u_i$
d	$log(wage_i) = \beta_0 + \beta_1 educ_i + u_i$

11.	Gegeben sei folgendes Modell: $zufriedenheit_i = \beta_0 + \beta_1 verheiratet_i + u_i$. Sie gehen davon aus, dass es eine (unbeobachtete) Variable $kind_i$ gibt, die positiv mit $verheiratet$ und negativ 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	**X** wird β_1 unterschätzt.

12.	Im linearen Wahrscheinlichkeitsmodell
a	**X** können vorhergesagte Werte der abhängigen Variable > 1 auftreten.
b	kann man die Koeffizientengröße nicht sinnvoll interpretieren, wohl aber die Richtung des Zusammenhangs.
c	liegt zwangsläufig Homoskedastie vor.
d	können keine t-Statistiken berechnet werden.

13.	Welche Aussage bezüglich des linearen Regressionsmodells $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ist richtig?
a	Heteroskedastie führt zu falsch geschätzten Regressionskoeffizienten.
b	Ausgelassene Variablen führen zu Homoskedastie.
c	**X** Der Mittelwert der vorhergesagten Werte entspricht dem Mittelwert der abhängigen Variable.
d	Die Konstante verringert sich durch Aufnahme zusätzlicher erklärender Variablen zwangsläufig.

14.	Gegeben ist das Regressionsmodell: $arbeitsstunden_i = \beta_0 + \beta_1 mann_i + \beta_2 bildungsjahre_i + \beta_3 mann_i \cdot bildungsjahre_i + u_i$, wobei $mann_i$ den Wert 1 annimmt, falls die befragte Person i ein Mann ist, $arbeitsstunden_i$ die wöchentliche Arbeitszeit von i misst und $bildungsjahre_i$ die Schulzeit von i in Jahren misst. Welche der Aussagen ist richtig?
a	β_3 beschreibt den Zusammenhang zwischen Schulzeit und Arbeitszeit für Männer.
b	β_3 kann nicht sinnvoll interpretiert werden.
c	β_1 ist der Unterschied der Arbeitszeit zwischen Männern und Frauen am Mittel der Daten.
d	**X** β_2 beschreibt den Zusammenhang zwischen Schuljahren und Arbeitszeit für Frauen.

15.	Bei einem t-Test für einen Steigungsparameter mit der Hypothese $H_0 : \beta_1 = 0$
a	entsprechen die Freiheitsgrade der Größe der verwendeten Stichprobe.
b	**X** kann der empirische t-Wert aus $\hat{\beta}_1$ und dem Standardfehler von $\hat{\beta}_1$ berechnet werden.
c	wird getestet, ob β_1 sich signifikant von der Konstanten unterscheidet.
d	wird ein einseitiger t-Test durchgeführt.

16.	Ein $R^2 > 0,50$ bedeutet, dass
a	kein Problem ausgelassener Variablen vorliegt.
b	**X** über 50% der Variation in der abhängigen Variable durch die erklärenden Variablen erklärt werden.
c	jede einzelne erklärende Variable signifikant mit der abhängigen Variable korreliert.
d	Heteroskedastie kein Problem im Modell darstellt.

17.	Gegeben sei die Regression $y_i = \beta_0 + \beta_1 x_i + u_i$. Sie schätzen dieses Modell mit einer Stichprobe mit 9887 Beobachtungen und erhalten $\hat{\beta}_1 = 2,762$ und $se(\hat{\beta}_1) = 0,20$. Welche Aussage ist richtig?
a	$\hat{\beta}_1$ ist in jedem Fall der kausale Effekt von x auf y.
b	Der Standardfehler deutet auf das Vorliegen von Heteroskedastie hin.
c	**X** $\hat{\beta}_1$ ist auf dem 1%-Niveau statistisch signifikant von Null verschieden.
d	Es kann keine Aussage bezüglich der Signifikanz von $\hat{\beta}_1$ getroffen werden.

18.	Sie schätzen die lineare Regression $Alkoholkonsum_i = \beta_0 + \beta_1 Alter_i + u_i$, wobei $Alkoholkonsum_i$ die Anzahl der Tage pro Woche, in denen Person i Alkohol trinkt, und $Alter_i$ das Alter von Person i in Jahren angibt. Das 95%-Konfidenzintervall von $\hat{\beta}_1$ lautet [-1,672;2,002]. Welche der Aussagen ist richtig?
a	**X** $\hat{\beta}_1$ ist am 5%-Niveau nicht statistisch signifikant von Null verschieden.
b	$\hat{\beta}_1$ beschreibt den kausalen Effekt des Alters auf den Alkoholkonsum.
c	Man kann keine Aussage über die statistische Signifikanz von $\hat{\beta}_1$ treffen.
d	$\hat{\beta}_0$ ist statistisch signifikant von Null verschieden.

19.	Sie verfügen über einen Querschnittsdatensatz aus dem Jahr 2013 mit der Variable age_i, die das Alter von Person i in Jahren misst. Sie errechnen die Variable $exper_i$ als Maß für die potenzielle Berufserfahrung als $Befragungsjahr - age_i - 15$ und schätzen das lineare Regressionsmodell $wage_i = \beta_0 + \beta_1 age_i + \beta_2 exper_i + u_i$. Welches Problem tritt bei der Schätzung auf?
a	Es liegt Heteroskedastie vor.
b	**X** Es liegt Multikollinearität vor.
c	Es liegt Homoskedastie vor.
d	Die Schätzkoeffizienten lassen sich nicht interpretieren.

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

21.	Sie regressieren in einer KQ-Schätzung das monatliche Bruttoeinkommen in Euro auf das Alter: $inc_i = \beta_0 + \beta_1 age_i + u_i$. Ihre geschätzten Koeffizienten sind: $\hat{\beta}_0 = 452$ und $\hat{\beta}_1 = 67$. Mit zunehmenden Alter steigt die Varianz des Einkommens. Welche Aussage trifft auf die Regression zu?
a	Es liegt Homoskedastie vor.
b	Mit zunehmenden Alter sinkt ceteris paribus im Mittel das Einkommen.
c	**X** Ein 45-jähriger hat ein vorhergesagtes Einkommen von 3467 Euro.
d	Im Modell wird keine Konstante geschätzt.

22.	Gegeben sei folgendes Modell: $wage_i = \beta_0 + \beta_1 male_i + u_i$, wobei $wage_i$ der individuelle Stundenlohn von i ist und $male_i$ eine Dummyvariable, die den Wert 0 annimmt, falls i eine Frau ist und den Wert 1 annimmt, falls i ein Mann ist. Welche Aussage ist richtig?
a	$\beta_0 - \beta_1$ ist der durchschnittliche Stundenlohn von Männern.
b	**X** β_0 ist der durchschnittliche Stundenlohn von Frauen.
c	$\beta_0 - \beta_1$ ist der durchschnittliche Stundenlohn von Frauen.
d	β_1 ist der durchschnittliche Stundenlohn von Männern.

23.	Sie möchten auf dem 5%-Signifikanzniveau testen, ob sich ein Koeffizient $\hat{\beta}_1$ von 0 unterscheidet. Der Datensatz, mit dem Sie die Schätzung durchführen, enthält 6402 Beobachtungen. Wie lautet in diesem Fall der kritische Wert der t-Statistik?
a	**X** $t_{kritisch} = 1,960$
b	$t_{kritisch} = 1,645$
c	$t_{kritisch} = 2,326$
d	$t_{kritisch} = 1,658$

24.	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	**X** H_0 verworfen wird, obwohl sie zutrifft.
d	H_0 nicht verworfen wird, obwohl sie nicht zutrifft.

25.	Unter den Gauss-Markov-Annahmen sind die Störterme eines linearen Regressionsmodells
a	**X** homoskedastisch.
b	normalverteilt.
c	standardnormalverteilt.
d	log-normalverteilt.

26.	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 umgekehrt U-förmigen Zusammenhang zwischen $\hat{y}_i$ und x_i hin?
a	$\hat{\beta}_0 < 0$ und $\hat{\beta}_2 > 0$.
b	$\hat{\beta}_0 > 0$ und $\hat{\beta}_2 < 0$.
c	$\hat{\beta}_1 < 0$ und $\hat{\beta}_2 > 0$.
d	**X** $\hat{\beta}_1 > 0$ und $\hat{\beta}_2 < 0$.

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

28.	Eine KQ-Schätzung $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ liefert $\hat{y}_i = 5 - 2x_i$. Welchen Wert hat y_1 für $x_1 = 2$ und $\hat{u}_1 = 2$?
a	-1
b	1
c	**X** 3
d	4

29.	Ein Konfidenzintervall wird breiter
a	mit steigendem Signifikanzniveau α.
b	**X** mit steigendem Standardfehler $se(\beta_j)$.
c	mit steigendem Schätzwert $\hat{\beta}_j$.
d	mit steigender Stichprobengröße.

30.	Beim F-Test
a	**X** gilt bei zweiseitigen Hypothesentests bezüglich einzelner Parameter $F_{\alpha,1,n-k-1} = t^2_{\frac{\alpha}{2},n-k-1}$.
b	kann die Teststatistik einen negativen Wert annehmen.
c	steigt der Wert der Teststatistik mit steigendem Signifikanzniveau α.
d	wird die Nullhypothese verworfen, wenn F kleiner als der kritische Wert ist.

31.	Wenn die Störterme u eines multiplen Regressionsmodells nicht normalverteilt sind, dann
a	können keine Teststatistiken berechnet werden.
b	**X** können t- und F-Tests für kleine Stichproben das falsche Ergebnis liefern.
c	ist der KQ-Schätzer verzerrt.
d	ist der KQ-Schätzer inkonsistent.

32.	Die Aufnahme irrelevanter Variablen in ein Regressionsmodell
a	verringert den Wert des R^2.
b	erhöht den Wert des angepassten Bestimmtheitsmaßes.
c	erhöht die Residuenquadratsumme SSR.
d	**X** verringert die Präzision der Schätzung.

33.	Der Standardfehler von $\hat{\beta}_j$
a	ist keine Zufallsvariable.
b	ist ein Schätzer für den Standardfehler der Regression.
c	sinkt mit steigendem unerklärten Teil der Variation in der abhängigen Variable.
d	**X** steigt mit steigender Residuenquadratsumme (SSR).

34.	Was ändert sich bei einer Umskalierung der abhängigen Variable?
a	Die t-Teststatistiken aus den Signifikanztests der einzelnen Parameter.
b	Das R^2 der Schätzung.
c	Das angepasste Bestimmtheitsmaß der Schätzung.
d	**X** Die Standardfehler der Parameter.

35.	Ein Chow-Test
a	**X** für zwei Gruppen kann auf Basis der Schätzung eines einzigen Regressionsmodells durchgeführt werden.
b	kann nicht mit Hilfe von Interaktionstermen durchgeführt werden.
c	hat eine t-verteilte Teststatistik.
d	hat die Alternativhypothese, dass die Steigungsparameter für verschiedene Gruppen identisch sind.

36.	Logarithmieren der abhängigen Variablen
a	schließt Beobachtungen mit $y = 1$ von der Schätzung aus.
b	**X** senkt die Streuung der abhängigen Variablen.
c	steigert den Einfluss der Ausreißerbeobachtungen auf die Schätzergebnisse.
d	erhöht die Streuung der abhängigen Variablen.

37.	Die Schätzung des Modells $Analphabetenquote = \beta_0 + \beta_1 male + u$ (wobei $Analphabetenquote$ in 0-100% gemessen wird) ergibt $\widehat{\beta}_1 = -7$. Daher gilt:
a	Frauen haben eine um 7 Prozent höhere Analphabetenquote als Männer.
b	Männer haben eine um 7 Prozent höhere Analphabetenquote als Frauen.
c	**X** Frauen haben eine um 7 Prozentpunkte höhere Analphabetenquote als Männer.
d	Männer haben eine um 7 Prozentpunkte höhere Analphabetenquote als Frauen.

38.	Unter den Annahmen des klassischen linearen Modells folgt der nicht standardisierte KQ-Schätzer
a	einer F-Verteilung.
b	einer t-Verteilung.
c	**X** einer Normalverteilung.
d	einer Standardnormalverteilung.

39.	Sie führen nacheinander einen rechtsseitigen und einen linksseitigen t-Test durch. Wie unterscheiden sich die kritischen Werte im Betrag, wenn beide Tests für das gleiche Modell, die gleiche Stichprobe und das gleiche Signifikanzniveau durchgeführt werden?
a	Der kritische Wert des rechtsseitigen Tests ist größer.
b	**X** Der kritische Wert ist in beiden Tests gleich groß.
c	Der kritische Wert des linksseitigen Tests ist größer.
d	Die Antwort hängt von dem in der Nullhypothese unterstellten Wert ab.

40.	Ein konsistenter Schätzer
a	kann nicht verzerrt sein.
b	kann nicht unverzerrt sein.
c	erfüllt notwendigerweise die Annahmen MLR.1-MLR.4.
d	**X** erfüllt $plim\ \hat{\beta} = \beta$.