Bachelorprüfung WS 2016/2017 - 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: **[8,5 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen Waffenbesitz und Wählerverhalten in den USA. Hierfür stehen Ihnen folgende Daten zur Verfügung:

	Staat hat für Trump gestimmt 0/1	Anteil der Waffenbesitzer in %
Arizona	1	24
Kalifornien	0	14
Kansas	1	34
Mississippi	1	52
New Jersey	0	7
Texas	1	27

Sie stellen folgendes lineares Wahrscheinlichkeitsmodell auf:

$$Trump_i = \beta_0 + \beta_1 Waffenbesitz_i + u_i$$

a) Berechnen Sie $\hat{\beta}_0$ und $\hat{\beta}_1$. (4,5 Punkte)

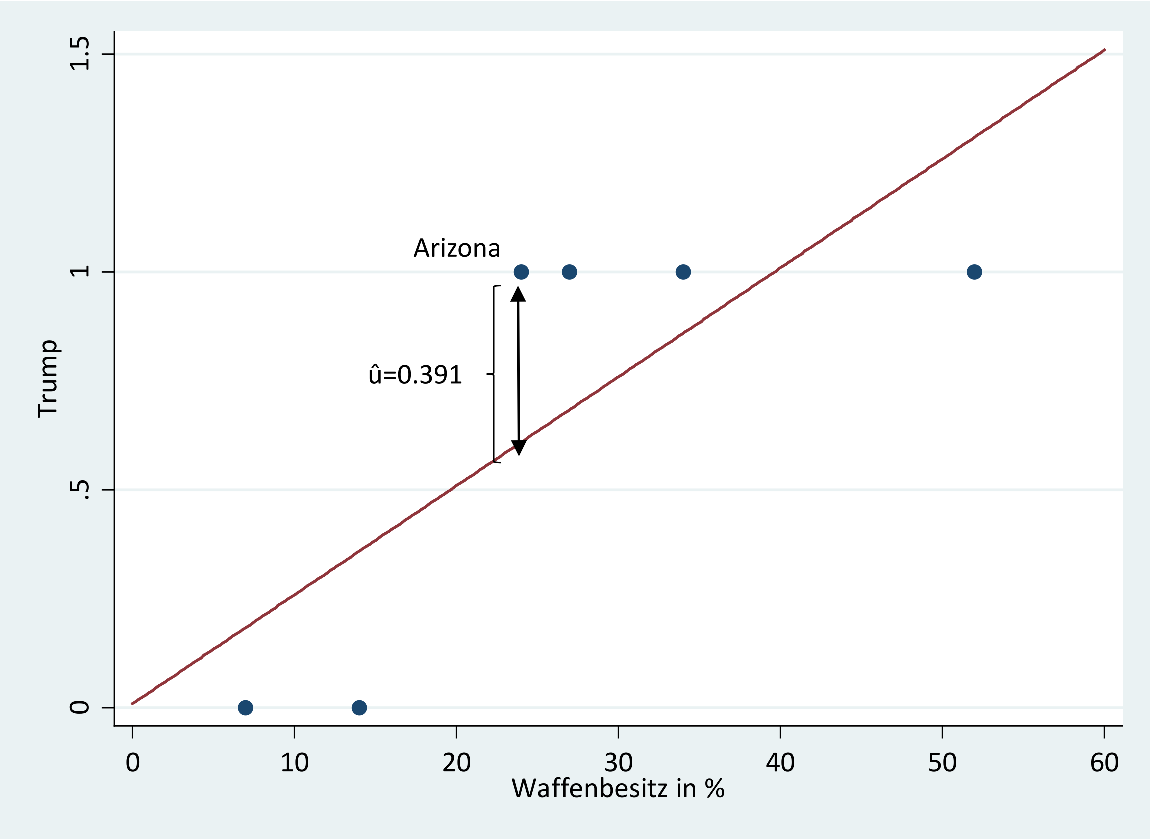
Hinweise: $\overline{Waffenbesitz} = 26,3$; $\overline{Trump} = \frac{2}{3}$.
Runden Sie alle Zwischenschritte auf die <u>*dritte*</u> *Nachkommastelle.*

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

- $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i y_i - n\cdot\bar{x}\cdot\bar{y} = 24\cdot 1 + 14\cdot 0 + 34\cdot 1 + 52\cdot 1 + 7\cdot 0 + 27\cdot 1 - 6\cdot 26,3\cdot\frac{2}{3} = 31,8$
- $\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\cdot\bar{x}^2 = 24^2 + 14^2 + 34^2 + 52^2 + 7^2 + 27^2 - 6\cdot 26,3^2 = 1259,86$
- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{31,8}{1259,86} = 0,025$
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 0,667 - 0,025\cdot 26,3 = 0,01$

b) Zeichnen Sie die Datenpunkte und die Regressionsgerade in ein Koordinatensystem. Berechnen Sie das Residuum für Arizona und kennzeichnen Sie es in Ihrer Zeichnung. (4 Punkte)
Falls Sie in Teilaufgabe a) zu keinem Ergebnis gekommen sind, verwenden Sie folgende Schätzgleichung:
$\widehat{Trump}_i = -0,5 + 0,01\cdot Waffenbesitz_i$

- Achsenbeschriftung und Skalierung korrekt
- Achsenabschnitt korrekt
- Steigung korrekt
- Datenpunkte korrekt eingetragen
- Residuum Arizona: $\hat{u} = y - \hat{y} = 1 - (0,01 + 0,025\cdot 24) = 0,39$
- Residuum Arizona, falls kein Ergebnis bei a): $1 - (-0,5 + 0,01\cdot 24) = 1,26$
- Residuum Arizona korrekt eingezeichnet



Aufgabe 2: **[9 Punkte]**

Sie interessieren sich für die Determinanten des Bierkonsums in verschiedenen Ländern. Sie haben folgende Informationen für 30 OECD Staaten für das Jahr 2009:

$Biermenge_i$	Durchschnittlicher Pro-Kopf-Verbrauch von Bier in Litern in Land i
$Bierpreis_i$	Durchschnittlicher Bierpreis in US-Dollar pro Liter in Land i
P_i	Wechselkursbereinigtes Preisniveau in Land i, relativ zu den USA (P_{USA}=1)
$log(EK_i)$	Natürlicher Logarithmus des Pro-Kopf-Einkommens in Land i

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Biermenge_i = \beta_0 + \beta_1 Bierpreis_i + \beta_2 (Bierpreis_i)^2 + \beta_3 P_i + \beta_4 log(EK_i) + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	-419,998	258,252	-1,626	0,117
Bierpreis	-45,989	11,814	-3,893	0,001
Bierpreis^2	3,948	1,829	2,159	0,041
P	11,202	???	???	???
log(EK)	57,298	27,137	2,111	0,045

a. Abhängige Variable: Biermenge

a) Interpretieren Sie $\hat{\beta}_4$ inhaltlich und statistisch. (2 Punkte)

- Inhaltliche Interpretation: Länder mit einem 1% höheren Pro-Kopf-Einkommen haben c.p. im Mittel einen um $\frac{\hat{\beta}_4}{100} = 0{,}573$ Liter höheren Pro-Kopf-Bierkonsum.
- Statistische Interpretation: Der Koeffizient ist auf dem 5%-Niveau statistisch signifikant von Null verschieden (p-Wert: 0,045<0,05).

b) Bei welchem Bierpreis ist der Bierkonsum am niedrigsten? (1,5 Punkte)

$$\frac{\partial \widehat{Biermenge}}{\partial Bierpreis} \stackrel{!}{=} 0$$

$$\Leftrightarrow \quad \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot Bierpreis^* = 0$$

$$\Leftrightarrow \quad -45{,}989 + 2 \cdot 3{,}948 \cdot Bierpreis^* = 0 \Leftrightarrow Bierpreis^* = 5{,}824$$

- Der Bierkonsum wird bei einem Preis von 5,82 US-Dollar minimiert.

c) Die USA haben ein Pro-Kopf-Einkommen von 57.220 Dollar, ein Liter Bier kostet durchschnittlich 2,60 Dollar. Wie hoch ist der erwartete Pro-Kopf-Bierkonsum der USA? Geben Sie Ihren Rechenweg an. (2 Punkte)

- $\widehat{Biermenge}_{USA} = \hat{\beta}_0 + \hat{\beta}_1 Bierpreis_{USA} + \hat{\beta}_2 (Bierpreis_{USA})^2 + \hat{\beta}_3 P_{USA} + \hat{\beta}_4 log(EK_{USA})$
- $\widehat{Biermenge}_{USA} = -419{,}998 - 45{,}989 \cdot 2{,}60 + 3{,}948 \cdot 2{,}60^2 + 11{,}202 \cdot 1 + 57{,}298 \cdot log(57.220)$
- $\widehat{Biermenge}_{USA} = 126{,}001$

d) Berechnen Sie den Standardfehler des Koeffizienten $\hat{\beta}_3$. (3,5 Punkte)
Hinweise:

i. Der Standardfehler der Regression beträgt 3,43.
ii. Das Bestimmtheitsmaß der Regression von P_i auf die restlichen erklärenden Variablen beträgt 0,73.
iii. Die Varianz der Variable P_i beträgt 0,086.

- $SE(\hat{\beta}_3) = \sqrt{\frac{\hat{\sigma}^2}{SST_j \cdot (1-R_j)}}$
- $\hat{\sigma}^2 = SER^2 = 3{,}43^2 = 11{,}765$
- $SST_j = \sum_{i=1}^{n} (x_i - \bar{x})^2 = N \cdot Var(P_i) = 30 \cdot 0{,}086 = 2{,}58$
- $1 - R_j = 1 - 0{,}73 = 0{,}27$
- $\sqrt{\frac{\hat{\sigma}^2}{SST_j \cdot (1-R_j)}} = \sqrt{\frac{11{,}765}{2{,}58 \cdot 0{,}27}} = 4{,}11$

Aufgabe 3: **[16,5 Punkte]**

Als Marketingmanager eines Modekonzerns möchten Sie die Determinanten einer erfolgreichen Winterschlussverkauf-Woche (WSV-Woche) untersuchen. Hierfür haben Sie Informationen zu der WSV-Woche 2016 aus 431 Filialen eingeholt.

$Umsatz_i$	Umsatz in der WSV-Woche 2016 in Filiale i (in Tausend Euro)
$Temp\_1_i$	=1 wenn Durchschnittstemperatur vor Filiale i in WSV Woche $\leq -5°C$, sonst 0
$Temp\_2_i$	=1 wenn Durchschnittstemperatur vor Filiale i in WSV Woche $> -5°C$ und $\leq 5°C$, sonst 0
$Temp\_3_i$	=1 wenn Durchschnittstemperatur vor Filiale i in WSV Woche $> 5°C$, sonst 0
$Marketing_i$	Werbeausgaben für WSV-Woche im Einzugsgebiet von Filiale i (in Tausend Euro)

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$log(Umsatz_i) = \beta_0 + \beta_1 Temp\_2_i + \beta_2 Temp\_3_i + \beta_3 Marketing_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,530(a)	,281	,284	24,586

a. Einflussvariablen: (Konstante), Temp_2, Temp_3, Marketing

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	-0,009	8,324	-0,001	0,999
Temp_2	0,047	0,027	1,741	0,082
Temp_3	-0,069	0,026	-2,654	0,008
Marketing	0,115	0,026	4,423	0,000

a. Abhängige Variable: log(Umsatz)

a) Welche Auswirkung hätte die zusätzliche Aufnahme der Variable $Temp\_1$ in Ihr Modell für die Schätzung? Begründen Sie Ihre Aussage mit der relevanten Gauss-Markov Annahme. (2 Punkte)

- Bei Aufnahme der Variable $Temp\_1$ wäre der KQ-Schätzer nicht mehr berechenbar.
- Alternativ: Die Konstante könnte nicht mitgeschätzt werden.
- Es würde perfekte Multikollinearität vorliegen (MLR.3).

b) Interpretieren Sie den geschätzten Koeffizienten von $Temp\_3$ inhaltlich und statistisch. (2 Punkte)

- Inhaltliche Interpretation: Bei einer Durchschnittstemperatur von über $5°C$ ist der Umsatz c.p. im Mittel um 6,9% niedriger als bei einer Durchschnittstemperatur von höchstens $-5°C$.
- Statistische Interpretation: Der Koeffizient ist auf dem 1%-Niveau statistisch signifikant von Null verschieden (p-Wert: $0{,}008 < 0{,}01$).

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

- Das Modell erklärt 28,1% der Variation im Umsatz der Filialen während der WSV-Woche.

d) In Kleinstadt A lag die Durchschnittstemperatur während der WSV-Woche 2016 bei $7°C$, in Kleinstadt B bei $2°C$. Um wie viel unterscheidet sich der Umsatz von Kleinstadt A im Vergleich zu Kleinstadt B durch den Temperaturunterschied, wenn sich A und B sonst nicht unterscheiden? Wo ist der Umsatz höher? (2 Punkte)

- Unterschied Kleinstadt A ($Temp\_3 = 1$)-Kleinstadt B ($Temp\_2 = 1$) $= \hat{\beta}_2 - \hat{\beta}_1 = -0{,}069 - 0{,}047 = -0.116$
- Bei einer Temperatur von $7°C$ (Kleinstadt A) ist der Umsatz in der WSV-Woche ceteris paribus um 11.6% niedriger als bei $2°C$ (Kleinstadt B).

e) Sie möchten testen, ob die Durchschnittstemperatur insgesamt einen signifikanten Einfluss auf den Umsatz der Filialen in der WSV-Woche hat. Geben Sie das Testverfahren und die Hypothesen an. Geben Sie die Teststatistik an und setzen Sie die bekannten Werte ein. Welcher Wert fehlt und wie können Sie ihn bestimmen? Geben Sie Freiheitsgrade, kritischen Wert und die Entscheidungsregel für das 1%-Signifikanzniveau an. (5,5 Punkte) *Hinweis: Berechnung der Teststatistik nicht notwendig.*

- Testverfahren: F-Test
- H_0: $\beta_1 = \beta_2 = 0$, H_1: mindestens ein $\beta_j \neq 0$ mit $j = 1,2$
- Teststatistik: $F_{emp} = \frac{(R_u^2 - R_r^2)/q}{R_u^2/(n-k-1)} = \frac{(0{,}281 - R_r^2)/2}{0{,}281/(431-3-1)}$ ($q = 2$, $R_u^2 = 0{,}281$)
- R_r^2 ergibt sich aus Schätzung des folgenden Modells: $log(Umsatz_i) = \alpha_0 + \alpha_1 Marketing_i + v_i$
- Freiheitsgrade: $n - k - 1 = 431 - 3 - 1 = 427$
- Kritischer Wert: $F_{kritisch} = F_{0{,}01;2;427} = 4{,}61$
- Entscheidungsregel: Wenn $F_{emp} > F_{kritisch}$ kann die Nullhypothese auf dem 1%-Niveau verworfen werden.

f) Ein Kollege vermutet, dass die Annahme $E(u|Temp\_2, Temp\_3, Marketing) = 0$ in Ihrem Modell verletzt sein könnte. Erläutern Sie diese Annahme und beschreiben Sie die Auswirkung einer Verletzung dieser Annahme auf die geschätzten Koeffizienten. Beschreiben Sie anhand eines Beispiels, warum die Annahme hier verletzt sein könnte. (4 Punkte)

- Die Annahme $E(u|Temp\_2, Temp\_3, Marketing) = 0$ beschreibt die mittlere bedingte Unabhängigkeit des Störterms von den erklärenden Variablen.
- Bei Verletzung dieser Annahme werden die Koeffizienten verzerrt geschätzt.

- Beispiel: In der obigen Schätzung wird z.B. die Wettbewerbssituation der einzelnen Filiale nicht berücksichtigt. Die Wettbewerbssituation korreliert vermutlich mit den Marketingausgaben und hat einen Einfluss auf den Umsatz während der WSV-Woche. [2P] Der Koeffizient β_3 wird deshalb unterschätzt.

Aufgabe 4: **[16 Punkte]**

Sie interessieren sich für die Determinanten der Lebenserwartung bei Geburt. Ihnen steht ein Querschnittsdatensatz für 65 Länder aus dem Jahr 2009 mit folgenden Informationen zur Verfügung:

$Lebenserwartung_i$	Lebenserwartung bei Geburt im Land *i* in Jahren
$Armut_i$	=1, wenn Land *i* arm ist, =0 wenn reich
$Gesundheitsausg_i$	Ausgaben der Regierung im Land *i* für Gesundheit pro Kopf in Tausend US-Dollar
$Alphabet_i$	Alphabetisierungsrate im Land *i* in % der Gesamtbevölkerung (kodiert 0-100)

Sie schätzen das folgende lineare Regressionsmodell mit SPSS:

$$Lebenserwartung_i = \beta_0 + \beta_1 Armut_i + \beta_2 Gesundheitsausg_i + \beta_3 Alphabet_i + \varepsilon_i \quad \text{(Modell I)}$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
Konstante	50,744	3,572	14,21	0,000
Armut	-8,510	1,815	-4,69	0,000
Gesundheitsausg	0,071	0,018	3,87	0,000
Alphabet	0,159	0,046	3,49	0,001

a. Abhängige Variable: Lebenserwartung

a) Interpretieren Sie den geschätzten Koeffizienten $\hat{\beta}_3$ inhaltlich. (1 Punkt)

- Inhaltlich: $\hat{\beta}_3 = 0,159$: Wenn die Alphabetisierungsrate um 1 Prozentpunkt steigt, steigt die Lebenserwartung c.p. im Mittel um 0,159 Jahre.

b) Welchen Wert würde $\hat{\beta}_2$ annehmen, wenn die Gesundheitsausgaben nicht in Tausend US-Dollar, sondern in Tausend Euro gemessen worden wären? *Hinweis:* Nehmen Sie einen Wechselkurs von 1 US-Dollar=1,5 Euro an. (1,5 Punkte)

- Umskalierung der unabhängigen Variable *Gesundheitsausg*:
  - Gesundheitsausgaben in US-Dollar: $Gesundheitsausg$
  - Gesundheitsausgaben in Euro: $\widetilde{Gesundheitsausg} = 1,5 \cdot Gesundheitsausg$

$$\widehat{Lebenserwartung}_i = \hat{\beta}_0 + \hat{\beta}_1 Armut_i + \hat{\beta}_2 Gesundheitsausg_i + \hat{\beta}_3 Alphabet_i$$

$$\Leftrightarrow \widehat{Lebenserwartung}_i = \hat{\beta}_0 + \hat{\beta}_1 Armut_i + \underbrace{\frac{\hat{\beta}_2}{1,5}}_{\tilde{\beta}_2} \cdot \underbrace{1,5 \cdot Gesundheitsausg_i}_{\widetilde{Gesundheitsausg_i}} + \hat{\beta}_3 Alphabet_i$$

$$\Rightarrow \tilde{\beta}_2 = \frac{0,071}{1,5} = 0,0473$$

c) Ein Kommilitone behauptet, dass die Lebenserwartung bei einem Anstieg der Gesundheitsausgaben in armen Ländern stärker zunimmt als in reichen Ländern. Um dies zu überprüfen, nehmen Sie einen Interaktionsterm in Ihr Modell auf und schätzen dieses erneut mit SPSS. Ihr Regressionsmodell und die Schätzergebnisse lauten wie folgt:

$$Lebenserwartung_i = \beta_0 + \beta_1 Armut_i + \beta_2 Gesundheitsausg_i + \beta_3 Alphabet_i$$
$$+ \beta_4 (Gesundheitsausg_i \cdot Armut_i) + \varepsilon_i \quad \text{(Modell II)}$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
Konstante	50,012	3,578	13,98	0,000
Armut	-5,875	2,539	-2,31	0,024
Gesundheitsausg	0,068	0,017	4,00	0,000
Alphabet	0,168	0,046	3,67	0,000
Gesundheitsausg · Armut	-0,282	0,152	-1,86	0,067

a. Abhängige Variable: Lebenserwartung

i. Teilen Sie die Meinung Ihres Kommilitonen? Begründen Sie Ihre Antwort. (2 Punkte)

- $\frac{\Delta \widehat{Lebenserwartung}}{\Delta Gesundheitsausg \Delta Armut} = \hat{\beta}_4$.
- $\hat{\beta}_4 = -0,282 < 0$.
- Interpretation: Da $\hat{\beta}_4 < 0$, steigt die Lebenserwartung bei einem Anstieg der Gesundheitsausgaben für arme Länder weniger als für reiche Länder (sie sinkt sogar, da $\hat{\beta}_2 + \hat{\beta}_4 < 0$). [0,5P] Die Schätzergebnisse widersprechen somit der Behauptung des Kommilitonen.

ii. Sie möchten testen, ob die Lebenserwartung für reiche Länder bei einer Erhöhung der Gesundheitsausgaben um 1.000 US-Dollar pro Kopf um mehr als 0,03 Jahre steigt. 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)

- Testverfahren: einseitiger (rechtsseitiger) t-Test
- Hypothesen: $H_0 : \beta_2 \leq 0,03$, $H_1 : \beta_2 > 0,03$

- Teststatistik: $t = \frac{\hat{\beta}_2 - 0,03}{se(\hat{\beta}_2)} = \frac{0,068 - 0,03}{0,017} = 2,235$
- Freiheitsgrade: $n - k - 1 = 65 - 4 - 1 = 60$
- Kritischer Wert c: $c = t_{\alpha;n-k-1} = t_{0,05;60} = 1,671$
- Testentscheidung: Da $t_{empirisch} = 2,235 > 1,671 = c$ kann die Nullhypothese auf dem 5%-Niveau verworfen werden. Die Vermutung wird bestätigt.

iii. Wie würde Ihre Testentscheidung für das 10%-Signifikanzniveau ausfallen? Begründen Sie Ihre Antwort. (1 Punkt)

- Die Nullhypothese kann auf dem 10%-Niveau verworfen werden.
- Der kritische Wert beim 10%-Signifikanzniveau ist kleiner als beim 5%-Signifikanzniveau und damit ist der empirische Wert der Teststatistik auch hier größer als der kritische Wert.

iv. Das angepasste Bestimmtheitsmaß $\overline{R}^2$ aus Modell I beträgt $0,663$. Berechnen Sie das angepasste Bestimmtheitsmaß $\overline{R}^2$ für Modell II, wobei $SSE = 5235,717$ und $SSR = 2375,140$ gilt. Welchem der beiden Modelle würden Sie den Vorrang geben? Runden Sie alle Zwischenschritte auf die dritte Nachkommastelle. (3 Punkte)

- $\overline{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{2375,140/(65-4-1)}{(5235,717 + 2375,140)/(65-1)} = 1 - \frac{39,586}{118,9200} = 0,667.$
- Modell II würde der Vorrang gegeben werden, da das $\overline{R}^2$ größer ist als in Modell I.

v. Sie sollen ein Konfidenzintervall für β_1 aus Modell II berechnen. Ihr Kommilitone hat bereits angefangen und die folgende Gleichung aufgestellt:

$[\hat{\beta}_1 - 2,000 \cdot se(\hat{\beta}_1); \hat{\beta}_1 + 2,000 \cdot se(\hat{\beta}_1)]$

Benennen Sie das Signifikanzniveau, zu dem das Konfidenzintervall berechnet wird. Berechnen Sie das Konfidenzintervall und interpretieren Sie Ihr Ergebnis. (3 Punkte)

- $2,000 = t_{\frac{0,05}{2};65-4-1}$ $\Rightarrow$ Das Konfidenzintervall wird für $\alpha = 0,05$ berechnet.
- $[\hat{\beta}_1 - 2,000 \cdot se(\hat{\beta}_1); \hat{\beta}_1 + 2,000 \cdot se(\hat{\beta}_1)] = [-5,875 - 2,000 \cdot 2,539; -5,875 + 2,000 \cdot 2,539]$ $= [-10,953; -0,797]$
- Für wiederholte Stichproben liegt in 95% der Fälle der wahre Wert innerhalb der auf diese Weise berechneten Intervallgrenzen.

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.	Gegeben ist das Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, wobei x_i und y_i binäre Variablen (0/1) sind. Zwei weitere Variablen sind definiert als $z_i = 1 - y_i$ und $w_i = 1 - x_i$. Wie verändert sich der Wert des R^2, wenn y_i durch z_i oder x_i durch w_i umkodiert wird?
a	Das R^2 steigt, wenn nur die erklärende Variable umkodiert wird.
b	Das R^2 steigt, wenn nur die abhängige Variable umkodiert wird.
c	**X** Das R^2 steigt nicht, wenn eine der Variablen umkodiert wird.
d	Das R^2 steigt nur, wenn beide Variablen umkodiert werden.

2.	Das angepasste $\bar{R}^2$
a	**X** kann bei Aufnahme zusätzlicher erklärender Variablen sinken.
b	gibt den Anteil der Variation in y an, die durch die x-Variablen erklärt wird.
c	kann Werte größer 1 annehmen.
d	nimmt Werte zwischen 0 und 1 an.

3.	Gegeben sei folgendes Modell: $income_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$. In welchem der folgenden Modelle ist dieses Modell genestet?
a	$log(income_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$
b	$hwage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$
c	$income_i = \beta_0 + \beta_1 educ_i + \beta_2 educ_i^2 + u_i$
d	**X** $income_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 female_i + u_i$

4.	Durch die Aufnahme eines irrelevanten Regressors in ein multiples lineares Regressionsmodell
a	werden die KQ-Schätzer verzerrt.
b	**X** sinkt die Präzision der Schätzung.
c	verändern sich die Regressionsergebnisse nicht.
d	verlieren unverzerrte Schätzer ihre kausale Interpretierbarkeit.

5.	Für das Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ testen Sie $H_0 : \beta_1 = 3$ vs. $H_1 : \beta_1 \neq 3$. Auf einem Signifikanzniveau von 10% führt eine Teststatistik von $1,70$
a	bei $n = 30$ zur Ablehnung von H_0.
b	bei $n = 32$ nicht zur Ablehnung von H_0.
c	**X** bei $n = 31$ nicht zur Ablehnung von H_0.
d	bei $n = 33$ nicht zur Ablehnung von H_0.

6.	Im linearen Wahrscheinlichkeitsmodell beschreibt $\hat{y}$ die Wahrscheinlichkeit, dass
a	y=0.
b	**X** y=1.
c	y=1%.
d	x=1.

7.	Der Standardfehler von $\hat{\beta}_j$
a	steigt bei einer präziseren Schätzung.
b	ist keine Zufallsvariable.
c	ist ein Schätzer für die Varianz σ^2.
d	**X** steigt mit steigender Residuenquadratsumme.

8.	Gegeben sei das Regressionsmodell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$. Wenn $\beta_1 > E(\hat{\beta}_1)$ gilt, ist β_1
a	effizient geschätzt.
b	unverzerrt geschätzt.
c	überschätzt.
d	**X** unterschätzt.

9.	Bei einem F-Test mit der Hypothese $\beta_3 = 0$
a	kann die Teststatistik sowohl positiv als auch negativ sein.
b	ist die Anzahl der Zählerfreiheitsgrade größer als die Anzahl der Nennerfreiheitsgrade.
c	kann das Bestimmtheitsmaß R^2 nicht zur Berechnung der Teststatistik verwendet werden.
d	**X** entspricht die Teststatistik dem quadrierten t-Wert des geschätzten Koeffizienten.

10.	Für die Schätzgleichung $\hat{y}_i = 1,5 - 0,5 \cdot x_i$ beträgt das Residuum für die Beobachtung $(x_1, y_1) = (1,3)$
a	0.
b	1.
c	**X** 2.
d	4.

11.	Typ-I-Fehlerwahrscheinlichkeiten
a	sind für F-Tests irrelevant.
b	werden durch t-Werte beschrieben.
c	**X** beschreiben die Wahrscheinlichkeit, die wahre Nullhypothese zu verwerfen.
d	werden durch die Aufnahme zusätzlicher Kontrollvariablen verringert.

12.	Unter den Gauss-Markov Annahmen
a	ist KQ das beste, erwartungstreue, nichtlineare Schätzverfahren.
b	ist die Varianz der Störterme heteroskedastisch.
c	folgt der KQ-Schätzer der t-Verteilung.
d	**X** ist der KQ-Schätzer konsistent.

13.	Sie schätzen das Modell $log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 log(x_{2i}) + \beta_3 exp(x_{3i}) + u_i$. Welcher Koeffizient gibt eine Semielastizität an?
a	$\hat{\beta}_0$
b	**X** $\hat{\beta}_1$
c	$\hat{\beta}_2$
d	$\hat{\beta}_3$

14.	In einer quadratischen Spezifikation $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ hat der Graph der Beziehung zwischen x und $\hat{y}$ die Form eines umgedrehten U, wenn
a	**X** $\hat{\beta}_1 > 0$ 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 $\hat{\beta}_2 = 0$.

15.	Wenn $Var(X) = 25$ dann ist $Var(5X) =$
a	5.
b	25.
c	125.
d	**X** 625.

16.	Bei kleiner Stichprobengröße führen χ^2-verteilte Störterme zu
a	einem zu hoch geschätzten Achsenabschnittsparameter.
b	verzerrten Steigungsparametern.
c	einer inkonsistenten Schätzung.
d	**X** falschen Verteilungsannahmen bei t-Tests.

17.	Welche der folgenden Informationen lässt sich mit einer binären Variablen vollständig abbilden?
a	Das Jahr der Einschulung der Person i.
b	Das Bundesland, in dem Person i eingeschult wurde.
c	**X** Das Geschlecht der Person i.
d	Die Farbe der Schultüte der Person i.

18.	Für das Modell $log(Lohn_i) = \beta_0 + \beta_1 Alter_i + \beta_2 Alter_i^2 + u_i$ schätzen Sie folgende statistisch signifikante Koeffizienten: $\hat{\beta}_0 = 6,73$; $\hat{\beta}_1 = 0,055$; $\hat{\beta}_2 = -0,0006$. Der marginale Effekt des Alters auf den Lohn
a	beträgt $5,5\%$.
b	beträgt $5,5\% - 0,06\% \cdot Alter_i$.
c	**X** ist negativ bei Alter 50.
d	ist positiv bis Alter 92.

19.	Eine erklärende Variable wird als endogen bezeichnet, wenn sie korreliert mit
a	der abhängigen Variable.
b	anderen unabhängigen Variablen.
c	**X** dem Störterm.
d	der Konstante.

20.	Im Modell $y_i = \beta_0 + \beta_1 x_1 + u_i$ führen Sie einen t-Test und einen F-Test auf Signifikanz von $\hat{\beta}_1$ durch. Welche Aussage trifft zu?
a	Die Wahrscheinlichkeit eines Typ 1-Fehlers ist beim F-Test niedriger.
b	Die Wahrscheinlichkeit eines Typ 2-Fehlers ist beim F-Test niedriger.
c	Der kritische Wert beim F-Test wird für 2 Zählerfreiheitsgrade bestimmt.
d	**X** Der p-Wert der F-Statistik gibt eine zuverlässige Aussage über die Signifikanz von $\hat{\beta}_1$.

21.	Wenn die abhängige Variable eines linearen Regressionsmodells binär ist,
a	nennt man das Modell dichotom.
b	können die geschätzten Koeffizienten nicht kausal interpretiert werden.
c	**X** liegt Heteroskedastie vor.
d	ist es sinnvoll, ohne Konstante zu schätzen.

22.	Wenn Sie in einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ die Variable y_i durch $\frac{1}{10} \cdot y_i$ ersetzen,
a	**X** verringert sich der Wert der Residuenquadratsumme um den Faktor 100.
b	verzehnfacht sich der Wert von $\hat{\beta}_1$.
c	sinkt der Wert des R^2 um den Faktor 10.
d	steigt der Wert von $se(\hat{\beta}_0)$ auf das Zehnfache.

23.	Ein größeres R^2 impliziert, dass die beobachteten Werte näher an
a	**X** der Regressionsgerade liegen.
b	dem Mittelwert der erklärenden Variablen liegen.
c	dem Mittelwert der abhängigen Variablen liegen.
d	dem Ursprung liegen.

24.	Sie schätzen das Modell $log(Lohn_i) = \beta_0 + \beta_1 \cdot Frau_i + u_i$. Als Ergebnis erhalten Sie $\hat{\beta}_0 = 7,95$ und $\hat{\beta}_1 = -0,29$. Was ist der auf eine Nachkommastelle exakt geschätzte prozentuale Lohnunterschied zwischen Männern und Frauen?
a	0,3%
b	**X** 25,2%
c	29,0%
d	33,6%

25.	Die Zufallsvariablen X und Y haben Mittelwerte $\bar{X} = 3$ und $\bar{Y} = 9$. Der Korrelationskoeffizient von X und Y, ρ_{XY}, ist 1. Welchen Wert hat y_i, wenn $x_i = 4$?
a	10
b	12
c	16
d	**X** Die gegebenen Informationen reichen nicht aus, um den Wert eindeutig zu berechnen.

26.	Die Nicht-Berücksichtigung einer relevanten Variable in einem multiplen linearen Regressionsmodell führt zu
a	**X** Fehlspezifikation des Modells.
b	Homoskedastie.
c	Dummy-Variable-Trap.
d	Multikollinearität.

27.	Eine t-verteilte Zufallsvariable
a	konvergiert bei steigender Stichprobe gegen die χ^2-Verteilung.
b	folgt für positive Werte einer F-Verteilung.
c	entsteht durch die Summierung mehrerer standardnormalverteilter Variablen.
d	**X** hat für große Stichproben mit hoher Wahrscheinlichkeit einen Wert nahe 0.

28.	Wenn das angepasste Bestimmtheitsmaß $\overline{R}^2$ den Wert 0,49 annimmt, dann
a	wird 49% der Variation der erklärenden Variable erklärt.
b	**X** ist das Bestimmtheitsmaß $R^2 > 0,49$.
c	wird 51% der Variation in der abhängigen Variablen nicht erklärt.
d	gilt für das Bestimmtheitsmaß $R^2 = 0,49 \cdot \frac{(n-1)}{(n-k-1)}$.

29.	Ein Typ-2 Fehler beschreibt die Wahrscheinlichkeit,
a	eine wahre Nullhypothese abzulehnen.
b	**X** eine falsche Nullhypothese nicht abzulehnen.
c	eine wahre Nullhypothese nicht abzulehnen.
d	eine falsche Nullhypothese abzulehnen.

30.	Interaktionsterme zwischen zwei Variablen
a	schließen bei einer Dummy-Variable und einer stetigen Variable konstante partielle Effekte aus.
b	**X** erlauben bei zwei Dummy-Variablen die Bestimmung unterschiedlicher Achsenabschnitte.
c	generieren bei zwei stetigen Variablen stets konstante partielle Effekte.
d	erlauben bei zwei Dummy-Variablen die Bestimmung unterschiedlicher Steigungsparameter.

31.	Die Schätzung des Modells $einkommen_i = \beta_0 + \beta_1 berufserfahrung_i + u_i$ ergibt bei 188 Beobachtungen $\hat{\beta}_1 = 4,8$ und $se(\hat{\beta}_1) = 2,4$. Welche Aussage trifft für die Schätzung von β_1 zu?
a	$\hat{\beta}_1$ ist auf dem 5%-Niveau signifikant kleiner als Null.
b	Der kritische Wert für einen rechtsseitigen t-Test mit $\alpha = 0,05$ beträgt $1,96$.
c	Die Teststatistik für den linksseitigen t-Test mit H_0: $\beta_1 \geq 0$ lautet -2.
d	**X** Der p-Wert für den Test der H_0: $\beta_1 = 0$ ist kleiner als $0,05$.

32.	Wenn für den Schätzer $\hat{\theta}$ gilt: $\lim_{n\to\infty} P(\lvert\hat{\theta} - \theta\rvert > \varepsilon) > 0$, dann ist $\hat{\theta}$
a	unverzerrt.
b	verzerrt.
c	konsistent.
d	**X** inkonsistent.

33.	Für ein Modell mit Konstante, 2 Steigungsparametern und 228 Beobachtungen schätzen Sie eine Residuenquadratsumme von 50.625. Welchen Wert hat die geschätzte Standardabweichung des Störterms?
a	50
b	**X** 15
c	225
d	1

34.	Welche Aussage trifft für die Schätzung des Modells $Lohn_i = \beta_0 + \beta_1 abitur_i + \beta_2 frau_i + \beta_3 abitur_i \cdot frau_i + u_i$ zu? Bei $abitur_i$ und $frau_i$ handelt es sich um Dummy-Variablen.
a	Der marginale Effekt von $abitur_i$ für Männer lautet: $\beta_1 + \beta_3$.
b	Die Nullhypothese beim Chow-Test auf Geschlechterunterschiede lautet: $\beta_2 = \beta_3 \geq 0$.
c	**X** Der vorhergesagte Lohn für Frauen ohne Abitur lautet: $\beta_0 + \beta_2$.
d	β_2 misst den Lohnunterschied von Männern und Frauen mit Abitur.

35.	Ein Konfidenzintervall
a	hat stets positive Intervallgrenzen, wenn $\hat{\beta}_j > 0$.
b	**X** wird mit steigendem α schmaler.
c	wird mit steigender Präzision breiter.
d	wird mit sinkender Stichprobengröße schmaler.

36.	Querschnittsdaten
a	enthalten wiederholte Messungen für jede Beobachtungseinheit.
b	beschreiben die Entwicklung einzelner Variablen über die Zeit.
c	werden auch als Paneldaten bezeichnet.
d	**X** enthalten mehr Beobachtungseinheiten als Erhebungszeitpunkte.

37.	Im Regressionsmodell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ wird x_2 durch $5 \cdot x_2$ ersetzt. Welche Aussage trifft zu?
a	**X** Der Standardfehler von $\hat{\beta}_2$ sinkt um den Faktor 5.
b	$\hat{\beta}_1$ und $\hat{\beta}_2$ sinken um den Faktor 5.
c	Die t-Statistik des Signifikanztests von $\hat{\beta}_2$ steigt um den Faktor 5.
d	Die p-Werte von $\hat{\beta}_1$ und $\hat{\beta}_2$ steigen um den Faktor 5.

38.	Was erhöht die Effizienz einer Schätzung?
a	Kleinere Varianz in den erklärenden Variablen.
b	Größere Varianz in irrelevanten erklärenden Variablen.
c	**X** Zusätzliche Beobachtungen.
d	Teilen der abhängigen Variablen durch 10.

39.	Sie schätzen das Modell $uni_i = \beta_0 + \beta_1 elternbild_i + \beta_2 frau_i + \beta_3 abitur_i + u_i$, wobei uni_i, $abitur_i$ und $frau_i$ Dummy-Variablen sind. Sie wollen wissen, ob sich der Zusammenhang zwischen Migranten und Nichtmigranten unterscheidet. Was ist der kritische Wert beim Chow-Test am 5 %-Niveau bei einem vollständig interagierten Modell mit 37 Beobachtungen?
a	2,92
b	2,60
c	**X** 2,70
d	2,95

40.	Schätzt man ein einfaches lineares Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$ mit der KQ-Methode, so
a	geht die geschätzte Regressionsgerade stets durch den Ursprung.
b	wird die Summe der vertikalen Abstände der Datenpunkte zur geschätzten Regressionsgerade minimiert.
c	liegen die Residuen stets unterhalb der geschätzten Regressionsgerade.
d	**X** liegt der Punkt $(\bar{x}, \bar{y})$ auf der geschätzten Regressionsgerade.