Bachelorprüfung SS 2014 - MUSTERLÖSUNG

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen. Es wird nur der Lösungsbogen eingesammelt.
Bewertung:	Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
Erlaubte Hilfsmittel:	• Formelsammlung (ist der Klausur beigefügt) • Tabellen der statistischen Verteilungen (sind der Klausur beigefügt) • Taschenrechner • Fremdwörterbuch
Wichtige Hinweise:	• Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert. • Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1: **[14 Punkte]**

Sie interessieren sich für Siegwahrscheinlichkeiten im Fußball und haben Daten zu 64 Länderspielen der deutschen Fußballnationalmannschaft mit folgenden Informationen gesammelt:

$Sieg_i$	$= 1$, wenn deutsche Mannschaft das Spiel i gewonnen hat; $= 0$ sonst
$Alter_i$	Durchschnittsalter der deutschen Mannschaft bei Spiel i
$Alter2_i$	Quadriertes Durchschnittsalter der deutschen Mannschaft bei Spiel i
$Fans_i$	Anteil Fans der deutschen Mannschaft an allen Zuschauern im Stadion bei Spiel i (in Prozent)
$Heimspiel_i$	$= 1$, wenn Spiel i in Deutschland stattfindet, $= 0$ sonst

Sie stellen folgendes lineares Regressionsmodell auf und schätzen dieses anschließend mit SPSS:

$$Sieg_i = \beta_0 + \beta_1 Alter_i + \beta_2 Alter2_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	-0,921	0,350	-2,634	0,006
$Alter$	0,147	0,034	4,262	0,000
$Alter2$	-0,003	0,001	-2,840	0,003

a. Abhängige Variable: *Sieg*

a) Berechnen Sie den marginalen Effekt des Alters auf die Siegwahrscheinlichkeit. Bei welchem Durchschnittsalter der Mannschaft wird die Siegwahrscheinlichkeit maximiert? Wie hoch ist die maximal durch Wahl des Alters erreichbare erwartete Siegwahrscheinlichkeit? (4 Punkte)

- Berechnung marginaler Effekt: $\frac{\Delta \widehat{Sieg_i}}{\Delta Alter_i} = \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 Alter_i = 0,147 - 2 \cdot 0,003 \cdot Alter_i$ [1P]
- $0,147 - 2 \cdot 0,003 Alter_i = 0 \Leftrightarrow Alter_i = 24,5$ [1P]
- Die Siegwahrscheinlichkeit wird maximal für ein Durchschnittsalter der Mannschaft von 24,5 Jahren.
- $\widehat{Sieg}_{max} = -0,921 + 0,147 \cdot 24,5 - 0,003 \cdot (24,5)^2 = 0,88$. [1P]
- Die maximal erwartete Siegwahrscheinlichkeit beträgt 88%. [1P]

b) Welche Werte würden $\hat{\beta}_1$ und $\hat{\beta}_2$ annehmen, wenn das Durchschnittsalter der Mannschaft nicht in Jahren, sondern in Monaten gemessen worden wäre? (2 Punkte)

- Umskalierung der unabhängigen Variable $Alter$:
- $\widehat{Sieg_i} = \hat{\beta}_0 + \hat{\beta}_1 Alter_i + \hat{\beta}_2 Alter2_i \Leftrightarrow \widehat{Sieg_i} = \hat{\beta}_0 + \frac{\hat{\beta}_1}{12} \cdot 12 \cdot Alter_i + \frac{\hat{\beta}_2}{144} \cdot 144 \cdot Alter2_i$
- $\tilde{\beta}_1 = \frac{\hat{\beta}_1}{12} = \frac{0,147}{12} = 0,01225$ [1P]
- $\tilde{\beta}_2 = \frac{\hat{\beta}_2}{144} = \frac{-0,003}{144} = -0,000021$ [1P]

Sie erweitern das Modell um die Variablen $Fans_i$ und $Heimspiel_i$:

$$Sieg_i = \beta_0 + \beta_1 Alter_i + \beta_2 Alter2_i + \beta_3 Fans_i + \beta_4 Heimspiel_i + u_i$$

c) SPSS berechnet [0,009;0,013] als 95%-Konfidenzintervall für β_3. Berechnen und interpretieren Sie inhaltlich den Effekt von einem Anstieg des Anteils der Fans der deutschen Mannschaft um 0,5 Prozentpunkte auf die Siegwahrscheinlichkeit der Mannschaft. Ist dieser Effekt statistisch signifikant von Null verschieden auf dem 5% Signifikanzniveau? (3 Punkte)

- Berechnung $\hat{\beta}_3$: $\frac{0{,}013+0{,}009}{2} = 0{,}011$ [1P]
- Inhaltliche Interpretation: Steigt der Anteil der eigenen Fans im Stadion um 0,5 Prozentpunkte, so erhöht sich c.p. im Durchschnitt die Siegwahrscheinlichkeit um $0{,}011 \cdot 0{,}5 \cdot 100 = 0{,}55$ Prozentpunkte. [1P]
- Da das Konfidenzintervall die Null nicht beinhaltet, ist der Effekt statistisch signifikant von Null verschieden auf dem 5%-Niveau. [1P]

d) Stellen Sie die Modellgleichung auf, mit welcher Sie testen können, ob sich sowohl der Alterseffekt als auch der Faneffekt zwischen Heim- und Auswärtsspielen unterscheidet. (2 Punkte)

- $Sieg_i = \beta_0 + \beta_1 Alter_i + \beta_2 Alter2_i + \beta_3 Fans_i + \beta_4 Heimspiel_i + \delta_1 Heimspiel_i \cdot Alter_i + \delta_2 Heimspiel_i \cdot Alter2_i + \delta_3 Heimspiel_i \cdot Fans_i + u_i$ [2P]

e) Sie vermuten, dass der Anteil der Fans der gegnerischen Mannschaft im Stadion während des Spiels ($Fans\_Gegner_i$) einen Einfluss auf die Siegwahrscheinlichkeit hat und nehmen die Variable in das Modell auf. Welches Problem tritt bei der Schätzung auf? Wie lässt es sich lösen? Begründen Sie Ihre Antwort knapp. (3 Punkte)
Hinweis: Nehmen Sie an, dass nur Fans der beiden Mannschaften anwesend sind.

- Es besteht das Problem der perfekten Multikollinearität. [1P]
- Die Variablen $Fans_i$ und $Fans\_Gegner_i$ sowie die Konstante sind linear abhängig [1P]; es ist nicht möglich, beide Variablen und die Konstante gleichzeitig in die Regression einzuschließen.
- Lösung: Entweder die Konstante oder $Fans_i$ oder $Fans\_Gegner_i$ aus dem Modell nehmen. [1P]

Aufgabe 2: **[11 Punkte]**

Sie interessieren sich für die Determinanten von Freibadbesuchen und befragen hierzu 3900 Grundschüler in Deutschland. Ihr Datensatz enthält folgende Informationen:

$Freibad_i$	Anzahl an Freibadbesuchen pro Woche von SchülerIn i
$Junge_i$	=1, wenn Schüler; =0, wenn Schülerin
$Hausaufgaben_i$	Wöchentliche Anzahl an Stunden, die SchülerIn i für Hausaufgaben aufwendet
$Klasse12_i$	=1, wenn SchülerIn i in Klasse 1 oder 2; =0, wenn SchülerIn i in Klasse 3 oder 4

Sie schätzen folgendes lineares Regressionsmodell mit SPSS:

$$Freibad_i = \beta_0 + \beta_1 Junge_i + \beta_2 Hausaufgaben_i + \beta_3 Klasse12_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	2,064	0,023	90,06	0,000
$Junge$	0,075	0,012	6,26	0,000
$Hausaufgaben$	-0,491	0,213	-2,303	0,011
$Klasse12$	-0,169	0,012	-13,89	0,000

a. Abhängige Variable: *Freibad*

a) Interpretieren Sie den Koeffizienten von $Klasse12$ sowohl inhaltlich als auch statistisch. (2 Punkte)

- Ein(e) Grundschüler(in) der 1. oder 2. Klasse besucht das Freibad c.p. im Durchschnitt 0,169 mal weniger pro Woche als ein(e) Grundschüler(in) aus der 3. oder 4. Klasse. [1P]
- Der Koeffizient ist auf dem 1%-Signifikanzniveau statistisch signifikant von Null verschieden. (p-Wert: 0,000<0,01) [1P]

b) Testen Sie auf dem 1%-Signifikanzniveau, ob der Unterschied in der Freibadnutzung zwischen Jungen und Mädchen kleiner als 0,1 ist. Geben Sie Testverfahren, Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

- Testverfahren: einseitiger t-Test [0,5P]
- Hypothesen: $H_0 : \beta_1 \geq 0,1, H_1 : \beta_1 < 0,1$ [1P]
- Teststatistik: $t = \frac{\hat{\beta}_1 - 0,1}{se(\hat{\beta}_1)} = \frac{0,075-0,1}{0,012} = -2,083$ [1P]
- Kritischer Wert: $t_{\alpha;n-k-1} = t_{0,01;3900-3-1} = -2,326$ [1P]
- Testentscheidung: Da $t_{empirisch} = -2,083 > -2,326 = t_{kritisch}$ kann die Nullhypothese auf dem 1% Niveau nicht verworfen werden, [1P] d.h. der Unterschied ist wohl nicht kleiner als $0,1$.

c) Angela geht in die 4. Klasse und wendet 4 Stunden pro Woche für Hausaufgaben auf, Michael geht in die 2. Klasse und wendet eine Stunde pro Woche für Hausaufgaben auf. Wieviel Zeit müsste Angela nach obiger Schätzung weniger für Hausaufgaben aufwenden, damit die erwartete wöchentliche Freibadnutzung für beide Kinder exakt übereinstimmt? (4,5 Punkte)

- Unterschied in der Freibadnutzung zwischen Angela und Michael:
  $\Delta Freibad = (2,064 + 0,075 - 0,491 - 0,169) - (2,064 - 0,491 \cdot 4) = 1,379$ [2P]
  $\Rightarrow$ Bei gegebenen Charakteristika besucht Michael das Freibad 1,379 mal pro Woche mehr als Angela.
- Ausgleich durch Hausaufgaben von Angela:
  $\Delta Freibad = \hat{\beta}_2 \Delta Hausaufgaben \Leftrightarrow \Delta Hausaufgaben = \frac{1,379}{-0,491} = -2,809$ [2P]
  $\Rightarrow$ Angela müsste 2,8 h/Woche weniger Zeit für Hausaufgaben verwenden, um genauso oft das Freibad zu nutzen wie Michael. [0,5P]

Aufgabe 3: **[15 Punkte]**

Sie interessieren sich dafür, ob die Schulklassengröße einen Einfluss auf die Löhne im Erwachsenenalter hat. Sie verfügen über einen US-amerikanischen Datensatz mit den folgenden Variablen für 767 Schüler:

$lohn_i$	Stundenlohn von Person i im Alter 27 (in US$)
$klein_i$	=1, wenn Person i in kleiner Klasse war ($<$ 17 Schüler); =0 sonst
$frau_i$	=1, wenn Frau; =0 sonst
$eltern_i$	Einkommen der Eltern von Person i (in 1000 US$)

Sie schätzen folgendes lineares Regressionsmodell mit SPSS:

$$log(lohn_i) = \beta_0 + \beta_1 klein_i + \beta_2 frau_i + \beta_3 eltern_i + u_i$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	6,013	0,088	68,32	0,000
$klein$	0,049	0,006	8,16	0,000
$frau$	-0,206	0,042	4,90	0,000
$eltern$	0,013	???	???	???

a. Abhängige Variable: $log(lohn)$

Das R^2 der Schätzung beträgt $0,245$.

a) Interpretieren Sie den geschätzten Koeffizienten von $klein$ sowohl inhaltlich als auch statistisch. (2 Punkte)

- Der Stundenlohn einer Person, die in einer kleinen Klasse war, ist c.p. im Mittel ca. 4,9% höher als der Stundenlohn einer Person, die nicht in einer kleinen Klasse war. [1P]
- Der Koeffizient ist auf dem 1%-Signifikanzniveau statistisch signifikant von Null verschieden. [1P]

b) Berechnen und interpretieren Sie inhaltlich den genauen (!) Effekt der Variable $frau$ auf den Stundenlohn. Runden Sie auf die 3. Nachkommastelle. (2 Punkte)

- $\hat{\beta}_2 = -0,206$
- Der genaue Effekt beträgt $e^{\hat{\beta}_2} - 1 = -0,186$. [1P]
- Eine Frau verdient c.p. im Mittel $18,6$ Prozent weniger pro Stunde als ein Mann. [1P]

c) Wie können Sie bestimmen, ob die Variable $eltern$ einen signifikanten Erklärungsbeitrag in Ihrem Modell liefert? Führen Sie einen entsprechenden Test am 1%-Niveau durch, indem Sie Nullhypothese, Alternativhypothese, Freiheitsgrade, Teststatistik, kritischen Wert und Testentscheidung angeben. Leistet die Variable einen signifikanten Erklärungsbeitrag? (Hinweis: das R^2 des Modells $log(lohn_i) = \beta_0 + \beta_1 klein_i + \beta_2 frau_i + u_i$ hat den Wert 0,181.) (6 Punkte)

1. **Nullhypothese:** $H_0 : \beta_3 = 0$ [0,5P]
2. **Alternativhypothese:** $H_1 : \beta_3 \neq 0$ [0,5P]
3. **Freiheitsgrade (aus unrestringiertem Modell):** $n-k-1 = 767-3-1 = 763$ [0,5P] und $q = 1$ [0,5P]
4. **Teststatistik:** $F_{empirisch} = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)} = \frac{(0{,}245-0{,}181)/1}{(1-0{,}245)/(767-3-1)} = \frac{0{,}064}{0{,}00098952} = 64{,}68$ [1P]
5. **Kritischer Wert:** $F_{kritisch} = F_{0{,}01;1{,}763} = 6{,}63$ [1P]
6. **Testentscheidung:** Da $F_{empirisch} > F_{kritisch}$ wird die Nullhypothese auf dem 1% Niveau verworfen. [1P] Die Variable *eltern* leistet einen signifikant von Null verschiedenen Erklärungsbeitrag. [1P]

d) Nennen Sie eine ausgelassene Variable, die den Koeffizienten $\hat{\beta}_1$ verzerren könnte. Stellen Sie das vollständige Modell auf, begründen und bestimmen Sie die Richtung der Verzerrung ausführlich und geben Sie an, ob β_1 über- oder unterschätzt wird. (5 Punkte).

- Beispiel: Fähigkeit des Schülers (*ability*) [0,5P].
- Wahres Modell: $log(lohn_i) = \beta_0 + \beta_1 klein_i + \beta_2 frau_i + \beta_3 eltern_i + \beta_4 ability_i + u_i$ [1P]
- Das Weglassen der Variable könnte $\hat{\beta}_1$ verzerren, weil individuelle Fähigkeiten mit Löhnen [0,5P] und [0,5P] mit Klassengröße [0,5P] korreliert sein könnten (z.B. wenn die fähigeren Schüler in kleine Klassen eingeteilt werden).
- Die Verzerrung ist positiv [0,5P], wenn $cov(klein_i, ability_i) > 0$ [0,5P] und $\beta_4 > 0$ [0,5P], d.h. β_1 würde überschätzt. [0,5 P]
- (Auch möglich: Die Verzerrung ist negativ [0,5P], wenn $cov(klein_i, ability_i) < 0$ [0,5P] und $\beta_4 > 0$ [0,5P] (oder umgekehrt), d.h. β_1 würde unterschätzt. [0,5P])

Aufgabe 4: **[10 Punkte]**

a) Leiten Sie formal für das Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ den KQ-Schätzer $\hat{\beta}_0$ her. (4 Punkte)

- $min \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ [1P]
- Nach $\hat{\beta}_0$ ableiten und gleich Null setzen: $2 \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$ [1P]
- Auflösen ergibt $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ [2P]

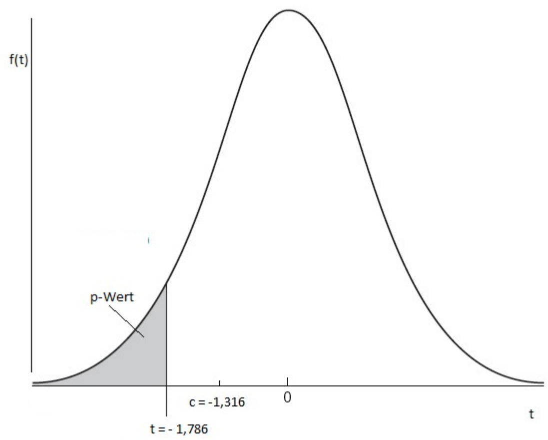
b) Nennen Sie zwei Charakteristika eines konsistenten Schätzers. (2 Punkte)

Für einen konsistenten Schätzer gilt, dass mit steigender Stichprobengröße

- die Varianz des Schätzers sinkt. [1P]
- der Schätzer gegen den wahren Wert konvergiert. [1P]

c) Die Hypothesen für einen t-Test für den Parameter β lauten: $H_0 : \beta \geq 0$ und $H_1 : \beta < 0$, mit α=0,1 und $n-k-1 = 25$. Der geschätzte Parameter $\hat{\beta}$ beträgt $-0{,}125$ und der dazu gehörige Standardfehler beträgt

$se(\hat{\beta}) = 0,070$. Skizzieren Sie die dazu passende Verteilungsfunktion der t-Statistik und kennzeichnen Sie in der Abbildung den p-Wert sowie den von Ihnen ermittelten kritischen und empirischen Wert. Beschriften Sie zudem die Achsen. (4 Punkte)



- Teststatistik: $t = \frac{-0,125}{0,070} = -1,786$. [1P]
- Kritischer Wert: $-c_{0,10;25} = -1,316$. [1P]
- Je 0,5P für korrekte Achsenbeschriftung, 0,5P je für Kennzeichnung von p-Wert und kritischem Wert.

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.	Eine KQ-Schätzung liefert $\hat{y}_i = 4,9 - 1,4x_{1i} + 3,1x_{2i}$. Welchen Wert hat das Residuum für die Beobachtung $(y_1; x_{11}; x_{21}) = (16; -1; 4)$?
a	0,1.
b	3,4.
c	**X** -2,7.
d	-1,8.

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

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

4.	Sie betrachten zwei Zufallsvariablen X und Y. Welche der folgenden Aussagen ist korrekt?
a	Sind X und Y positiv korreliert, so kann die Kovarianz von X und Y sowohl positiv als auch negativ sein.
b	**X** Sind X und Y unabhängig, so ist der Korrelationskoeffizient gleich Null.
c	Sind X und Y unkorreliert, so sind X und Y auch unabhängig.
d	Sind X und Y unabhängig, so ist der Erwartungswert von X und der Erwartungswert von Y gleich Null.

5.	In einem einfachen linearen Regressionsmodell $y_i = \beta_0 + \beta_1 x_i + u_i$
a	ist β_1 positiv, wenn $corr(x,y) < 0$.
b	ist β_1 negativ, wenn $corr(x,y) > 0$.
c	**X** ist β_1 negativ, wenn $corr(x,y) < 0$.
d	Keine der Antworten ist korrekt.

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

7.	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	werden die Abstände der Datenpunkte zur geschätzten Regressionsgerade minimiert.
c	ist die Summe der Residuen ungleich 0.
d	**X** liegt der Punkt $(\bar{y}, \bar{x})$ auf der geschätzten Regressionsgerade.

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

9.	Man testet, ob ein Steigungsparameter signifikant von -9 verschieden ist, indem man
a	überprüft, ob das Konfidenzintervall den Wert 9 einschließt oder nicht.
b	**X** überprüft, ob das Konfidenzintervall den Wert -9 einschließt oder nicht.
c	den Wert 9 und den Schätzwert addiert und diesen sich ergeben Wert mit dem kritischen Wert der t-Verteilung vergleicht.
d	den Wert 9 vom Schätzwert subtrahiert und diesen sich ergeben Wert mit dem kritischen Wert der t-Verteilung vergleicht.

10.	Für das Modell $lohn_i = \beta_0 + \beta_1 alter_i + \beta_2 alter_i^2 + u_i$ ergibt eine KQ-Schätzung $\hat{\beta}_1 = 0,28$ und $\hat{\beta}_2 = -0,004$. In welchem Alter wird der durchschnittliche Lohn maximiert?
a	32.
b	**X** 35.
c	38.
d	40.

11.	Gegeben ist folgendes Modell: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$. Man testet $H_0 : \beta_2 \leq 2$ vs. $H_1 : \beta_2 > 2$. Bei einem Signifikanzniveau von 5% führt eine Teststatistik von $1,71$
a	bei $n = 20$ zur Ablehung von H_0.
b	**X** bei $n = 30$ zur Ablehung von H_0.
c	bei $n = 35$ nicht zur Ablehung von H_0.
d	bei $n = 40$ nicht zur Ablehung von H_0.

12.	Das Problem perfekter Multikollinearität
a	kann durch Vergrößerung der Stichprobe reduziert werden.
b	führt zu inkonsistenten Schätzergebnissen.
c	führt zu ineffizienten Schätzergebnissen.
d	**X** kann durch Auslassen der linear abhängigen Variablen gelöst werden.

13.	Sie beobachten das Gewicht für Männer und Frauen in Nord- und Südeuropa. Sie schätzen folgendes Modell: $Gewicht_i = \beta_1 Frau_i + \beta_2 Mann_i + \beta_3 Nord_i + \beta_4 Frau_i \cdot Nord_i + u_i$. Was misst der geschätzte Parameter für β_3?
a	**X** Den Mittelwertunterschied des Gewichtes für Männer zwischen Nord- und Südeuropa .
b	Den Mittelwertunterschied des Gewichtes zwischen Männern und Frauen aus Südeuropa.
c	Den Mittelwert des Gewichtes für Männer.
d	Den Mittelwert des Gewichtes für Männer aus Nordeuropa.

14.	Wenn eine diskrete Zufallsvariable X die Werte -3, 6, 5 und 2 mit konstanter Wahrscheinlichkeit $\frac{1}{4}$ annimmt, so gilt
a	$E(X) = 9$.
b	**X** $E(X) = \frac{10}{4}$.
c	$E(X^2) = 74$.
d	$E(X^2) = 10$.

15.	Wie hoch ist die Gesamtvariation SST über alle Beobachtungen gegeben $(y_1;x_1) = (3;3), (y_2;x_2) = (-1;4), (y_3;x_3) = (4;5)$?
a	0.
b	6.
c	**X** 14.
d	20.

16.	Der Standardfehler eines geschätzten Koeffizienten $se(\hat{\beta}_j)$
a	**X** ist eine Zufallsvariable mit Verteilungseigenschaften.
b	ist bei Vorliegen von normalverteilten Störtermen konsistent geschätzt.
c	kann bei hoher Multikollinearität nicht berechnet werden.
d	gibt an, wie stark die Beobachtungen um die Regressionsgerade streuen.

17.	Sie schätzen folgende Konsumfunktion mit KQ: $konsum_i = \beta_0 + \beta_1 log(einkommen_i) + u_i$. (*Hinweis*: *konsum* und *einkommen* sind in Euro gemessen.) $\hat{\beta}_1$ beträgt 400. Welche Aussage ist richtig?
a	Wenn das Einkommen um 1000 Euro ansteigt, steigt der Konsum um 400 Euro.
b	Wenn das Einkommen um 1 Euro ansteigt, steigt der Konsum um 0,4%.
c	**X** Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 4 Euro.
d	Wenn das Einkommen um 1% ansteigt, steigt der Konsum um 0,4%.

18.	Sie haben Daten über Löhne von Männern und Frauen aus Ost- und Westdeutschland. Bei welcher Modellspezifikation kommt es zum dummy variable trap?
a	$Lohn_i = \beta_0 + \beta_1 Mann_i + \beta_2(Mann_i \cdot Ost_i) + \beta_3(Frau_i \cdot Ost_i) + u_i$
b	$Lohn_i = \beta_0 + \beta_1(Mann_i \cdot Ost_i) + \beta_2(Frau_i \cdot Ost_i) + \beta_3(Mann_i \cdot West_i) + u_i$
c	$Lohn_i = \beta_1 Mann_i + \beta_2 Frau_i + \beta_3(Mann_i \cdot Ost_i) + u_i$
d	**X** $Lohn_i = \beta_1 Mann_i + \beta_2 Frau_i + \beta_3(Mann_i \cdot Ost_i) + \beta_4(Mann_i \cdot West_i) + u_i$

19.	Im linearen Wahrscheinlichkeitsmodell
a	**X** gibt die Prognose $\hat{y}$ die Wahrscheinlichkeit für das Ereignis $y = 1$ an.
b	wird die stetige abhängige Variable in Prozentpunkten gemessen.
c	werden nur binär kodierte Variablen als erklärende Variablen verwendet.
d	können kausale Effekte nicht identifiziert werden.

20.	Um die Präzision einer Schätzung zu erhöhen
a	sollte man ohne Konstante schätzen.
b	kann man auch irrelevante erklärenden Variablen im Modell verwenden.
c	sollte man die Spezifikation mit einem hohen Bestimmtheitsmaß wählen.
d	**X** sollte man erklärende Variablen mit hoher Variation verwenden.

21.	Im einfachen linearen Regressionsmodell berechnet sich der Steigungsparameter $\hat{\beta}_1$ als Verhältnis der
a	Kovarianz von x und y zur Standardabweichung von x.
b	Kovarianz von x und y zum Mittelwert von x.
c	**X** Kovarianz von x und y zur Varianz von x.
d	Kovarianz von x und y zum quadrierten Mittelwert von x.

22.	Der Standardfehler der Regression (SER) ist ein Schätzer für
a	die Streuung der einzelnen Steigungsparameter.
b	**X** die Streuung der auf x bedingten Störterme.
c	die Standardabweichung der Konstante.
d	Keine der Antworten ist korrekt.

23.	Wird ein irrelevanter Regressor (d.h. ein Regressor, dessen wahrer Koeffizient exakt gleich 0 ist) einem multiplen Regressionsmodell hinzugefügt, so
a	**X** sinkt das korrigierte Bestimmtheitsmaß $\bar{R}^2$.
b	verändern sich die Regressionsergebnisse nicht.
c	steigt die Präzision der Schätzung.
d	kann man diesen nutzen, um einen Test auf ausgelassene Variablen durchzuführen.

24.	Ein Typ 2-Fehler tritt auf, wenn man die Nullhypothese
a	verwirft, obwohl diese zutrifft.
b	verwirft, obwohl diese falsch ist.
c	**X** nicht verwirft, obwohl diese falsch ist.
d	nicht verwirft, obwohl diese zutrifft.

25.	Ein Schätzer $\hat{\beta}_1$ für den unbekannten Parameter β_1 im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ ist konsistent, wenn
a	$N \rightarrow \infty$.
b	**X** $plim(\hat{\beta}_1) = \beta_1$.
c	u normalverteilt ist.
d	$E[\hat{\beta}_1] = \beta_1$.

26.	Im Modell $log(y_i) = \beta_0 + \beta_1 log(x_{1i}) + \beta_2 x_{2i} + u_i$
a	**X** wird β_1 als Elastizität interpretiert.
b	wird β_2 als Elastizität interpretiert.
c	werden sowohl β_1 als auch β_2 als Elastizitäten interpretiert.
d	Keine der Antworten ist korrekt.

27.	Wird die Nullhypothese eines Chow-Tests abgelehnt, so
a	liegen keine signifikante Unterschiede in den Steigungsparametern verschiedener Gruppen vor.
b	liegt ein Endogenitätsproblem vor.
c	**X** fallen Regressionskoeffizienten für verschiedene Gruppen unterschiedlich aus.
d	müssen zwei verschiedene Regressionen geschätzt werden.

28.	Wird im Modell $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ mit $\beta_0 \neq 0$ die Konstante β_0 nicht mitgeschätzt, so
a	sinkt die Zahl der Freiheitsgrade.
b	**X** werden die Steigungsparameter verzerrt geschätzt.
c	ändern sich die Konfidenzintervalle der Steigungsparameter nicht.
d	kann man das R^2 als Maß der Schätzgüte verwenden.

29.	Eine Typ 1-Fehlerwahrscheinlichkeit beschreibt die Wahrscheinlichkeit,
a	**X** 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.

30.	Ein nicht-linearer Zusammenhang zwischen zwei Variablen
a	kann mit der Kovarianz bestimmt werden.
b	kann mit dem Korrelationskoeffizienten abgebildet werden.
c	kann nicht in einem multiplen Regressionsmodell abgebildet werden.
d	**X** Keine der Antworten ist korrekt.

31.	Wenn die Störterme einer multiplen Regression nicht normalverteilt sind, dann
a	ist der KQ-Schätzer inkonsistent.
b	sind t- und F-Tests auch in kleinen Stichproben gültig.
c	können keine Hypothesentests durchgeführt werden.
d	**X** kann der KQ-Schätzer dennoch unverzerrt sein.

32.	Das korrigierte Bestimmtheitsmaß $\bar{R}^2$ im einfachen linearen Regressionsmodell
a	nimmt ausschließlich negative Werte an.
b	**X** berücksichtigt die zur Schätzung benötigten Freiheitsgrade.
c	gibt das Verhältnis von unerklärter Variation zur Gesamtvariation an.
d	nimmt ausschließlich positive Werte an.

33.	Das Problem ausgelassener Variablen entsteht im Modell $y_i = \beta_0 + \beta_1 x_{1i} + u_i$, wenn eine Variable x_2
a	mit x_1 korreliert.
b	mit y korreliert.
c	**X** mit x_1 und u korreliert.
d	Keine der Antworten ist korrekt.

34.	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	$\beta_1 + \beta_3 x_2$
b	$\beta_1 x_1 + \beta_3 x_2$
c	$\frac{\beta_1}{x_1} + \beta_3 \frac{1}{x_1} x_2$
d	**X** $\beta_1 \frac{1}{x_1} + \beta_3 x_2$

35.	Bei konsistenten Schätzverfahren
a	ist der Erwartungswert des Schätzers mit dem wahren Wert identisch.
b	**X** liegt der Wahrscheinlichkeitsgrenzwert des Schätzers umso näher am wahren Wert, je größer die Stichprobe.
c	ist der Schätzer unverzerrt.
d	ist Multikollinearität bei steigender Stichprobengröße kein Problem.

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

37.	Zwei Koeffizienten sind gemeinsam signifikant, wenn
a	die 0 in beiden Konfidenzintervallen enthalten ist.
b	beide Koeffizienten einzeln signifikant sind.
c	mindestens einer der beiden Koeffizienten signifikant ist.
d	**X** Keine der Antworten ist korrekt.

38.	Gilt $E[u\|x] = 0$, dann
a	ist der Störterm gleich 0, unabhängig von x.
b	kann der KQ-Schätzer nicht kausal interpretiert werden.
c	**X** ist die ceteris paribus Interpretation angemessen.
d	können ausgelassene Variablen den KQ-Schätzer verzerren.

39.	F-Tests
a	**X** können verwendet werden, um Hypothesen für einzelne Parameter zu testen.
b	können bei mehr als 10 Nennerfreiheitsgraden nicht durchgeführt werden.
c	gelten nur bei normalverteilten Störtermen.
d	werden basierend auf einer t-Verteilung berechnet.

40.	Durch Umskalierung der abhängigen Variable
a	ändert sich der Wert des Bestimmtheitsmaßes.
b	ändert sich die Anzahl der verwendeten Freiheitsgrade.
c	können die p-Werte sinken.
d	**X** verändert sich das Konfidenzintervall der Konstante.