## Aufgabe 1: [11 Punkte]

Sie vermuten, dass das Niveau der Wohnungsmieten in Universitätsstädten unter anderem vom Anteil der Studierenden an den Einwohnern abhängt. Ihnen liegen Daten für 64 amerikanische Universitätsstädte vor, mit denen Sie folgendes Modell schätzen:

$$rent = \beta_0 + \beta_1 \cdot pctstu + \beta_2 \cdot avginc + u$$

rent = durchschnittlich zu zahlende Miete pro Wohnung in US-Dollars

pctstu = Anteil der Studierenden an den Einwohnern gemessen in Prozent (0-100%)

avginc = durchschnittliches Jahreseinkommen pro Einwohner in 1000 US-Dollars

SPSS liefert Ihnen folgenden Output:

**ANOVA(b)**

| Modell | | Quadratsumme | df | Mittel der Quadrate | F | Signifikanz |
|---|---|---|---|---|---|---|
| 1 | Regression | 378290,39 | 2 | 189145,199 | 47,389 | ,000(a) |
| | Residuen | 243472,04 | 61 | 3991,345 | | |
| | Gesamt | 621762,43 | 63 | | | |

a Einflußvariablen : (Konstante), avginc, pctstu
b Abhängige Variable: rent

**Koeffizienten(a)**

| Modell | | Nicht standardisierte Koeffizienten | | Standardisierte Koeffizienten | T | Signifikanz |
|---|---|---|---|---|---|---|
| | | B | Standardfehler | Beta | | |
| 1 | (Konstante) | 78,064 | 39,590 | | 1,972 | ,053 |
| | pctstu | 1,699 | ,608 | ,233 | 2,797 | ,007 |
| | avginc | 12,567 | 1,291 | ,811 | 9,734 | ,000 |

a Abhängige Variable: rent

a) Interpretieren Sie die Größe der Koeffizienten von pctstu und avginc. (2 Punkte)

*Steigt der Studentenanteil unter den Einwohnern einer Universitätsstadt um einen Prozentpunkt, so steigt die erwartete Durchschnittsmiete ceteris paribus um circa 1,70 \$.*

*Steigt das jährliche Durchschnittseinkommen pro Einwohner in einer Universitätsstadt um 1000\$, so erhöht sich die erwartete Durchschnittsmiete dort ceteris paribus um circa 12,57 \$.*

b) Berechnen Sie das 95% Konfidenzintervall für den Koeffizienten von avginc. (3 Punkte)

$$KI = [\hat{\beta}_2 - t_{\alpha/2,n-k-1} \cdot se(\hat{\beta}_2);\ \hat{\beta}_2 + t_{\alpha/2,n-k-1} \cdot se(\hat{\beta}_2)]$$
$$KI = [12{,}567 - 2{,}00 \cdot 1{,}291; 12{,}567 + 2{,}00 \cdot 1{,}291] = [9{,}985; 15{,}149]$$

c) Berechnen und interpretieren Sie den $R^2$ - Wert. (2 Punkte)

$$R^2 = \frac{378290{,}39}{621762{,}43} = 0{,}61$$

*Circa 61 % der Variation in rent wird durch das Modell erklärt.*

d) Wie ändern sich jeweils $\hat{\beta}_2$, $se(\hat{\beta}_2)$ und der p-Wert von $\hat{\beta}_2$, wenn das durchschnittliche Jahreseinkommen pro Einwohner in \$ statt in 1000 \$ gemessen wäre? Beeinflusst diese Transformation auch die Schätzgüte des Modells? (4 Punkte)

$$\tilde{\beta}_2 = 12{,}567 : 1000 = 0{,}012567 \ (1)$$
$$se(\tilde{\beta}_2) = 1{,}291 : 1000 = 0{,}001291 \ (1)$$

*p-Wert bleibt gleich, da t-Statistik =* $[\hat{\beta}_2 : 1000]/[se(\hat{\beta}_2) : 1000]$ *gleich bleibt.*
*Schätzgüte wird nicht beeinflusst. (* $R^2$ *bleibt gleich)*

**Aufgabe 2:** **[8 Punkte]**

Sie entschließen sich, die abhängige Variable *rent* in logarithmierter Form zu verwenden:

$$\log(rent) = \beta_0 + \beta_1 \cdot pctstu + \beta_2 \cdot avginc + u$$

Die Schätzung liefert folgende Ergebnisse:

**Koeffizienten(a)**

| Modell | | Nicht standardisierte Koeffizienten | | Standardisierte Koeffizienten | T | Signifikanz |
|---|---|---|---|---|---|---|
| | | B | Standardfehler | Beta | | |
| 1 | (Konstante) | 5,339 | ,085 | | 63,150 | ,000 |
| | pctstu | ,004 | ,001 | ,272 | 3,086 | ,003 |
| | avginc | ,024 | ,003 | ,776 | 8,806 | ,000 |

a Abhängige Variable: log(rent)

a) Interpretieren Sie erneut die Größe der Koeffizienten von *pctstu* und *avginc*. (2 Punkte)

*Steigt der Studentenanteil unter den Einwohnern einer Universitätsstadt um einen Prozentpunkt, so steigt die erwartete Durchschnittsmiete ceteris paribus um ca. 0,4%.*

*Steigt das jährliche Durchschnittseinkommen pro Einwohner in einer Universitätsstadt um 1000 \$, so erhöht sich die erwartete Durchschnittsmiete dort ceteris paribus um circa. 2,4%*

b) Testen Sie anhand eines t-Tests am 5% Signifikanzniveau, ob der Koeffizient von *avginc* signifikant größer Null ist. Geben Sie die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert der t-Verteilung, die Anzahl der Freiheitsgrade sowie die Testentscheidung an. (3 Punkte)

$$H_0 : \beta_2 \leq 0 \ ; \ H_1 : \beta_2 > 0$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0{,}024}{0{,}003} = 8{,}00$$

$$t_{0,05;64-2-1} = 1{,}671$$

*1,671 < 8,00*

*Der Koeffizient von avginc ist signifikant größer Null.*

c) Wie hoch ist laut diesem Modell die erwartete Durchschnittsmiete in einer Universitätsstadt in der 30% der Einwohner Studierende sind und das jährliche Durchschnittseinkommen pro Kopf 20 000 $ beträgt? (3 Punkte)
Hinweis: es gilt $\hat{\sigma}^2 = 0{,}02$

$$\hat{\log}(rent) = 5{,}339 + 0{,}004 \cdot 30 + 0{,}024 \cdot 20 = 5{,}939$$

$$\hat{rent} = \exp(\hat{\sigma}^2 / 2) \cdot \exp(\hat{\log}(rent)) = 379{,}63$$

*Die erwartete Durchschnittsmiete beträgt 379,63$.*

**Aufgabe 3:** **[13 Punkte]**

Sie möchten untersuchen, wovon das Rauchverhalten abhängt. Dazu betrachten Sie Daten zu 807 Individuen und schätzen das Modell:

$$cigs = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot male + \beta_3 \cdot age + \beta_4 \cdot age2 + \beta_5 \cdot log(income) + u$$

cigs = Anzahl gerauchter Zigaretten pro Tag

educ = Anzahl in Ausbildung verbrachter Jahre

male = 1 wenn männlich, = 0 wenn weiblich

age = Alter in Jahren

age2 = Alter in Jahren quadriert

log(income) = Logarithmiertes Jahreseinkommen in US-Dollars

SPSS liefert Ihnen folgenden Output:

**ANOVA(b)**

| Modell | | Quadratsumme | df | Mittel der Quadrate |
|---|---|---|---|---|
| 1 | Regression | 6800,243 | 5 | 1360,049 |
| | Residuen | 144953,440 | 801 | 180,966 |
| | Gesamt | 151753,683 | 806 | |

a Einflußvariablen : (Konstante), educ, male age, age2, log(income)
b Abhängige Variable: cigs

**Koeffizienten[a]**

| Modell | | Nicht standardisierte Koeffizienten | | Standardisierte Koeffizienten | T | Signifikanz |
|---|---|---|---|---|---|---|
| | | B | Standardfehler | Beta | | |
| 1 | (Konstante) | -5,889 | 6,828 | | -,862 | ,389 |
| | educ | -,516 | ,168 | -,115 | -3,076 | ,002 |
| | male | -,161 | 1,455 | -,004 | -,111 | ,912 |
| | age | ,782 | ,161 | ,970 | 4,855 | ,000 |
| | age2 | -,009 | ,002 | -1,048 | -5,202 | ,000 |
| | log(income) | ,730 | ,728 | ,038 | 1,002 | ,316 |

a. Abhängige Variable: cigs

a) Beantworten Sie die folgenden Fragen. (3 Punkte)

(1) Was versteht man allgemein unter irrelevanten Variablen?

*Eine erklärende Variable ist irrelevant, wenn ihr Bevölkerungsparameter gleich Null ist.*

(2) Welche Auswirkungen haben irrelevante Variablen für die Schätzergebnisse?

*Die Varianz der übrigen geschätzten Parameter steigt.*

(3) Beurteilen Sie anhand der vorliegenden Schätzergebnisse, ob *male* eine irrelevante Variable ist.

*Der Parameter von male ist insignifikant. Das deutet darauf hin, dass male eine irrelevante Variable ist.*

b) Testen Sie auf dem 5% Signifikanzniveau, ob das von Ihnen geschätzte Modell signifikant ist. Geben Sie an, welchen Test Sie durchführen, die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert sowie die Testentscheidung. (4 Punkte)

*F-Test auf Gesamtsignifikanz des Modells*
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
*$H_1$: mind. ein Koeffizient ≠ 0 (1P)*

$$F = \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)} = \frac{(151753{,}7 - 144953{,}4)/5}{144953{,}4/801} = \frac{1360{,}06}{180{,}97} = 7{,}52$$

$F_{0,05;5;801} = 2{,}21$

*$F > F_{krit}$ ⇨ $H_0$ muss auf dem 5% Signifikanzniveau verworfen werden. Das Modell ist statistisch signifikant.*

c) Sie überlegen, die Variable *log(mincome)* (logarithmierter Monatsverdienst in US-Dollars) zusätzlich in Ihr Modell aufzunehmen und

$$cigs = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot male + \beta_3 \cdot age + \beta_4 \cdot age2 + \beta_5 \cdot log(income) + \beta_6 \cdot log(mincome) + u$$

zu schätzen. Die Regression

$$log(mincome) = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot male + \beta_3 \cdot age + \beta_4 \cdot age2 + \beta_5 \cdot log(income) + u$$

ergibt einen $R_6^2$-Wert von 0,99. Sollten Sie die Variable *log(mincome)* in Ihr Modell aufnehmen? Begründen Sie. (2 Punkte)

*Der sehr hohe $R_6^2$-Wert deutet auf nahezu perfekte Multikollinearität hin. Die Variation in log(mincome) wird beinahe gänzlich durch die übrigen Regressoren erklärt. Der eigene Erklärungsbeitrag von log(mincome) ist vernachlässigbar gering. Die Variable sollte nicht ins Modell aufgenommen werden.*

d) Welche drei Faktoren beeinflussen $Var(\hat{\beta}_6)$? Erwarten Sie auf Basis der in c) angegebenen Ergebnisse eine präzise Schätzung von $\hat{\beta}_6$? Begründen Sie kurz Ihre Antwort. (4 Punkte)

*Sörtermvarianz; Variation von log(mincome) in der Stichprobe; Anteil der Variation in log(mincome) die durch die anderen Regressoren erklärt werden kann;*
*$R_6^2$ hoch -> unpräzise Schätzung*

**Aufgabe 4:** **[18 Punkte]**

Ihr Datensatz enthält Informationen über NBA Basketballspieler. Sie schätzen folgendes Modell:

$$pkt = \beta_0 + \beta_1 \cdot erf + \beta_2 \cdot erf2 + u$$

pkt = pro Spiel durchschnittlich erzielte Punktzahl

erf = Ligazugehörigkeit in Jahren

erf2 = Ligazugehörigkeit in Jahren, quadriert

Sie erhalten folgende Ergebnisse mit SPSS:

**Koeffizienten(a)**

| Modell | | Nicht standardisierte Koeffizienten | | | |
|---|---|---|---|---|---|
| | | B | Standardfehler | T | Signifikanz |
| 1 | (Konstante) | 6,279 | ,937 | 6,700 | ,000 |
| | erf | 1,342 | ,329 | 4,077 | ,000 |
| | erf2 | -,078 | ,024 | -3,240 | ,001 |

a Abhängige Variable: pkt

a) Beantworten Sie folgende Fragen: (8 Punkte)

(1) Wie unterscheidet sich der marginale Effekt der Ligaerfahrung für Personen mit fünf und zehn Jahren Ligaerfahrung?

$$\frac{\partial \widehat{pkt}}{\partial erf} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot erf = 1{,}342 - 2 * 0{,}078 * erf$$

*2*5*(-0,078)=-0,78*

*Der marginale Effekt der Ligaerfahrung ist für Spieler mit 10 Jahren Ligaerfahrung um 0,78 Punkte niedriger als für Spieler mit 5 Jahren.*

(2) Bei welchem Wert für die Variable *erf* ist der marginale Effekt der Ligazugehörigkeit Null?

$0 = 1{,}342 - 2 \cdot 0{,}078 \cdot erf$ ; $erf_0 = 8{,}6$

*Bei einer Ligazugehörigkeit von 8,6 Jahren hat ein weiteres Jahr Erfahrung keinen Effekt auf die Punktzahl.*

(3) Bei welchem Wert für die Variable *erf* ist der marginale Effekt der Ligazugehörigkeit maximal? Begründen Sie Ihr Ergebnis kurz.

*Der marginale Effekt nimmt linear in erf ab. Daher ist er bei erf = 0 maximal.*

(4) Fertigen Sie eine grobe Skizze des geschätzten Zusammenhangs zwischen den pro Spiel erzielten Punkten (y-Achse) und Erfahrung (x-Achse) an.

*Skizze eines umgekehrt U-förmigen Kurvenverlaufs*

b) Sie haben in Ihrer Regression nicht für das Alter kontrolliert. (5 Punkte)

(1) Welche Beziehungen müssen zwischen den Modellvariablen und der weggelassen Variable *Alter* bestehen, damit die Parameter Ihres Modells trotz Fehlen der Variable *Alter* mit KQ unverzerrt geschätzt werden können?

*Alter darf nicht mit erf oder erf2 korreliert sein (1 P) oder (nicht: ‚und')*
*darf nicht mit pkt korreliert sein*

(2) Wie plausibel sind diese Bedingungen? Begründen Sie.

*unplausibel, da Alter und Erfahrung positiv korreliert sind*

*unplausibel, da älter werden (im Gegensatz zu mehr Erfahrung sammeln) negativ mit erzielter Punktzahl korreliert sein sollte (andere plausible Erklärungen möglich)*

c) Welche Annahmen müssen mindestens erfüllt sein, damit die Parameter einer Kleinstquadrateschätzung konsistent geschätzt werden können? Geben Sie für jede der Annahmen eine kurze verbal Beschreibung. (5 Punkte)

*MLR.1 : das Modell ist linear in den Parametern*
*MLR.2 : die Stichprobe ist zufällig gezogen*
*MLR.3 : keine perfekte Kollinearität zwischen Regressoren*
*MLR.4' : der Erwartungswert des Störterms ist Null und die Kovarianzen zwischen dem Störterm und jedem einzelnen Regressor sind Null*

*+1P für das Wissen, dass genau MLR.1 – MLR.4' Voraussetzung für Konsistenz sind (d.h. kein Punkt, wenn zu wenig oder zu viele Annahmen aufgeführt werden)*

**Aufgabe 5:** **[26 Punkte]**

Wahr oder falsch? Tragen Sie für jede der folgenden Aussagen ein „w" für „wahr" oder ein „f" für „falsch" ein. Für jede richtige Antwort gibt es 0,5 Punkte, für jede falsche Antwort werden 0,5 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

| | |
|---|---|
| w | Das Signifikanzniveau eines Tests beschreibt die Wahrscheinlichkeit, dass man $H_0$ verwirft, obwohl $H_0$ wahr ist. |
| w | Die Ablehnungsregion eines t-Tests mit der Nullhypothese $H_0$: $\beta_j=0$ und der Alternativhypothese $H_1$:$\beta_j>0$ befindet sich in der graphischen Darstellung der t-Verteilung am „rechten Ende". |
| w | Im log-linearen Modell misst $100 \cdot \hat{\beta}_1$ die prozentuale Änderung von y bei Änderung von $x_1$ um eine Einheit. |
| f | Wenn der marginale Effekt der abhängigen Variablen vom Wert der Ausprägung der Variablen selbst abhängt, dann wurden Interaktionsterme verwendet. |
| w | Im multiplen Modell implizieren die Annahmen MLR.1 – MLR.5, dass der KQ-Schätzer asymptotisch normalverteilt ist. |
| f | Das Gauss-Markov Theorem gilt nur für kleine Stichproben. |
| f | Wenn $\hat{\beta}_j$ wegen ausgelassener erklärender Variablen verzerrt geschätzt ist, ist der Intervallschätzer eine unverzerrte Alternative. |
| w | Es gibt Modelle, die nichtlinear in den Variablen sind und gleichzeitig linear in den unbekannten Parametern. |
| f | Bei Regressionen durch den Ursprung hat der Achsenabschnittsparameter den Wert eins. |
| f | Ein Regressor ist endogen, wenn er mit dem Störterm unkorreliert ist. |
| w | Medianwerte sind weniger von Ausreißern beeinflusst als Mittelwerte. |
| f | Da beim F-Test mehrere lineare Restriktionen getestet werden, hat der p-Wert eine andere Bedeutung als beim t-Test einer linearen Restriktion. |
| w | Ist ein Schätzverfahren konsistent, so entspricht der geschätzte Wert bei unendlich großer Stichprobe dem Bevölkerungsparameter. |

| | |
|---|---|
| w | Der p-Wert entspricht dem Signifikanzniveau eines Tests, bei dem die berechnete Teststatistik dem kritischen Wert entspricht. |
| w | Der Schätzwert für die Konstante in einem Regressionsmodell ist eine Realisation einer Zufallsvariable. |
| w | Bei immer größer werdenden Stichproben tendiert die Varianz konsistenter Schätzer gegen Null. |
| f | Je höher das $R^2$, desto höher ist die Wahrscheinlichkeit, beim F-Test auf Gesamtsignifikanz eines Modells $H_0$ zu verwerfen. |
| f | Das 95%-Konfidenzintervall ist breiter als das 99%-Konfidenzintervall. |
| f | $\exp(x_1+x_2)=\exp(x_1)+\exp(x_2)$ |
| f | Das Einfügen irrelevanter Regressoren in ein Modell erhöht die Varianz des Störterms. |
| w | Dummy-Variablen können zur Bewertung von Politikmaßnahmen genutzt werden. |
| f | Die Unverzerrtheit der KQ-Schätzer für die Steigungsparameter im multiplen Modell wird durch die Annahme $\mathrm{Var}(u\|x_1,x_2,..,x_k) = 0$ sichergestellt. |
| w | Die KQ-Schätzung eines linearen Modells ist auch dann möglich, wenn die erklärende Variable in logarithmierter Form vorliegt. |
| f | Wenn X den Erwartungswert von Y nicht beeinflusst, sind X und Y korreliert. |
| f | Die Kovarianz kann nur Werte zwischen -1 und +1 annehmen. |
| w | Das $\overline{R}^2$ kann steigen, wenn eine zusätzliche erklärende Variable berücksichtigt wird. |
| f | Für diskrete Zufallsvariablen kann keine kumulative Verteilungsfunktion berechnet werden. |
| f | Das Weglassen relevanter Variablen führt in der Regel zu einer Unterschätzung der KQ-Parameter der anderen Regressoren. |
| w | Wenn die Annahme der Normalverteilung der KQ-Schätzer nicht gilt, sind Konfidenzintervalle unter den Annahmen MLR1-MLR5 weiterhin asymptotisch interpretierbar. |
| f | $\mathrm{Var}(aX+b)=a*\mathrm{Var}(X)$ |
| w | Die Nullhypothese bezieht sich auf den unbekannten Bevölkerungsparameter, nicht auf den geschätzten Wert. |
| f | Benutzt man Dummy-Variablen, um die Effekte der vier Jahreszeiten zu kontrollieren und berücksichtigt lediglich 3 Indikatorvariablen in der Modellspezifikation, so kann über die nicht berücksichtigte Jahreszeit keine Aussage getroffen werden. |
| w | Bei Schätzung der Spezifikation $y=\beta_0+\beta_1x+\beta_2x^2$ ist der marginale Effekt von x auf y nicht konstant. |
| f | Je größer der berechnete t-Wert, desto wahrscheinlicher wird, dass die zugehörige $H_0$ gilt. |
| w | Im log-linearen Modell ist $100 \cdot \hat{\beta}_1$ als Semi-Elastizität interpretierbar. |
| w | Die Nullhypothese $H_0$: $\beta_1=\beta_2$ kann mit dem t-Test getestet werden. |
| f | Im Modell mit logarithmierter abhängiger Variable dürfen keine Dummy-Variablen als erklärende Variablen verwendet werden. |
| f | Tests zu Linearkombinationen von Parametern können nicht mit dem t-Test durchgeführt werden. |
| w | Wenn X und Y unabhängig voneinander sind, dann ist $E(X\|Y)=E(X)$. |
| w | Ein Regressor ist exogen, wenn er mit dem Störterm unkorreliert ist. |
| w | Unter „omitted variable bias" versteht man eine Verzerrung des KQ-Schätzers, die auftreten kann, wenn relevante erklärende Variablen im Modell fehlen. |
| f | Ein einseitiger t-Test auf dem 5% Niveau hat eine größere Ablehnungsregion als ein einseitiger t-Test auf dem 10% Niveau. |
| w | Die F-Verteilung kann nur positive Werte annehmen. |
| w | Das Konfidenzintervall um präzise geschätzte Parameter ist enger als das um Parameter mit großem Standardfehler. |
| w | Ein Chow-Test kann in Form eines F-Tests durchgeführt werden. |
| f | Das Gauss-Markov-Theorem macht eine Aussage zu nichtlinearen Schätzverfahren. |
| w | Wird eine logarithmierte abhängige Variable umskaliert, ändert sich bei Schätzung des Modelles lediglich der Achsenabschnitt. |

| | |
|---|---|
| w | F- und t-Test können bei zweiseitigen Tests zum gleichen Ergebnis kommen. |
| f | Ist die Stichprobengröße zu klein, so sind die Parameter verzerrt. |
| f | Im multiplen linearen Modell werden für jeden Regressor separate $R^2$-Werte errechnet. |
| w | Die Zählerfreiheitsgrade des F-Tests hängen nicht von der Zahl der Beobachtungen ab. |
| w | Von zwei unverzerrten Schätzern für θ (W und V) ist W effizienter, wenn Var(W)<Var(V). |

**Aufgabe 6:** **[14 Punkte]**

Welche Antwort ist richtig? Kreuzen Sie nur **eine Antwort** pro Aufgabe an. Falls mehrere Aussagen korrekt sind, kreuzen Sie **nur** die entsprechende **Antwortkombination** an. Für jede richtige Antwort gibt es 1 Punkt. Für falsche Antworten werden keine Punkte abgezogen.

| | | |
|---|---|---|
| 1. | Die ceteris paribus Annahme im multiplen linearen Regressionsmodell | |
| a | ❑ | kann mittels des Lagrange-Multiplier Test überprüft werden. |
| b | ❑ | gilt für Männer, aber nicht für Frauen. |
| c | ❑ | ist eine asymptotische Eigenschaft. |
| d | ❑ | a und b. |
| e | ❑ | a, b und c. |
| f | ☒ | keine der Antworten. |
| 2. | Unverzerrtheit der Steigungsparameter im KQ-Verfahren | |
| a | ❑ | gilt als Eigenschaft des Schätzverfahrens nur in großen Stichproben. |
| b | ❑ | gilt als Eigenschaft des Schätzverfahrens auch bei kleiner Stichprobe. |
| c | ❑ | wird als finite sample property bezeichnet. |
| d | ❑ | gilt unabhängig von der Verteilung der Störterme. |
| e | ❑ | a, b und d. |
| f | ☒ | b, c und d. |
| 3. | Bei der Auswahl von Regressoren | |
| a | ❑ | gilt es, over-controlling zu vermeiden. |
| b | ❑ | spielt omitted variable bias keine Rolle. |
| c | ❑ | kann es im Sinne einer präziseren Schätzung sinnvoll sein, eine zusätzliche Variable ins Modell aufzunehmen, die nicht mit den berücksichtigten korreliert ist. |
| d | ❑ | a und b. |
| e | ☒ | a und c. |
| f | ❑ | a, b und c. |
| 4. | Die t-Verteilung | |
| a | ❑ | hat eine Varianz die vom gewählten Signifikanzniveau abhängt. |
| b | ❑ | weist für Mittelwert und Median den gleichen Wert auf. |
| c | ❑ | wird bei nur einer zu testenden Restriktion asymmetrisch. |
| d | ❑ | nähert sich mit steigender Zahl an Freiheitsgraden der Normalverteilung an. |
| e | ☒ | b und d. |
| f | ❑ | Alle Antworten sind zutreffend. |
| 5. | Bei welchen der folgenden Modelle ist aufgrund ihrer Spezifikation mindestens eine der Gauss-Markov Annahmen zum linearen Regressionsmodell verletzt? | |
| a | ❑ | $\log(y) = \beta_0 + \beta_1 \log(x) + \beta_2 \log(x^2) + u$ |

| | | |
|---|---|---|
| b | ☐ | $\log(y) = \beta_0 + \beta_1 \log(x) + \beta_2 [\log(x)]^2 + u$ |
| c | ☐ | $y = \beta_0 + x^{\beta_1} + u$ |
| d | ☐ | $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ |
| e | ☒ | a und c. |
| f | ☐ | a, b und c. |
| 6. | Ein sehr hoher p-Wert bei einem Test auf statistische Signifikanz | |
| a | ☐ | kann auf einen betraglich großen Koeffizienten zurückzuführen sein. |
| b | ☒ | kann auf eine sehr ungenaue Schätzung zurückzuführen sein. |
| c | ☐ | deutet auf ein zu niedriges Signifikanzniveau hin. |
| d | ☐ | deutet auf statistische Signifikanz des Koeffizienten hin. |
| e | ☐ | a und b. |
| f | ☐ | c und d. |
| 7. | Im Modell $y_i = \beta_1 x_i + u_i$ und bei Gültigkeit der Annahmen MLR.1 bis MLR.5 wird $\beta_1$ mittels $\sum_i x_i y_i \Big/ \sum_i x_i^2$ | |
| a | ☐ | unverzerrt geschätzt, wenn der Populationsparameter der Regressionskonstante $\beta_0 = 0$. |
| b | ☐ | unverzerrt geschätzt, wenn der Populationsparameter des Steigungsparameters $\beta_1 = 0$. |
| c | ☐ | unverzerrt geschätzt, wenn $\bar{x} = 0$. |
| d | ☐ | effizient geschätzt, wenn $\beta_0 = 0$ |
| e | ☒ | a, c und d. |
| f | ☐ | b und d. |
| 8. | Gilt die Normalverteilungsannahme der Störterme nicht, | |
| a | ☐ | kann die asymptotische Gültigkeit von t- und F-Tests selbst unter zusätzlichen Annahmen nicht aus dem zentralen Grenzwertsatz abgeleitet werden. |
| b | ☐ | ist der KQ-Schätzer inkonsistent. |
| c | ☐ | ist der KQ-Schätzer nicht mehr BLUE, da er an Effizienz einbüßt. |
| d | ☐ | a und c. |
| e | ☐ | a und b. |
| f | ☒ | keine der Antworten. |
| 9. | Mittels KQ vorhergesagte Werte | |
| a | ☐ | haben im einfachen Regressionsmodell am Mittelwert der Daten einen erwarteten Vorhersagefehler von Null. |
| b | ☐ | haben die kleinste Varianz, wenn die abhängige Variable an ihrem Mittelwert betrachtet wird. |
| c | ☐ | sind Zufallsvariablen. |
| d | ☐ | keine der Antworten. |
| e | ☒ | a und c. |
| f | ☐ | b und c. |
| 10. | Geht in die Spezifikation eines Modells eine Dummy-Variable ein, | |
| a | ☐ | ist das $R^2$ nicht mehr berechenbar. |
| b | ☐ | fällt die Schätzgüte des Modells $(R^2)$ nicht. |
| c | ☐ | bewirkt dies eine Parallelverschiebung der Regressionsgerade im Ausmaß $\beta_0 + \beta_{Dummy}$. |
| d | ☐ | misst diese einen Niveauunterschied für zwei Gruppen. |
| e | ☐ | a, b und c. |

| | | |
|---|---|---|
| f | ☒ | b und d. |
| 11. | Wird das Modell $\log(y) = \beta_0 + \beta_1 \log(x) + u$ geschätzt, | |
| a | ❑ | misst $\beta_1$ die absolute Änderung in y bei Änderung von x um ein Prozent. |
| b | ❑ | misst $\beta_1$ die absolute Änderung in y bei Änderung von x um einen Prozentpunkt. |
| c | ❑ | misst $\beta_1$ die relative Änderung in y bei Änderung von x um einen Prozentpunkt. |
| d | ☑ | misst $\beta_1$ die relative Änderung in y bei Änderung von x um ein Prozent. |
| e | ❑ | keine der Antworten. |
| f | ❑ | b und d. |
| 12. | Wenn sich der Erwartungswert der Störterme in einer linearen Regression je nach Ausprägung der Regressoren unterscheidet | |
| a | ❑ | spricht man von Homoskedastie. |
| b | ❑ | müssen die Koeffizienten des Modells als kausale Effekte interpretiert werden. |
| c | ❑ | sind die geschätzten Parameter unverzerrt, solange die Störterme normalverteilt sind. |
| d | ❑ | muss das Modell mehrmals geschätzt werden. |
| e | ❑ | a und d. |
| f | ☒ | keine der Antworten. |
| 13. | Die Verwendung von Interaktionstermen zwischen unterschiedlichen erklärenden Variablen | |
| a | ❑ | ist zwingende Voraussetzung für die Durchführung des Chow-Tests. |
| b | ❑ | ist nur möglich, wenn die ceteris paribus Bedingung erfüllt ist. |
| c | ☒ | ermöglicht, für verschiedene Gruppen unterschiedliche Steigungsparameter zu berechnen. |
| d | ❑ | ist nur für Dummy-Variablen sinnvoll. |
| e | ❑ | a, b und c. |
| f | ❑ | a und c. |
| 14. | Perfekte Multikollinearität | |
| a | ❑ | führt zu einer sehr genauen Schätzung der Effekte der kollinearen Regressoren. |
| b | ❑ | führt zu einer sehr ungenauen Schätzung der Effekte der kollinearen Regressoren. |
| c | ☒ | tritt auf, wenn bei einer Schätzung mit Dummy Variablen und Konstante keine Referenzkategorie gewählt wird. |
| d | ❑ | tritt auf, wenn die abhängige Variable und mindestens einer der Regressoren in ein und derselben Einheit gemessen werden. |
| e | ❑ | b und c. |
| f | ❑ | b, c und d. |