Aufgabe 1: [8 Punkte]

Leiten Sie die Kleinstquadrateschätzer für β_0 und β_1 im einfachen linearen Modell aus dem Minimierungskalkül $\min\sum_{i=1}^{n} u_i^2$ ab.

1)

$$\min_{\beta_0,\beta_1} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

Ableitung nach β_0:

$$2 \cdot \sum (y_i - \beta_0 - \beta_1 x_i) \cdot (-1) = 0$$

Ableitung nach β_1:

$$2 \cdot \sum (y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) = 0$$

2)
Umformen der FOC für $\widehat{\beta_0}$:

$$\sum y_i - \sum \widehat{\beta_0} - \sum \widehat{\beta_1} x_i = 0$$

$$n\bar{y} - n\widehat{\beta_0} - n\widehat{\beta_1}\bar{x} = 0$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

3)
Umformen der FOC für $\widehat{\beta_1}$:

$$2 \cdot \sum (y_i - \bar{y} + \widehat{\beta_1}\bar{x} - \widehat{\beta_1} x_i) \cdot (-x_i)$$

$$\sum (y_i - \bar{y}) \cdot x_i = \sum (y_i - \bar{y}) \cdot (x_i - \bar{x})$$

$$\sum (x_i - \bar{x}) \cdot x_i = \sum (x_i - \bar{x})^2$$

Aufgabe 2: [15,5 Punkte]

Sie schätzen eine Lohnregression für 525 Beobachtungen auf Basis des Modells: $wage = \beta_0 + \beta_1 \cdot exper + u$. Dabei misst wage den Stundenlohn in Dollar und exper die Berufserfahrung in Jahren. Sie erhalten folgende Ergebnisse mit SPSS:

ANOVA(b)

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	88,916	1	88,916	6,583	,011(a)
	Residuen	7063,665	523	13,506		
	Gesamt	7152,581	524			

a Einflußvariablen : (Konstante), exper

b Abhängige Variable: wage

Koeffizienten(a)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		B	Standardfehler	Beta	T	Signifikanz
1	(Konstante)	5,384	,258		20,894	,000
	exper	,030	,013	,111	2,308	,016

a Abhängige Variable: wage

a) Interpretieren Sie die Parameter $\hat{\beta}_0$ und $\hat{\beta}_1$ inhaltlich. Sind die Parameter bei einer Irrtumswahrscheinlichkeit von 1% statistisch signifikant? *(3 Punkte)*

- $\hat{\beta}_0$*: ist auf dem 1% Niveau statistisch signifikant von Null verschieden. Eine Person ohne Berufserfahrung verdient im Durchschnitt 5,384 Dollar pro Stunde.*

- $\hat{\beta}_1$*: ist auf dem 1% Niveau nicht statistisch signifikant von Null verschieden. Mit jedem zusätzlichen Jahr Berufserfahrung steigt der durchschnittliche Stundenlohn um 3 Cent.*

b) Sie haben in Ihrer Schätzung nicht für das Ausbildungsniveau der Personen (educ) kontrolliert und befürchten eine Verzerrung wegen ausgelassener Variablen. *(5,5 Punkte)*

b1) Erklären Sie allgemein, unter welchen Bedingungen eine solche Verzerrung entsteht und welche Parameter sie betrifft. Welche Annahme der Kleinstquadrate-Methode könnte im Beispiel verletzt sein?

- eine Verzerrung wegen ausgelassener Variablen entsteht, wenn wichtige Regressoren im geschätzten Modell fehlen und die ausgelassenen Variablen mit den eingeschlossenen korreliert sind

- die Verzerrung betrifft alle geschätzten Parameter des unvollständigen Modells

- die Annahme SLR.4: E(u|exper)=0 könnte verletzt sein

b2) Wird der Steigungsparameter im Beispiel über- oder unterschätzt? Begründen Sie Ihre Aussagen und nutzen Sie die nachfolgende Korrelationsmatrix.

Korrelationen

		wage	educ	exper
wage	Korrelation nach Pearson	1	,405(**)	,111(*)
educ	Korrelation nach Pearson	,405(**)	1	-,301(**)
exper	Korrelation nach Pearson	,111(*)	-,301(**)	1

educ misst die Ausbildung der Person in Jahren

- die Verzerrung hat das Ausmaß $\hat{\beta}_2\tilde{\delta}_1$ *wobei* $\hat{\beta}_2$ *der geschätzte Parameter von educ und* $\tilde{\delta}_1$ *der Steigungsparameter einer Regression von educ auf exper ist*

- educ und exper sind negativ korreliert, somit ist $\tilde{\delta}_1$ *negativ*

- Ausbildung und Stundenlohn sind positiv korreliert, somit ist $\hat{\beta}_2$ *positiv und* β_1 *im Beispiel unterschätzt*

c) Sie beschließen, das Modell aus a) zu ändern und schätzen stattdessen das Modell: $wage = \beta_0 + \beta_1 \cdot exper + \beta_2 \cdot educ + u$, wobei educ die Ausbildung einer Person in Jahren misst. Als SPSS Output erhalten Sie:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,474(a)	,224	,221	3,26000

a Einflußvariablen : (Konstante), exper, educ

ANOVA(b)

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	1604,967	2	802,484	75,509	,000(a)
	Residuen	5547,614	522	10,628		
	Gesamt	7152,581	524			

a Einflußvariablen : (Konstante), exper, educ
b Abhängige Variable: wage

Koeffizienten(a)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-3,380	,769		-4,398	,000
	exper	,070	,011	,257	6,355	,000
	educ	,644	,054	,483	11,944	,000

a Abhängige Variable:wage

Führen Sie anhand dieser Information sowie der relevanten Größen aus Aufgabe a) einen **F-Test** auf statistische Signifikanz der Variable educ durch (α=5%). Geben Sie Nullhypothese, Alternativhypothese, Teststatistik, kritischen Wert, Freiheitsgrade und Schlusslogik an. *(4 Punkte)*

- H_0: $\beta_2 = 0$ H_1: $\beta_2 \neq 0$

- $$F_{emp} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k-1)} = \frac{(7063{,}665 - 5547{,}614)/1}{5547{,}614/(525-2-1)} = 142{,}65$$

- $F_{krit1;522}$=3,84

- $F_{emp} > F_{krit} \rightarrow H_0$ *verwerfen → Ausbildung hat einen signifikanten Einfluss auf den Lohn*

d) Unter welchen Umständen kann ein geschätzter Parameter des Modells in c) als kausaler Effekt auf den Stundenlohn interpretiert werden? Diskutieren Sie kurz, ob diese Umstände für den Ausbildungsparameter erfüllt sind. *(3 Punkte)*

- *ein Parameter ist als kausaler Effekt interpretierbar, wenn bei Änderung des Regressors gleichzeitig alle anderen Einflüsse auf die abhängige Variable im Erwartungswert konstant sind, also alle anderen Regressoren und der Störterm. <u>Alternativ:</u> kausaler Effekt, wenn Annahme MLR.4 erfüllt ist*
- *sowohl die Ausbildung als auch der Lohn werden z.B. von der Fähigkeit einer Person beeinflusst*
- *unterschiedliche Ausbildungsniveaus gehen somit im Erwartungswert mit unterschiedlichen Fähigkeiten einher, welche im Störterm enthalten sind*
- *bei Änderung des Ausbildungsregressors ändert sich daher Erwartungswert des Störterms und der geschätzte Ausbildungsparameter ist verzerrt und nicht mehr als kausaler Effekt interpretierbar*

Aufgabe 3: **[13 Punkte]**

Im weiteren Verlauf Ihrer Untersuchung fügen Sie noch einige erklärende Variablen zu Ihrem Modell hinzu:

female	= 1, wenn die Person weiblich ist, 0 sonst
married	= 1, wenn die Person verheiratet ist, 0 sonst
fem_educ	= female · educ
fem_marr	= female · married

Sie schätzen nun das Modell:

$$wage = \beta_0 + \beta_1 \cdot exper + \beta_2 \cdot educ + \beta_3 \cdot female + \beta_4 \cdot fem\_educ + \beta_5 \cdot married + \beta_6 \cdot fem\_marr + u$$

SPSS liefert folgenden Output:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,590(a)	,348	,340	2,99972

a Einflußvariablen : (Konstante), fem_marr, educ, exper, married, fem_educ, female

Koeffizienten(a)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		B	Standardfehler	Beta	T	Signifikanz
1	(Konstante)	-3,065	,898		-3,414	,001
	exper	,054	,011	,197	4,948	,000
	educ	,610	,063	,458	9,627	,000
	female	,593	1,276	,080	,465	,642
	fem_educ	-,083	,098	-,143	-,845	,399
	married	2,065	,408	,273	5,065	,000
	fem_marr	-2,681	,545	-,315	-4,917	,000

a Abhängige Variable: wage

a) Wie unterscheidet sich der erwartete Lohn *(3 Punkte)*

a1) für verheiratete und unverheiratete Männer,

- $\Delta wage = \hat{\beta}_5$
- =2,065
- *bei gegebener Ausbildung und gegebener Berufserfahrung verdienen verheiratete Männer im Erwartungswert* ***2,07$ mehr pro Stunde*** *als unverheiratete Männer.*

a2) für verheiratete und unverheiratete Frauen,

- $\Delta wage = \hat{\beta}_5 + \hat{\beta}_6$
- $= 2{,}065 + (-2{,}681) = -0{,}616$
- *bei gegebener Ausbildung und gegebener Berufserfahrung verdienen verheiratete Frauen im Erwartungswert* ***61ct. weniger pro Stunde*** *als unverheiratete Frauen.*

wenn die genannten Personengruppen in den Merkmalen Berufserfahrung und Ausbildung übereinstimmen? Benennen Sie das Ergebnis und zeigen Sie Ihren Rechenweg.

b) Wie lassen sich die eingefügten Variablen zu Geschlecht und Familienstand in einer grafischen Darstellung des geschätzten Zusammenhangs zwischen wage und exper interpretieren? (Setzen Sie zur Vereinfachung educ=0.) Beantworten Sie die Frage verbal oder mithilfe einer Skizze. *(3 Punkte)*

- *Die eingefügten Dummy-Variablen führen dazu, dass die Achsenabschnitte der Regressionsgeraden nicht für alle Gruppen identisch sind.*
- *Sie führen zu genau* *<u>vier</u>* *unterschiedlichen Achsenabschnitten (Männer/Frauen jeweils verheiratet/ unverheiratet)*
- *In Skizze:* **<u>2 Punkte</u>** *dafür, dass eine Parallelverschiebung einer Geraden dargestellt wird, mit exper auf der x-Achse und wage auf der y-Achse*
- **<u>1 Punkt</u>** *für Darstellung von genau 4 parallelen Geraden für Männer/Frauen jeweils verheiratet /unverheiratet*

c) Betrachten Sie die Effekte von Ausbildung und Geschlecht. *(3 Punkte)*

c1) Ist der aus dieser Stichprobe geschätzte marginale Effekt der Ausbildung auf den erwarteten Lohn (ceteris paribus) für Frauen größer, kleiner oder genauso groß wie für Männer? (Hinweis: Sie müssen Ihre Aussage nicht auf statistische Signifikanz überprüfen.) Erläutern Sie Ihr Ergebnis kurz!

- $\frac{\partial \hat{wage}}{\partial educ} = \hat{\beta}_2 + \hat{\beta}_4 \cdot female$ *(nicht gefordert)*

→ *marginaler Effekt* $= \hat{\beta}_2$ *für Männer und* $\hat{\beta}_2 + \hat{\beta}_4$ *für Frauen.*

- *Da* $\hat{\beta}_4 < 0$ *folgt, dass der marginale Effekt der Ausbildung auf den erwarteten Lohn für Frauen kleiner ist als für Männer.*

c2) Wie groß ist der Lohnunterschied zwischen unverheirateten Männern und Frauen mit jeweils 10 Jahren Ausbildung?

- $\frac{\partial \hat{wage}}{\partial female} = \hat{\beta}_3 + \hat{\beta}_4 \cdot educ + \hat{\beta}_6 \cdot married$ *(nicht gefordert)*

- *bei 10 Jahren Ausbildung: 0,593-0,083*10+0=-0,237 Der Lohnunterschied beträgt 24 ct. zu Lasten der Frauen*

d) Sie möchten wissen, ob die neu eigefügten Variablen zu Geschlecht und Familienstand ihr Modell gegenüber der Spezifikation in Aufgabe 1c) signifikant verbessert haben. Testen Sie dies auf dem 5%-Niveau mithilfe eines geeigneten Tests. Geben Sie die Nullhypothese, die Teststatistik, den kritischen Wert, Ihre Schlusslogik sowie Ihre Schlussfolgerung an. *(4 Punkte)*

- *benötigt: F-Test, ermitteln über* R^2

- $H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

- $F = \frac{(R^2_u - R^2_r)/q}{(1 - R^2_u /(n-k-1)}$ *oder hier:*

- $F = \frac{(0{,}348 - 0{,}224)/4}{(1-0{,}348)/(525-7)} = \frac{0{,}031}{8{,}855} = 24{,}628$

- $F_{krit4;518}=2{,}38$

- *da F>Fc kann die H0 verworfen werden.*

- *Das Modell wurde durch das Hinzufügen von Geschlechts- und Familienstandsvariablen signifikant verbessert* .

Aufgabe 4: **[12,5 Punkte]**

Sie schätzen die folgenden Spezifikationen für die gleiche Stichprobe (N=525):

Modell 1: $log(wage) = \beta_0 + \beta_1 \cdot exper + u$

Modell 2: $log(wage) = \alpha_0 + \alpha_1 \cdot log(exper) + u$

SPSS-Output Modell 1:

Koeffizienten(a)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta	B	Standardfehler
1	(Konstante)	1,549	,037		41,872	,000
	exper	,004	,002	,111		

a Abhängige Variable: log(wage)

SPSS-Output Modell 2:

Koeffizienten(a)

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	1,343	,056		24,198	,000
	log_exper	,117	,021	,234	5,516	,000

a Abhängige Variable: log(wage)

a) Interpretieren Sie für beide Modelle inhaltlich den Effekt der Berufserfahrung auf den Stundenlohn. *(2 Punkte)*

- Modell 1: Erhöhung der Berufserfahrung um ein Jahr, Erhöhung des erwarteten Lohns um 0,4%.

- Modell 2: Erhöhung der Berufserfahrung um 1%, Erhöhung des erwarteten Lohns um ca. 0,12%

b) Testen Sie für Modell 1 auf dem 1% Signifikanzniveau, ob der Koeffizient $\hat{\beta}_1$ signifikant von Null verschieden ist. Geben Sie die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert der t-Verteilung, die Anzahl der Freiheitsgrade sowie die Testentscheidung an. *(4 Punkte)*

- H_0: $\beta_1 = 0$, H_1: $\beta_1 \neq 0$

- $t = \dfrac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \dfrac{0{,}004}{0{,}002} = 2$

- $t_{0{,}01;523} = 2{,}576$

- $t < t_c$*: Die* H_0 *kann auf dem 1%-Signifikanzniveau nicht verworfen werden*

c) Wie ändert sich in Modell 1 der geschätzte Koeffizient $\hat{\beta}_1$, wenn *(2 Punkte)*

c1) die Berufserfahrung in Fünfjahresschritten statt in Jahren gemessen wird?

*- 0,004*5 = 0,02*

c2) der Stundenlohn in 100 Dollar statt in Dollar gemessen wird?

- Abhängige Variable logarithmiert, Koeffizient bleibt unverändert

d) Sie haben das 95%-Konfidenzintervall [0,075;0,158] für den Koeffizienten der Berufserfahrung $\hat{\alpha}_1$ in Modell 2 berechnet. *(4,5 Punkte)*

d1) Nennen Sie drei Faktoren, welche allgemein die Breite eines Konfidenzintervalls beeinflussen und beschreiben Sie deren Wirkungsrichtung.

- Varianz/Standardabweichung des geschätzten Koeffizienten: je präziser der Koeffizient geschätzt, desto enger das Konfidenzintervall [oder direkte Wirkung auf die Varianz des Koeffizienten: Beobachtungszahl, SST_x*, oder* σ^2*]*

- *Signifikanzniveau: je niedriger das Signifikanzniveau α, desto größer das Konfidenzintervall bei gegebener Zahl von Freiheitsgraden (desto größer wird der kritische t-Wert)*
- *Freiheitsgrade: je größer die Zahl der Freiheitsgrade bei gegebenem Signifikanzniveau, desto enger das Konfidenzintervall (desto kleiner der kritische t-Wert)*

d2) Wie interpretiert man ein Konfidenzintervall allgemein?

- *Interpretation: Würde man für viele wiederholt gezogene Stichproben die Berechnung des Konfidenzintervalls durchführen, enthielten 95% aller so konstruierter Konfidenzintervalle den wahren Bevölkerungsparameter.*

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.

f	Die Kovarianz kann nur Werte zwischen -1 und +1 annehmen.
w	Die t-Verteilung ist symmetrisch.
f	Wenn relevante erklärende Variablen in einer Spezifikation fehlen, können die Koeffizienten nie als kausale Effekte interpretiert werden.
w	Bei einer quadratischen Modellspezifikation $y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + u$ variiert der marginale Effekt von x auf y mit x.
f	Für diskrete Zufallsvariablen kann keine kumulative Verteilungsfunktion berechnet werden.
w	Die Varianz ist das Quadrat der Standardabweichung einer Zufallsvariablen.
f	Var(aX+b)=a*Var(X), wenn a und b Konstanten sind und X eine Zufallsvariable ist.
w	Medianwerte können sich vom arithmetischen Mittel unterscheiden.
f	Im linearen Modell gibt die Regressionskonstante den Mittelwert der abhängigen Variable an.
w	Bei linearen Regressionen wird die Schätzgüte z.B. mit dem Wert des R^2 gemessen.
w	Die Varianz des geschätzten Steigungsparameters hängt von der Streuung der erklärenden Variable ab.
w	Im Zufallsfehler spiegeln sich alle die Faktoren wider, die die abhängige Variable beeinflussen und im Regressionsmodell nicht berücksichtigt wurden.
f	Werden logarithmierte erklärende Variablen genutzt, so muss der geschätzte Steigungsparameter negativ sein.
w	Die Varianz des geschätzten Achsenabschnittsparameters wird vom Stichprobenumfang beeinflusst.
f	Wenn die Regressionsgerade horizontal verläuft, ist das Bestimmtheitsmaß 1.
f	Wenn im einfachen linearen Modell $\bar{x} = 0$ gilt, ist der Schätzer $\hat{\beta}_1 = \widehat{\mathrm{cov}}(x, y) / \widehat{\mathrm{var}}(x)$ nicht definiert.
f	Grundidee des KQ-Schätzers ist es, eine Linie so durch eine Punktwolke zu legen, dass die Summe der vertikalen Abweichungen der beobachteten Werte von der Linie minimiert wird.
f	Das R^2 ist im multiplen Modell gleich dem quadrierten Korrelationskoeffizienten zwischen x_i und $\hat{x}_i$.
f	Das Gauss-Markov-Theorem macht eine Aussage zu nichtlinearen Schätzverfahren.
w	Ein Regressor ist exogen, wenn er mit dem Störterm unkorreliert ist.
f	Das Weglassen relevanter Variablen führt in der Regel zu einer Überschätzung der KQ-Parameter der anderen Regressoren.
w	Das multiple Modell ist nicht schätzbar, wenn der Korrelationskoeffizient zwischen zwei Regressoren den Wert 1 annimmt.
f	Im multiplen Modell ist eine ceteris paribus Interpretation der Parameter nicht mehr

	möglich.
f	Die Gruppe der linearen Regressionsmodelle beinhaltet auch Modelle, die nichtlinear in den Parametern sind.
f	Hypothesen zu Linearkombinationen von Parametern können nicht mit dem t-Test getestet werden.
f	Ein einseitiger t-Test auf dem 5% Niveau hat eine größere Irrtumswahrscheinlichkeit als ein zweiseitiger t-Test auf dem 5% Niveau.
w	Der p-Wert ist das Signifikanzniveau eines Tests, bei dem der berechnete tatsächliche Wert der Teststatistik dem kritischen Wert entspricht.
w	Die Nullhypothese bezieht sich auf den unbekannten Bevölkerungsparameter.
f	Die Gauss-Markov Annahmen werden verletzt, wenn der Störterm u unabhängig von den erklärenden Variablen ist.
f	Ein statistisch signifikanter Koeffizient ist immer ökonomisch signifikant.
w	Die Nullhypothese H_0: β_1=-β_2 kann mit dem t-Test getestet werden.
w	Einzelne lineare Restriktionen können sowohl mit dem F- als auch mit dem t-Test getestet werden.
f	Die Teststatistik des Lagrange Multiplier Tests folgt der t-Verteilung.
w	Unter Gültigkeit der Gauss-Markov Annahmen ist der KQ-Schätzer der effiziente Schätzer in der Klasse der linearen unverzerrten Schätzer.
w	Folgt der Störterm nicht einer Normalverteilung, gilt die BLUE-Eigenschaft des KQ-Schätzers immer noch.
f	Die Annahme, dass die Kovarianz zwischen erklärenden Variablen und Störterm gleich Null ist, impliziert, dass kein Zusammenhang zwischen ihnen besteht.
w	Der Lagrange Multiplier Test eignet sich – wie der F-Test – zum gleichzeitigen Testen mehrerer Restriktionen im linearen Modell.
F	Nur Modellspezifikationen mit großer Parameterzahl können konsistent sein.
F	Bei kleinem R^2 sollten die Koeffizienten mit Vorsicht interpretiert werden, da sie in diesem Fall häufig verzerrt sind.
F	Quadratische Spezifikationen eignen sich nicht, um nicht-konstante marginale Effekte zu beschreiben.
f	Bei Umskalierung einer erklärenden Variable ändern sich die Koeffizienten aller erklärenden Variablen.
f	Der Koeffizient einer logarithmierten erklärenden Variable gibt eine Elastizität an, wenn die abhängige Variable negative Werte annimmt.
w	Negative Werte des angepassten $\overline{R}^2$ sind möglich.
w	Das nach KQ-Schätzung vorhergesagte $\hat{y}$ ist eine Zufallsvariable.
w	Wenn der partielle marginale Effekt einer erklärenden Variable vom Wert der Ausprägung einer anderen Variable abhängt, dann kann dies durch Interaktionsterme abgebildet werden.
w	Interaktionen zwischen Dummy-Variablen können genutzt werden, um zu testen, ob partielle Effekte von Dummy-Variablen für zwei Gruppen variieren.
f	Wenn binäre Variablen mit den Werten -1 und 1 kodiert werden, bleibt der Schätzer unverzerrt, verliert aber Effizienz.
w	Führt man eine Schätzung ohne Konstante durch, können die Schätzer hierdurch verzerrt werden.

w	Ein vollständig interagiertes Modell erlaubt Unterschiede zwischen den Parametern für zwei Gruppen.
w	Ein vollständig interagiertes Modell erlaubt Unterschiede zwischen den Parametern für drei Gruppen.
f	Binär kodierte Variablen können keine quantitativen Informationen abbilden.
F	Die Koeffizienten von Dummyvariablen nehmen Werte von entweder 0 oder 1 an.

Aufgabe 6: **[15 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.	Querschnittsdaten unterscheiden sich von Zeitreihendaten dadurch, dass	
a	☐	Querschnittsdaten zu einem Zeitpunkt erhoben werden.
b	☐	Querschnittsdaten eine gegebene Beobachtungseinheit mehrfach betrachten.
c	☐	Querschnittsdatensätze mehr Beobachtungen haben.
d	☐	Querschnittsdaten einzelne Beobachtungseinheiten nicht wiederholt betrachten.
e	☐	a und b.
f	☒	a und d.

2.	Multikollinearität	
a	☐	bezeichnet die Korrelation der exogenen Variablen mit der endogenen Variablen.
b	☐	bezeichnet die Korrelation der endogenen Variablen mit dem Störterm.
c	☐	bezeichnet die Korrelation von exogenen Variablen mit anderen exogenen Variablen.
d	☐	erhöht die Varianz der Schätzer für die Steigungsparameter.
e	☒	c und d.
f	☐	b und d.

3.	Unter den Gauss-Markov Annahmen	
a	❑	gilt die BLUE-Eigenschaft des KQ-Schätzers.
b	❑	ist der KQ-Schätzer der beste unverzerrte Schätzer.
c	❑	folgt der KQ-Schätzer der Normalverteilung.
d	❑	ist der KQ-Schätzer konsistent.
e	❑	a und b.
f	☒	a und d.

4.	Die „dummy variable trap“	
a	☒	ist ein Problem perfekter Multikollinearität.
b	❑	tritt auf, wenn in einer Schätzung mit Konstante eine Referenzgruppe gewählt wird.
c	❑	tritt auf, wenn in einer Schätzung ohne Konstante keine Referenzgruppe gewählt wird.
d	❑	lässt sich durch Skalierung vermeiden.
e	❑	b und c.
f	❑	keine der Antworten.

5.	Die Kovarianz von zwei Zufallsvariablen	
a	❑	kann nicht negativ sein.
b	❑	liegt zwischen -1 und 1.
c	❑	ist ein Maß nicht-linearer Zusammenhänge.
d	❑	nimmt für unabhängige Zufallsvariablen den Wert 1 an.
e	❑	b und c.
f	☒	keine der Antworten.

6.	Die t-Verteilung	
a	❑	beschreibt stets die Verteilung von R^2
b	❑	hat mit der Beobachtungszahl variierende Freiheitsgrade.
c	❑	ist symmetrisch um den Erwartungswert 0.
d	❑	konvergiert bei wachsenden Freiheitsgraden gegen die Normalverteilung.
e	☒	b, c und d.
f	❑	alle Antworten treffen zu.

7.	Die Varianz des Störterms in der Grundgesamtheit	
a	❑	ist eine Funktion der erklärenden Variablen.
b	❑	steigt mit steigender Stichprobengröße.
c	☒	kann anhand der Summe quadrierter Residuen approximiert werden.
d	❑	ist eine Konstante.
e	❑	a und b.
f	❑	c und d.

8.	Der Lagrange Multiplier Test	
a	❑	erlaubt, gleichzeitig mehrere lineare Restriktionen zu testen.
b	❑	nutzt eine Hilfsregression.
c	❑	hat eine R^2-verteilte Teststatistik.
d	❑	nutzt so viele Freiheitsgrade wie Restriktionen.
e	☒	a, b und d.
f	❑	alle Antworten treffen zu.

9.	Der Wert des angepassten R^2	
a	❑	liegt zwischen 0 und 1.
b	☒	hängt von der Anzahl der geschätzten Parameter ab.
c	❑	steigt, wenn zusätzliche Regressoren berücksichtigt werden.
d	❑	ist größer als der Wert des R^2.
e	❑	a und b.
f	❑	alle Antworten treffen zu.

10.	Das Gauss-Markov Theorem	
a	❑	basiert auf der Normalverteilungsannahme.
b	❑	macht eine Aussage zur Konsistenz der KQ Schätzer.
c	❑	macht eine Aussage zur Effizienz der KQ Schätzer.
d	❑	basiert auf der Annahme homoskedastischer Störterme.
e	☒	c und d.
f	❑	a und c.

11.	Zum Testen mehrerer linearer Restriktionen	
a	❑	werden t-Tests verwendet.
b	❑	werden Chow-Tests verwendet.
c	❑	werden F-Tests verwendet.
d	❑	werden LM-Tests verwendet.
e	☒	b, c und d.
f	❑	alle Antworten treffen zu.

12.	Die Präzision der Schätzung des Steigungsparameters β_1	
a	❑	steigt, wenn die Variation in y zunimmt.
b	☒	steigt, wenn die Variation in x_1 zunimmt.
c	❑	sinkt, wenn der Stichprobenumfang steigt.
d	❑	variiert mit dem Bevölkerungswert der Regressionskonstante.
e	❑	b und d.
f	❑	a und d.

13.	Die Eigenschaft der Konsistenz	
a	❑	kann unter schwächeren als den Gauss-Markov-Annahmen nachgewiesen werden.
b	❑	ist eine asymptotische Eigenschaft.
c	❑	gilt nur für verzerrte Schätzer.
d	❑	setzt voraus, dass ein Schätzer normalverteilt ist.
e	☒	a und b
f	❑	a, b und c.

14.	Werden irrelevante Variablen ins Modell aufgenommen, so	
a	❑	sinkt das multiple Bestimmtheitsmaß.
b	☒	verringert sich die Effizienz der Parameterschätzer.
c	❑	werden verzerrte Parameterschätzer ausgewiesen.
d	❑	erhöht sich die Effizienz der Parameterschätzer.
e	❑	b, c und d.
f	❑	b und c.

15.	Typ I-Fehlerwahrscheinlichkeiten	
a	☒	werden durch p-Werte beschrieben.
b	❑	sind für t- aber nicht für F-Tests relevant.
c	❑	sind bei einseitigen Tests nur halb so groß wie bei zweiseitigen Tests.
d	❑	beschreiben die Wahrscheinlichkeit, die Nullhypothese zu verwerfen.
e	❑	b und c.
f	❑	c und d.