Bachelorprüfung

Fach: Praxis der empirischen Wirtschaftsforschung

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

Name, Vorname	
Matrikelnr.	
E-Mail	
Studiengang	
Semester	
Datum	
Raum, Platznr.	
Letztes Semester freiwillige Hausarbeit abgegeben ❑ **Ja** ❑ **Nein**	
Unterschrift	

<u>Vorbemerkungen:</u>

Anzahl der Aufgaben: Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen.

Bewertung: Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.

Erlaubte Hilfsmittel:

- Formelsammlung (ist der Klausur beigefügt)
- Tabellen der statistischen Verteilungen (sind der Klausur beigefügt)
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1: **[14 Punkte]**

Sie interessieren sich für die Determinanten von Kindergesundheit, gemessen mit dem Geburtsgewicht des Kindes. Ihnen liegt ein Datensatz mit Informationen zu 1185 Kindern vor. Ihnen stehen folgende Informationen zur Verfügung:

weight	Geburtsgewicht in Gramm
male	=1 wenn das Kind männlich ist, =0 wenn weiblich
week	Schwangerschaftswoche der Geburt
smoke	=1 wenn die Mutter jemals geraucht hat, =0 wenn die Mutter niemals geraucht hat
height	Körpergröße des Kindes in cm
age	Alter der Mutter in Jahren

Sie schätzen die folgende Gleichung mit SPSS:

$$\text{weight}_i = \beta_0 + \beta_1 \text{week}_i + \beta_2 \text{male}_i + \beta_3 \text{smoke}_i + \beta_4 \text{height}_i + \varepsilon_i$$

Sie erhalten folgendes Schätzergebnis und unterstellen Sie die Gültigkeit der Gauss-Markov Annahmen:

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	-4895,307	234,722	-20,856	,000
	week	64,875	6,247	10,386	,000
	male	52,113	23,021	2,264	,024
	smoke	-29,390	24,960	-1,177	,239
	height	111,530	4,514	24,710	,000

a. Abhängige Variable: Geburtsgewicht des Kindes in Gramm

a) Interpretieren Sie den geschätzten Koeffizienten für β_3 inhaltlich und statistisch. (2 Punkte)

Wenn die Mutter jemals geraucht hat, so ist das Geburtsgewicht ihres Kindes c.p. im Mittel 29,39 Gramm niedriger als das Gewicht von Kindern, deren Mutter noch nie geraucht hat.

Der Koeffizient ist am 10%-Niveau nicht statistisch signifikant.

b) Wie ändert sich das Geburtsgewicht des Kindes, wenn die Geburt des Kindes 28 Tage zu früh stattfindet? (2,5 Punkte)

$$\Delta weight = \hat{\beta}_1 \times \Delta week = 64{,}875 \times (-4) = -259{,}5$$

Wenn das Kind 28 Tage eher zur Welt kommt, so verringert sich das Geburtsgewicht c.p. im Mittel um 259,5 Gramm.

c) Berechnen Sie das 95% Konfidenzintervall für den geschätzten Parameter $\hat{\beta}_1$ (week) und interpretieren sie es. (3,5 Punkte)

$$95\%\ Konfidenzintervall_{\hat{\beta}_1} = \left[\hat{\beta}_1 \pm c \cdot se\left(\hat{\beta}_1\right)\right]$$

$$c = t_{\frac{0{,}05}{2};\infty} = 1{,}96$$

$$[64{,}875 \pm 1{,}96 \cdot 6{,}247] = [52{,}631; 77{,}1191]$$

Bei wiederholter Stichprobenziehung liegt in 95% der Fälle der wahre Parameter β_1 innerhalb der auf diese Weise bestimmten Konfidenzintervalle.

d) Sie vermuten, dass das Alter der Mutter eine relevante Determinante für das Geburtsgewicht ist. Sie nehmen das Alter linear (*age*) und quadratisch (*age2*) in Ihr Modell auf und erhalten folgende Modellzusammenfassung ihrer Schätzung:

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,760[a]	,578	,576	387,931

a. Einflußvariablen : (Konstante), week, male, smoke, height, age, age2

i) Interpretieren Sie das Bestimmtheitsmaß. (1 Punkt)

57,8% der Variation im Geburtsgewicht wird durch das Modell erklärt.

ii) Testen Sie am 1%-Niveau, ob die beiden Parameter für das Alter der Mutter gemeinsam signifikant sind. Geben Sie Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und ihre Testentscheidung an. (Hinweis: Das Bestimmtheitsmaß der restringierten Schätzung beträgt 0,566) (5 Punkte)

$$weight_i = \beta_0 + \beta_1 week_i + \beta_2 male_i + \beta_3 smoke_i + \beta_4 height_i + \beta_5 age_i + \beta_6 age_i^2 + \varepsilon_i$$

H_0: $\beta_5 = \beta_6 = 0$

H_1: *mindestens ein* $\beta_j \neq 0$ *mit j=5,6*

Teststatistik

$$F = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-k-1)} = \frac{(0{,}578-0{,}566)/2}{(1-0{,}578)/(1185-6-1)} = \frac{0{,}006}{0{,}000358} = 16{,}75$$

Zählerfreiheitsgrade: 2

Nennerfreiheitsgrade: 1178

Kritischer Wert: $F_c = F_{0,01;2;1178} = 4{,}61$

Testentscheidung: Da $F > F_c$ *wird die Nullhypothese verworfen. Das Alter der Mutter hat am 1%-Niveau einen signifikanten Erklärungsgehalt.*

Aufgabe 2: **[15 Punkte]**

Sie interessieren sich für die Einflussfaktoren auf den Stundenlohn. Sie untersuchen eine Zufallsstichprobe mit 44 Beobachtungen von erwerbstätigen Männern. Es liegen Ihnen folgende Informationen vor:

lnw	Logarithmierter Stundenlohn in Euro
educ	Bildung in Jahren
exper	Berufserfahrung in Jahren
exper2	Berufserfahrung quadriert ($exper^2$)
age	Alter in Jahren
age2	Alter quadriert (age^2)

Sie stellen folgendes Modell auf:

$$\mathrm{lnw_i} = \beta_0 + \beta_1 \mathrm{educ_i} + \beta_2 \mathrm{exper_i} + \beta_3 \mathrm{exper_i^2} + \varepsilon_i$$

SPSS liefert Ihnen den folgenden Output:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	1,911	,305	6,272	,000
	educ	,062	,012	5,163	,000
	exper	,004	,027	,152	,880
	exper2	-4,914E-5	,001	-,091	,928

a. Abhängige Variable: lnw

a) Interpretieren Sie den Effekt der Bildung auf den Lohn statistisch und inhaltlich. (2 Punkte)

Erhöht sich die Bildung um ein Jahr, so erhöht sich der Lohn c.p. im Mittel um 6,2%.

Der Koeffizient ist hochsignifikant am 1%-Niveau.

b) Testen Sie am 5%-Niveau, ob die Bildungsrendite signifikant höher als 6% ist. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

Einseitiger t-Test

$H_0: \beta_1 \leq 0{,}06$

$H_1: \beta_1 > 0{,}06$

Teststatistik: $t = \frac{\hat{\beta}_1 - 0{,}06}{se(\hat{\beta}_1)} = \frac{0{,}062 - 0{,}06}{0{,}012} = 0{,}167$

Kritischer Wert: $t_{\alpha;n-k-1} = t_{0{,}05;44-3-1} = 1{,}684$

Entscheidung: Die Nullhypothese kann am 5%-Niveau nicht verworfen werden.

c) Nennen Sie eine weggelassene Variable, die den Koeffizienten $\hat{\beta}_1$ verzerren könnte und begründen Sie, warum in Ihrem Beispiel eine Verzerrung bestehen kann. Nennen Sie die Richtung der Verzerrung, geben Sie an ob β_1 über- oder unterschätzt wird und begründen Sie Ihre Antwort. (3,5 Punkte)

Beispiel: Motivation

Wahres Modell: $lnw_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 Motivation_i + \varepsilon_i$

Könnte $\hat{\beta}_1$ *verzerren, weil Motivation mit den Löhnen und mit der Bildung korreliert sein könnte.*

Die Verzerrung wird bestimmt durch:

$Bias(\hat{\beta}_1) = \beta_4 \frac{Cov(motivation_i, educ_i)}{Var(motivation_i)}$

Die Verzerrung ist positiv, wenn

$Cov(motivation_i, educ_i) > 0$ *und* $\beta_4 > 0$ *, d.h.* β_1 *würde überschätzt.*

d) Ein Kommilitone vermutet, dass zusätzlich zum Effekt der Berufserfahrung auch das Alter einen entscheidenden Einfluss auf die Löhne hat. Sie nehmen das Alter linear und quadratisch in Ihr Modell auf und erhalten folgenden Output:

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	1,912	,905	2,113	,041
	educ	,061	,030	2,033	,049
	exper	,004	,068	,058	,477
	exper2	-4,914E-5	,001	-,049	,480
	age	,145	,090	1,623	,111
	age2	-,001	,001	-1,590	,119

a. Abhängige Variable: lnw

i) Vergleichen Sie den Output mit der Schätzung des Ursprungsmodells. Welche Muster stellen Sie fest? (Hinweis: Sie müssen weder inhaltlich noch statistisch interpretieren) (1 Punkt)

Höhere Standardfehler als in der ersten Schätzung bei gleichbleibenden Koeffizienten (die Koeffizienten sind nicht mehr signifikant).

ii) Welches Problem ist durch die Aufnahme des Alters in die Regression entstanden? Welche Auswirkungen hat dies auf Präzision und Unverzerrtheit der Schätzung? (3 Punkte)

Multikollinearität

Multikollinearität führt zu ineffizienten Schätzern, die Präzision sinkt.

Unter Multikollinearität ist KQ immer noch unverzerrt

iii) Wie könnte man das Problem im Rahmen des gegebenen Modells beheben? (1 Punkt)

Erhöhung des Stichprobenumfangs

Aufgabe 3: **[19 Punkte]**

Sie interessieren sich für die Determinanten der Pendelzeit zur Arbeitsstätte. Der Datensatz beinhaltet folgende Angaben zu 187.552 erwerbstätigen Personen:

Pendelzeit	Zeitaufwand für den Weg zur Arbeitsstätte (Minuten pro Tag)
Einkommen	=1, wenn bis 1.000€; =2, wenn 1.000-2.000€; =3, wenn 2.000€ und mehr (pro Monat)
Alter	Alter (in Jahren)
Mann	=1, wenn männlich, =0 wenn weiblich
Kinder	Anzahl Kinder

a) Sie stellen folgendes Modell auf:

$$\text{Pendelzeit}_i = \beta_0 + \beta_1 \text{Alter}_i + \beta_2 \text{Kinder}_i + \beta_3 \text{Mann}_i + u_i$$

Sie wollen mit Hilfe eines Regressionsmodells prüfen, ob sich die Pendelzeit für Personen in den verschiedenen Einkommensgruppen statistisch signifikant unterscheidet. Geben Sie die Regressionsgleichung für eine Modellspezifikation an, die Ihnen das gewünschte Ergebnis liefert. Definieren Sie gegebenenfalls neue notwendige Variablen. (3 Punkte)

Dummy-Variablen für 2 Ausprägungen der Variable Monatseinkommen ins Modell einfügen:

$$Pendelzeit_i = \beta_0 + \beta_1 Alter_i + \beta_2 Kinder_i + \beta_3 Mann_i + \beta_4 Einkommen2_i + \beta_5 Einkommen3_i + u_i$$

wobei

$Einkommen2$, *Dummyvariable:= 1, wenn Einkommen=2, sonst =0*

$Einkommen3$, *Dummyvariable:= 1, wenn Einkommen=3, sonst =0*

b) Anschließend schätzen Sie das folgende Modell:

$$\text{Pendelzeit}_i = \beta_0 + \beta_1 \text{Alter}_i + \beta_2 \text{Kinder}_i + \beta_3 \text{Mann}_i + \beta_4 \text{Mann}_i \cdot \text{Kinder}_i + u_i$$

Sie erhalten den folgenden Output:

$$\widehat{\text{Pendelzeit}}_i = 26{,}80 - 0{,}10\ \text{Alter}_i - 1{,}41\ \text{Kinder}_i + 2{,}10\ \text{Mann}_i + 1{,}23\ \text{Mann}_i \cdot \text{Kinder}_i$$

i. Berechnen Sie den Standardfehler für $\hat{\beta}_3$, gegeben dass die Teststatistik für einen Signifikanztest ($H_0: \beta_3 = 0$) den Wert t_3 =19,92 annimmt. (1 Punkt)

$$se(\hat{\beta}_3) = \frac{\hat{\beta}_3}{t_3} = \frac{2{,}10}{19{,}92} = 0{,}105$$

ii. Wie ändert sich die mittlere Pendelzeit pro zusätzliches Kind für Männer in diesem Modell? Wie hoch ist dieser Effekt für Frauen? (2 Punkte)

$$\left.\frac{\widehat{Pendelzeit}}{\partial Kind}\right|_{Mann=1} = \hat{\beta}_2 + \hat{\beta}_4 = -1{,}41 + 1{,}23 = -0{,}18$$

$$\left.\frac{\widehat{Pendelzeit}}{\partial Kind}\right|_{Mann=0} = \hat{\beta}_2 = -1{,}41$$

Mit jedem weiteren Kind sinkt die mittlere Pendelzeit für Männer c.p. um 0,18 Minuten/Tag.

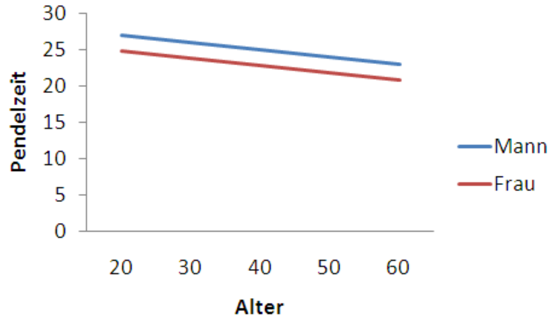
Für Frauensinkt die mittlere Pendelzeit pro zusätzliches Kind c.p. um 1,41 Minuten/Tag.

iii. Berechnen Sie die erwartete Pendelzeit für einen 35-jährigen Mann mit 2 Kindern. (2 Punkte)

$$\widehat{Pendelzeit}_i = 26{,}80 - 0{,}10 \cdot 35 - 1{,}41 \cdot 2 + 2{,}10 + 1{,}23 \cdot 2 = 25{,}04\ [Minuten]$$

Für einen 35-jährigen Mann mit 2 Kindern beträgt die erwartete Pendelzeit 25,04 Minuten

iv. Stellen Sie den Zusammenhang zwischen Alter und Pendelzeit für Männer und Frauen ohne Kinder graphisch dar. Beschriften Sie die Achsen in Ihrem Diagramm. (2 Punkte)



v. Welche Änderungen würden sich für die geschätzten Koeffizienten von β_1 und β_3 ergeben, wenn Sie die Geschlechtsvariable wie folgt kodiert hätten:

Frau =1, wenn weiblich, =0 wenn männlich. (2 Punkte)

$\hat{\beta}_1$: *Keine Änderung*

$\hat{\beta}_3$: *Änderung des Vorzeichens*

c) Ein Kollege schlägt Ihnen vor, für das vereinfachte Modell

$\text{Pendelzeit}_i = \beta_0 + \beta_1 \text{Alter}_i + \beta_2 \text{Kinder}_i + u_i$ zu prüfen, ob sich alle Parameter im Modell für Männer und Frauen unterscheiden.

i. Sie schätzen das Modell getrennt für Männer und Frauen. Welchen Test (1% Niveau) können Sie verwenden um die Vermutung Ihres Kollegen zu testen? Nennen Sie Testverfahren, Null- und Alternativhypothese, Freiheitsgrade, kritischen Wert und die Entscheidungsregel. (4,5 Punkte)

Chow-Test (auch zulässig: F-Test):

$H_0: \beta_{0,Mann=0} = \beta_{0,Mann=1}$ *und* $\beta_{1,Mann=0} = \beta_{1,Mann=1}$ *und* $\beta_{2,Mann=0} = \beta_{2,Mann=1}$

$H_1: H_0$ *gilt nicht*

Freiheitsgrade: Zähler=q=3 (Anzahl der Restriktionen),

Nenner= n–2(k+1)= 187552 - 2 (2+1) = 187546

$c = F_{1\%;3;187546} = F_{1\%;3;\infty} = 3{,}78$

Wenn der Wert der Teststatistik größer ist als $3{,}78$ *dann folgt, dass die* H_0 *verworfen wird und somit die Koeffizienten für Frauen und Männer signifikant verschieden sind.*

ii. Wie könnten Sie die Vermutung Ihres Kollegen prüfen, wenn Sie nicht getrennt für Männer und Frauen schätzen? Geben Sie die Regressionsgleichung für eine Modellspezifikation an, die Ihnen das gewünschte Ergebnis liefert, nennen Sie Testverfahren, Null- und Alternativhypothese. (2,5 Punkte)

-Schätzung eines vollständig interagierten Modells

$Pendelzeit_i = \beta_0 + \beta_1 Alter_i + \beta_2 Kinder_i + \beta_3 Mann_i + \beta_4 Alter_i \cdot Mann_i + \beta_5 Kinder_i \cdot Mann_i + u_i$ *(unrestringiertes Modell)*

-gemeinsame Schätzung des restringierten Modells für Männer und Frauen

$Pendelzeit_i = \beta_0 + \beta_1 Alter_i + \beta_2 Kinder_i + u_i$ *(restringiertes Modell)*

-F-Test auf gemeinsame Signifikanz

-Hypothesen:

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

H_1: *mindestens ein* $\beta_j \neq 0$ *mit j=3, 4, 5*

Aufgabe 4: **[27 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	Eine standardnormalverteilte Zufallsvariable hat den Mittelwert 0 und eine Standardabweichung von 1.
F	Bei der Alternativhypothese $H_1: \beta_1 \neq 3$ wird ein rechtsseitiger t-Test durchgeführt.
F	Ein positives quadriertes Residuum $\hat{u}_i^2$ nach einer KQ-Schätzung bedeutet, dass y_i unterschätzt wird.
F	Bei 100 Nennerfreiheitsgraden können F-Tests nicht durchgeführt werden.
F	Das angepasste Bestimmtheitsmaß $\overline{R}^2$ kann nicht verwendet werden, um genestete Modelle miteinander zu vergleichen.
W	Die Varianz von $\hat{\beta}$ kann nicht negativ werden.
F	Die Abbildung konkaver Zusammenhänge ist im linearen Regressionsmodell nicht möglich.
W	Bei perfekter Multikollinearität ist eine KQ-Schätzung nicht durchführbar.
F	Unter Homoskedastie sind die Parameterschätzer verzerrt.
F	Die t-Verteilung nähert sich mit steigender Stichprobengröße der F-Verteilung an.
F	Der geschätzte Koeffizient $\hat{\beta}$ lässt sich aus dem Produkt aus p-Wert und Standardfehler von $\hat{\beta}$ berechnen.
F	Ist die Varianz des Fehlerterms u für alle Beobachtungen identisch, so bezeichnet man ihn als heteroskedastisch.
W	Das Bestimmtheitsmaß (R^2) kann zur Berechnung der F-Statistik genutzt werden.
W	Bei $R^2 = 1$ gilt $\hat{u}_i = 0$ für jedes i.
W	Bei diskreten Zufallsvariablen X beschreibt die Wahrscheinlichkeitsdichtefunktion f(X=x) für jedes x die Wahrscheinlichkeit, mit der X den Wert x annimmt.
W	Je größer α, desto kleiner das Konfidenzintervall eines Schätzers.
F	Konsistenz erfordert mittlere bedingte Unabhängigkeit $E(u\|x) = 0$.
F	Im einfachen linearen Modell $y_i = \beta_0 + \beta_1 x_{1i} + u_i$ wird die Zielfunktion $\sum_{i=1}^{n} u_i^2$ maximiert.
W	F-Tests können verwendet werden, um Hypothesen betreffend einzelner Parameter zu testen.
W	Das Auslassen relevanter Variablen kann zu Verzerrungen der geschätzten Parameter führen.
F	Im linearen Modell ist die geschätzte Varianz des KQ-Schätzers für die Konstante gleich Null.
F	Durch Logarithmierung der abhängigen Variable gewinnen Ausreißerbeobachtungen an Bedeutung.
W	Wenn zwei Variablen x und z statistisch unabhängig sind, dann gilt $\mathrm{Cov}(x, z) = 0$.
F	Konsistenz und asymptotische Effizienz sind für $n \rightarrow 0$ definiert.
W	Wird nach einer KQ-Schätzung ein 95%-Konfidenzintervall (KI) für β_2 berechnet, so wird die Breite des KI durch die Varianz von $\hat{\beta}_2$ beeinflusst.
W	Wird das Modell $\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ auf Gesamtsignifikanz getestet, so lautet das restringierte Modell $\log(y) = \beta_0 +$ u.
W	Die Parameterschätzer von Dummyvariablen geben an, um wie viel sich die Achsenabschnittsparameter bei Teilgruppen unterscheiden.
F	Der Bevölkerungsparameter β_k ist eine Zufallsvariable.
W	Das Modell $y = \beta_0 + \beta_1 \sqrt{x} + u$ ist linear in den Parametern.
F	Ein $R^2 = 1$ kann ein Hinweis auf Multikollinearität im einfachen Regressionsmodell sein.
W	Wenn ein perfekter linearer Zusammenhang vorliegt, nimmt der Korrelationskoeffizient ρ den Wert |-1| an.
F	Perfekte Multikollinearität liegt vor, wenn eine Variable linear und quadratisch berücksichtigt wird.
W	Wird $y = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_3 + u$ geschätzt, so ergeben sich drei Bedingungen erster Ordnung.
F	Schätzt man eine Regressionsgleichung mit einem Interaktionsterm, dann muss der Interaktionsterm signifikant sein, um unverzerrte Schätzergebnisse zu erhalten.
F	Die BLUE-Eigenschaft gilt nicht, wenn die Residuen nicht normalverteilt sind.
W	Unter den Gauss-Markov Annahmen (MLR.1–MLR.5) ist der KQ-Schätzer asymptotisch effizient.
F	Die dummy variable trap entsteht durch Aufnahme einer irrelevanten Dummyvariable.
W	Im Modell $y = \beta_0 + \beta_1 x + u$ ist $\hat{\beta}_1 < 0$ wenn Cov(x,y)<0.
F	Ein Schätzer W des Parameters θ heißt unverzerrt, wenn für alle θ gilt: $plim(W) = \theta$.
W	Eine Bernoulli Zufallsvariable nimmt Werte von 0 oder 1 an.
W	Ein exogener Regressor ist mit dem Störterm unkorreliert.
F	Im einfachen Modell kann der KQ-Schätzer nicht berechnet werden, wenn $\bar{x} = 0$.

W	Paneldaten enthalten für eine Beobachtungseinheit wiederholte Messungen.
W	Die Eigenschaft der Unverzerrtheit gilt unabhängig von der Stichprobengröße.
F	Bei zwei unverzerrten Schätzern $\hat{\beta}_1$ und $\tilde{\beta}_1$ für einen unbekannten Bevölkerungsparameter β_1 ist $\tilde{\beta}_1$ effizient, wenn $E(\tilde{\beta}_1) < E(\hat{\beta}_1)$.
W	Eine Nullhypothese, die am 5% Niveau verworfen wurde, wird am 10% Niveau verworfen.
F	Bei Querschnittsdaten variieren die Schätzergebnisse mit der Reihenfolge der Beobachtungen.
F	Das R^2 fällt, wenn in die Schätzung eine irrelevante Variable zusätzlich aufgenommen wird.
W	Die Dichtefunktion der F-Verteilung ist rechtsschief.
F	Je höher das R^2, desto niedriger ist die Wahrscheinlichkeit beim F-Test auf Gesamtsignifikanz die H_0 zu verwerfen.
W	Durch Umskalierung der abhängigen Variable ändert sich auch der Wert der Konstanten.
F	Die Entwicklung der Arbeitslosenquote in 27 EU-Ländern in den Jahren 1995-2000 ist ein Beispiel für einen Querschnittsdatensatz.
W	Die Gauss-Markov Annahmen werden verletzt, wenn die funktionale Form falsch spezifiziert wurde.
W	Bei einem R^2 von 0,1 sind präzise Parameterschätzungen möglich.

Aufgabe 5: **[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.	Im Modell $\log(y)=\beta_0+\beta_1\log(x)+u$	
a	☑	wird β_1 als Elastizität interpretiert.
b	❑	wird β_1 als Semielastizität interpretiert.
c	❑	gibt β_1 den marginalen Effekt von x auf y an.
d	❑	gibt $\beta_1/100$ den marginalen Effekt von y auf x an.
e	❑	resultiert eine 1% Änderung in y aus einer 1% Änderung in x.
f	❑	a und e

2.	Für eine unverzerrte Schätzung des wahren Bevölkerungsparameters mittels KQ ist erforderlich,	
a	❑	dass der Erwartungswert der Konstante Null ist.
b	❑	dass die Störterme normalverteilt sind.
c	❑	dass der Erwartungswert des Störterms abhängig von den erklärenden Variablen ist.
d	❑	dass die Stichprobe gegen unendlich geht.
e	❑	b und d
f	☑	keine der genannten Antworten

3.	Gepoolte Querschnittsdaten	
a	❑	enthalten wiederholte Messungen für jede Beobachtungseinheit.
b	❑	sind Kombinationen von Zeitreihenerhebungen zu verschiedenen Erhebungszeitpunkten.
c	☑	enthalten für eine Beobachtungseinheit eine Messung.
d	❑	werden auch als Paneldaten bezeichnet.
e	❑	b und d
f	❑	keine der genannten Antworten

4.	Die Annahme $x \sim \text{i.i.d.}(\mu, \sigma^2)$ impliziert unter anderem, dass	
a	❑	$E(\mu) = 0$
b	❑	$E(\sigma^2) = 0$
c	❑	$E(\mu) = x$
d	☑	$Var(x) = \sigma^2$
e	❑	a und d
f	❑	keine der Antworten

5.	Interaktionsterme von stetigen Variablen mit binären Variablen werden verwendet,	
a	❑	um die dummy variable trap zu beheben.
b	❑	um unterschiedliche Achsenabschnittsparameter für Teilgruppen zu bestimmen.
c	☑	um unterschiedliche Steigungsparameter für Teilgruppen zu bestimmen.
d	❑	um Effekte qualitativer Variablen auf Signifikanz zu testen.
e	❑	b und c
f	❑	b und d

6.	Wenn $Cov(x, u) = 0$, dann	
a	❑	ist $E(u\|x) = E(u) = 0$.
b	❑	ist der KQ-Schätzer unverzerrt.
c	❑	ist die ceteris-paribus Interpretation in kleinen Stichproben legitim.
d	☑	besteht kein linearer Zusammenhang zwischen den erklärenden Variablen und den Störtermen.
e	❑	b und d
f	❑	a und c

7.	Die t-Verteilung	
a	❑	konvergiert mit steigender Beobachtungszahl gegen die Chi-Quadrat-Verteilung.
b	❑	konvergiert mit steigender Beobachtungszahl gegen die Poisson-Verteilung.
c	☑	konvergiert mit steigender Beobachtungszahl gegen die Standardnormalverteilung.
d	❑	ist eine asymmetrische Verteilung.
e	❑	a und d
f	❑	keine der Antworten

8.	Beim Signifikanzniveau α gilt für den der Typ 1 Fehler:	
a	❑	er wird begangen, wenn eine unzutreffende H_0 nicht verworfen wird.
b	❑	er wird begangen, wenn eine zutreffende H_0 verworfen wird.
c	❑	er tritt ein mit der Wahrscheinlichkeit 1-α.
d	❑	er tritt ein mit der Wahrscheinlichkeit α.
e	❑	a und d
f	☑	b und d

9	Beim Umskalieren der abhängigen Variable im Modell $y = \beta_0 + \beta_1 x + u$ ändern sich	
a	❑	die geschätzten Koeffizienten nicht.
b	❑	die Standardfehler der Koeffizienten.
c	❑	die t-Statistiken.
d	❑	die Konfidenzintervalle für beide Koeffizienten.
e	☑	b und d
f	❑	a und c

10.	Die SST (Sum of squares total)	
a	❑	wird verwendet, um die Stichprobenvarianz der abhängigen Variable zu berechnen.
b	❑	hängt von der Anzahl der Restriktionen im Modell ab.
c	❑	wird mittels einer KQ-Schätzung bestimmt.
d	❑	ist für genestete Modelle identisch.
e	☑	a und d
f	❑	b und c

11.	Die Eigenschaft der Unverzerrtheit	
a	❑	ist auch für kleine Stichproben definiert.
b	❑	ist nur für große Stichproben definiert.
c	❑	ist im linearen Regressionsmodell eine asymptotische Eigenschaft.
d	❑	gilt unter MLR.1–MLR.4.
e	❑	b und c
f	☑	a und d

12.	Wenn zwei genestete Modelle (1) und (2) verglichen werden, sollte das Modell (2) bevorzugt werden, wenn	
a	☐	$R_2^2 < R_1^2$.
b	☐	im Modell (2) alle Koeffizienten signifikant sind.
c	☐	$\bar{R}_1^2 > \bar{R}_2^2$.
d	☑	$\bar{R}_2^2 > \bar{R}_1^2$.
e	☐	b und c
f	☐	keine der Antworten

13.	Der KQ-Schätzer	
a	☐	minimiert die Summe der quadrierten vertikalen Abweichungen von der Regressionsgeraden.
b	☐	minimiert die Residuenquadratsumme.
c	☐	minimiert den unerklärten Teil der Streuung in der abhängigen Variable.
d	☐	maximiert R^2.
e	☐	a und b.
f	☑	alle Antworten sind zutreffend.

14.	Dummy-Variablen	
a	☐	nehmen nur Werte zwischen 0 und 1 an.
b	☑	nehmen typischerweise die Ausprägungen 0 und 1 an.
c	☐	können mehr als zwei Ausprägungen haben.
d	☐	sind für qualitative Informationen nicht definiert.
e	☐	a, b, c und d.
f	☐	keine der Antworten.

15.	Der Zusammenhang zwischen zwei Variablen	
a	☐	wird mit dem Standardfehler gemessen.
b	☑	wird mit dem Korrelationskoeffizienten als linearer Zusammenhang gemessen.
c	☐	wird nur mit experimentellen Daten verlässlich gemessen.
d	☐	wird mit der Standardabweichung gemessen.
e	☐	a und b.
f	☐	a und d.