Bachelorprüfung

Fach: Praxis der empirischen Wirtschaftsforschung

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

Name, Vorname			
Matrikelnr.			
E-Mail			
Studiengang			
Semester			
Datum			
Raum, Platznr.			
Freiwillige Hausarbeit geschrieben		❑ **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: **[13 Punkte]**

Sie interessieren sich für die Determinanten des Arbeitsangebots von Müttern mit Kindern im Vorschulalter. Der Datensatz beinhaltet folgende Angaben zu 3215 Frauen:

Stunden	Arbeitszeit (Stunden pro Woche)
Alter	Alter (in Jahren)
AlterK	Alter des jüngsten Kindes (in Jahren)
Oma	=1, wenn Großmutter wohnhaft im Haushalt, =0 sonst
Arblos	regionale Arbeitslosenquote (in Prozent)

Sie unterstellen die Gültigkeit der Gauss-Markov Annahmen und schätzen folgendes Modell:

$$\mathrm{Stunden_i} \ = \beta_0 + \beta_1 \mathrm{Alter_i} + \beta_2 \mathrm{Alter_i^2} + \beta_3 \mathrm{AlterK_i} + \beta_4 \mathrm{Oma_i} + \beta_5 \mathrm{Arblos_i} + \mathrm{u_i}$$

Sie schätzen das Modell mit SPSS und erhalten den folgenden Output:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	-7,227	6,042	-1,196	,232
	Alter	,306	,358	**?**	,393
	Alter^2	-,004	,005	-,667	,505
	AlterK	2,387	,127	18,822	,000
	Oma	1,599	1,217	1,314	,189
	Arblos	**?**	,057	8,917	,000

a. Abhängige Variable: Stunden

a) Berechnen Sie die beiden fehlenden Werte („?“). (2 Punkte)

$$t_1 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0{,}306}{0{,}358} = 0{,}854$$

$$\hat{\beta}_5 = t_3 \cdot se(\hat{\beta}_5) = 8{,}917 \cdot 0{,}57 = 0{,}511$$

b) Interpretieren Sie statistisch und inhaltlich den geschätzten Parameter für *Oma* ($\hat{\beta}_4$). (2 Punkte)

Mütter, in deren Haushalt eine Großmutter wohnt, arbeiten c.p. im Durchschnitt 1,599 Stunden pro Woche mehr als Mütter in deren Haushalt keine Großmutter wohnt. Der Parameter ist auf dem 10%-Niveau nicht signifikant von Null verschieden.

c) In welchem Alter der Mutter ist ihre Arbeitszeit maximal? (2 Punkte)

Marginaler Effekt von Alter : $\frac{\partial \widehat{Stunden}}{\partial Alter} = 0{,}306 - 2 \cdot 0{,}004 \cdot Alter = 0$

$Alter_{max} = \frac{0{,}306}{2 \cdot 0{,}004} = 38{,}25$

Im Alter von 38 Jahren bieten die Mütter c.p. die meiste Arbeit an.

d) Sie möchten prüfen, ob sich der Effekt der regionalen Arbeitslosenquote für Haushalte mit und ohne im Haushalt wohnende Großmutter statistisch signifikant unterscheidet. Geben Sie die Regressionsgleichung für eine Modellspezifikation an, die Ihnen das gewünschte Ergebnis liefert. (2 Punkte)

$$Stunden_i \ = \beta_0 + \beta_1 Alter_i + \beta_2 Alter_i^2 + \beta_3 AlterK_i + \beta_4 Oma_i + \beta_5 Arblos_i + \beta_6 Oma_i \cdot Arblos_i + u_i$$

e) Berechnen Sie unter der Annahme, dass es in der Region keine Arbeitslosigkeit gibt, die erwartete Arbeitszeit für eine 30-jährige Mutter eines 6-jährigen Kindes, wenn keine Großmutter im Haushalt wohnt. (2 Punkte)

$$\widehat{Stunden} = -7{,}227 + 0{,}306 \cdot 30 - 0{,}004 \cdot 30^2 + 2{,}387 \cdot 6 = 12{,}675$$

In diesem Fall beträgt die erwartete Arbeitszeit 12,675 Stunden

f) Das Bestimmtheitsmaß R^2 beträgt in Ihrem Ausgangsmodell 0,1392.

i) Interpretieren Sie diesen Wert. (1 Punkt)

13,92% der Variation der Arbeitszeit können durch das Modell erklärt werden.

ii) Berechnen Sie das angepasste Bestimmtheitsmaß $\overline{R^2}$. (2 Punkte)

$$\overline{R^2} = 1 - (1 - R^2)\frac{n-1}{n-k-1} = 1 - (1 - 0{,}1392) \cdot \frac{3215 - 1}{3215 - 5 - 1} = 0{,}1378$$

Das angepasste Bestimmtheitsmaß beträgt 0,1378

Aufgabe 2: **[18 Punkte]**

Sie wollen untersuchen, ob privat Versicherte oder gesetzlich Versicherte höhere Kosten für das Gesundheitssystem verursachen. Ihnen liegen Daten des Sozio-oekonomischen Panels (SOEP) mit 66794 Beobachtungen vor. Folgende Informationen sind gegeben:

docvisits	Anzahl der Arztbesuche in den letzten 3 Monaten
gkv	=1 wenn gesetzlich versichert, =0 wenn privat versichert
income	durchschnittliches Bruttomonatseinkommen in 1000 Euro
badhealth	Gesundheitszustand auf einer Skala von 1 bis 5, wobei 1=keinerlei Einschränkungen und 5= sehr schlecht
female	=1 wenn weiblich, =0 sonst
age	Alter in Jahren

Sie stellen folgendes Modell auf:

$$\mathrm{docvisits_i} = \beta_0 + \beta_1 \mathrm{gkv_i} + \mathrm{u_i}$$

a) Sie interessieren sich für den kausalen Effekt der Versicherung auf die Anzahl der Arztbesuche. Sie vermuten, dass durch das weggelassene Einkommen in der Schätzgleichung der Effekt der GKV verzerrt geschätzt wird. Unter welchen Bedingungen würde das Weglassen des Einkommens nicht zu einer Verzerrung des geschätzten Koeffizienten $\widehat{\beta}_1$ führen? (2 Punkte)

- *Wenn das Einkommen keinen Einfluss auf die Arztbesuche hat*
- *und/oder* $\mathrm{Cov(gkv, income) = 0}$.

b) Ein Kommilitone vermutet, dass auch das Weglassen des Gesundheitszustands eine verzerrte Schätzung des Koeffizienten $\widehat{\beta}_1$ bewirkt. Sie wollen die Richtung der Verzerrung durch die nicht aufgenommene Gesundheit feststellen und lassen sich folgende Korrelationen ausgeben:

Korrelationen

		Badhealth	gkv	docvisits
Badhealth	Korrelation nach Pearson	1	,044	,321
gkv	Korrelation nach Pearson	,044	1	-,019
docvisits	Korrelation nach Pearson	,321	-,019	1

In welche Richtung wäre der geschätzte Effekt für β_1 durch die nicht aufgenommene Variable *badhealth* verzerrt, wenn Sie das Modell schätzen? Begründen Sie Ihre Antwort. (3 Punkte)

Wenn das wahre Modell lautet: $\text{docvisits}_i = \beta_0 + \beta_1 \text{gkv}_i + \beta_2 \text{badhealth}_i + u_i$

Dann gilt: $E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\text{Cov}(\text{gkv}_i, \text{badhealth}_i)}{\text{Var}(\text{gkv}_i)}$

Die Verzerrung von $\hat{\beta}_1$ *ist gegeben durch:* $\text{Bias}(\hat{\beta}_1) = \beta_2 \frac{\text{Cov}(\text{gkv}_i, \text{badhealth}_i)}{\text{Var}(\text{gkv}_i)}$

Es liegt eine Überschätzung vor, wenn die Verzerrung positiv ist. Dies ist der Fall, wenn beide Elemente der Verzerrung das gleiche Vorzeichen haben.

Wenn schlechte Gesundheit bewirkt, dass mehr zum Arzt gegangen wird, dann ist β_2 positiv.

Cov(gkv,badhealth)>0, d.h. schlechte Gesundheit ist positiv mit einer gesetzlichen Versicherung korreliert (schlechte Risiken in der GKV)

Wenn somit alle Komponenten der Verzerrung positiv sind, ist $\text{Bias}(\hat{\beta}_1) > 0$ *und* β_1 *wird überschätzt.*

c) Sie erweitern Ihr Modell zu:

$$\text{docvisits}_i = \beta_0 + \beta_1 \text{gkv}_i + \beta_2 \text{income}_i + \beta_3 \text{badhealth}_i + \beta_4 \text{age}_i + \beta_5 \text{female}_i + u_i$$

Sie schätzen das Modell mit SPSS und erhalten den folgenden Output:

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	99216,6	5	19843,3	1583,0	,000[a]
	Nicht standardisierte Residuen	837214,5	66788	12,5		
	Gesamt	936431,1	66793			

a. Einflußvariablen : (Konstante), female, health, gkv, age, income

b. Abhängige Variable: docvisits

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		T	Sig.
		RegressionskoeffizientB	Standardfehler		
1	(Konstante)	-,113	,080	-1,417	,156
	gkv	-,388	,039	-10,036	,000
	income	,009	,008	1,182	,237
	badhealth	1,427	,016	86,769	,000
	age	-,007	,001	-5,062	,000
	female	,321	,029	10,993	,000

a. Abhängige Variable: docvisits

Interpretieren Sie den geschätzten Koeffizienten für β_1 statistisch und inhaltlich. (2 Punkte)

Die mittlere Anzahl der Arztbesuche ist c.p. bei gesetzlich Versicherten um 0,388 niedriger als bei privat Versicherten.

Der Koeffizient ist statistisch hochsignifikant am 1%-Niveau.

d) Berechnen und interpretieren Sie inhaltlich den geschätzten Effekt einer Verringerung des Einkommens um 2000 Euro auf die Anzahl der Arztbesuche. (2 Punkte)

$\Delta\widehat{docvisits_i} = \hat{\beta}_2 \Delta income$

$\Delta income = -2$

$\Delta\widehat{docvisits_i} = 0{,}009 \cdot (-2) = -0{,}018$

Wenn sich das Einkommen um 2000 Euro verringert, so sinkt die Anzahl der erwarteten Arztbesuche c.p. um 0,018 Besuche.

e) Welcher Koeffizient hätte sich für β_2 ergeben, wenn das Einkommen in 1000 Dollar gemessen würde? Zeigen Sie Ihren Rechenweg. (Hinweis: 1 Euro= 1,4389 US Dollar) (3 Punkte)

$$\widehat{docvisits_i} = \hat{\beta}_0 + \hat{\beta}_1 gkv_i + \hat{\beta}_2 \frac{income_i \cdot 1{,}4389}{1{,}4389} + \hat{\beta}_3 badhealth_i + \hat{\beta}_5 age_i + \hat{\beta}_6 female_i$$

$$\widehat{docvisits_i} = \hat{\beta}_0 + \hat{\beta}_1 gkv_i + \frac{\hat{\beta}_2}{1{,}4389} income_i \cdot 1{,}4389 + \hat{\beta}_3 badhealth_i + \hat{\beta}_5 age_i + \hat{\beta}_6 female_i$$

Neues $\tilde{\beta}_1 = \frac{\hat{\beta}_2}{1{,}4389} = \frac{0{,}009}{1{,}4386} = 0{,}006$

f) Wie hoch ist der Erklärungsgehalt des Modells? (2 Punkte)

$R^2 = \frac{SSE}{SST} = \frac{99216{,}6}{936431{,}1} = 0{,}106$

10,6% der Variation in den Arztbesuchen wird durch das Modell erklärt.

g) Berechnen Sie das 99% Konfidenzintervall für den Parameter $\hat{\beta}_1$ (gkv), zeigen Sie Ihren Rechenweg und interpretieren Sie das Ergebnis. (4 Punkte)

99% $\text{Konfidenzintervall}_{\hat{\beta}_1} = [\hat{\beta}_1 \pm c \cdot se(\hat{\beta}_1)]$

$c = t_{\frac{0{,}01}{2};\infty} = 2{,}576$

$[-0{,}388 \pm 2{,}576 \cdot 0{,}039] = [-0{,}488; -0{,}288]$

Bei wiederholter Stichprobenziehung liegt in 99% der Fälle der wahre Parameter innerhalb der auf diese Weise bestimmten Konfidenzintervalle.

a)

Aufgabe 3: **[20 Punkte]**

Sie interessieren sich für die Determinanten der subjektiven Lebenszufriedenheit. Der Datensatz beinhaltet folgende Angaben zu 5036 Personen:

satlife	Subjektive Lebenszufriedenheit, gemessen auf einer Skala von 1=sehr unzufrieden bis 10=sehr zufrieden
lnw	logarithmierter Stundenlohn
sathealth	Subjektive Zufriedenheit mit der Gesundheit, gemessen auf einer Skala von 1=sehr unzufrieden bis 10=sehr zufrieden
educ	Anzahl der Schuljahre
$educ^2$	quadrierte Anzahl der Schuljahre

Sie schätzen zunächst folgendes Modell:

Modell 1: $satlife_i = \beta_0 + \beta_1 lnw_i + \beta_2 sathealth_i + u_i$

Sie erhalten den folgenden Output:

	Modell 1			
	Koeffizient	Std. fehler	T	Sig.
lnw	0.366	0.031	11.54	0,000
sathealth	0.382	0.010	36.91	0,000
Konstante	2.930	0.147	19.92	0,000
R^2	0.2193			
$\overline{R}^2$	0.2190			
N	5036			

a) Interpretieren Sie den geschätzten Koeffizienten für β_1 statistisch und inhaltlich. (2 Punkte)

- *signifikant auf 1% Niveau, da p-Wert <0,01*
- *eine Erhöhung des Stundenlohnes um 1% ist c.p. im Mittel korreliert mit einem Zufriedenheitsgewinn von 0,00366 Punkten.*

b) Testen Sie auf dem 1% Signifikanzniveau, ob ein Anstieg der subjektiven Gesundheitszufriedenheit um eine Einheit die subjektive Lebenszufriedenheit um mehr als 0,2 Einheiten erhöht. Geben Sie Testverfahren, Null- und Alternativhypothese, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung an. (5 Punkte)

einseitiger t-Test, rechtsseitig
H_0: $\beta_2 \leq 0,2$

H_1: $\beta_2 > 0,2$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{0,382 - 0,2}{0,010} = 18,2$$
$$FG: n - k - 1 = 5036 - 2 - 1 = 5033$$
$$t_c = t_{\alpha,n-k-1} = t_{0,01;5033} = 2,326$$

Da $t_c < t$ kann die H_0 verworfen werden. Ein Anstieg der subjektiven Gesundheitszufriedenheit um eine Einheit erhöht die subjektive Lebenszufriedenheit um mehr als 0,2 Einheiten.

c) Einer ihrer Kommilitonen behauptet, dass die Anzahl der Schuljahre einen Einfluss auf die subjektive Lebenszufriedenheit hat. Sie berücksichtigen diese im folgenden Modell in quadratischer Form:

Modell 2: $\text{satlife}_i = \beta_0 + \beta_1 \text{lnw}_i + \beta_2 \text{sathealth}_i + \beta_3 \text{educ}_i + \beta_4 \text{educ}_i^2 + u_i$

Sie schätzen das Modell mit SPSS und erhalten den folgenden Output:

	Modell 2			
	Koeffizient	Std. fehler	T	Sig.
lnw	0.342	0.035	9.88	0.000
sathealth	0.382	0.010	36.81	0.000
educ	-0.046	0.077	-0.60	0.550
educ^2	0.002	0.003	0.80	0.426
Konstante	3.245	0.524	6.19	0.000
R^2	0.2299			

$\overline{R}^2$	0.2293
N	5036

Testen Sie *educ* und *educ*^{2} auf gemeinsame Signifikanz auf dem 1% Niveau. Geben Sie die Testverfahren, Null- und Alternativhypothese, die Anzahl der Freiheitsgrade, die Teststatistik den kritischen Wert und die Testentscheidung an. (5 Punkte)

F-Test:

H_0: $\beta_3 = \beta_4 = 0$

H_1: *H_0 trifft nicht zu*

Nenner-FG:
$n - k - 1 = 5036 - 4 - 1 = 5031$

Zähler-FG: 2

$$F = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-k-1)} = \frac{(0{,}2299-0{,}2193)/2}{(1-0{,}2299)/(5036-4-1)} = 34{,}62$$

$F_c = F_{0{,}01;2;5031} = 4{,}61$

$Wenn\ F > F_c$ *wird die H_0 verworfen.*

H_0 kann am 1% Signifikanzniveau abgelehnt werden. Die beiden Parameter sind gemeinsam signifikant von Null verschieden.

d) Hat Ihr Kommilitone Recht oder sind die Anzahl der Schuljahre in diesem Zusammenhang irrelevant? (2 Punkte)

Obwohl linearer Term und quadratischer Term einzeln insignifikant sind, sind beide gemeinsam signifikant. Somit handelt es sich bei educ nicht um eine irrelevante Variable.

e) Sie vermuten, dass sich alle Parameter im **Modell 1** zwischen Ost- und Westdeutschland unterscheiden. Wie können Sie diese Vermutung testen? Erläutern Sie ihre Vorgehensweise. Geben Sie Testverfahren, Teststatistik, und Freiheitsgrade an. (Hinweis: Sie müssen den Test nicht durchführen) (5 Punkte).

Lösungsvariante 1:

-getrennte Schätzung für Ost- und Westdeutschland, zusätzlich gepoolte Schätzung

-Ermittlung der SSR im gepoolten Modell und der SSR für beide Gruppen (Ost-/Westdeutschland)

-Chow-Test

$$\mathrm{F} = \frac{\mathrm{SSR_p} - (\mathrm{SSR_1} + \mathrm{SSR_2})/\mathrm{q}}{(\mathrm{SSR_1} + \mathrm{SSR_2})/(\mathrm{n} - 2(\mathrm{k}+1))}$$

-Zählerfreiheitsgrade: 3

-Nennerfreiheitsgrade: 5036-2(2+1)=5030

-falls F>c (kritischer Wert), dann mind. ein β_j zwischen Ost und West unterschiedlich

Lösungsvariante 2:

-gemeinsame Schätzung des Modells 1 für Ost- und Westdeutschland (restringiertes Modell)

-Schätzung eines vollständig interagierten Modells, d.h. Dummy-Variable für Region, zusätzlich alle Variablen (lnw, sathealth) interagiert mit Regionaldummy (unrestringiertes Modell)

-Ermittlung der SSR im restringierten und unrestringierten Modell

-F-Test auf gemeinsame Signifikanz

$$F = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(n-k-1)}$$

-Zählerfreiheitsgrade: 3

Nennerfreiheitsgrade: 5036-5-1=5030

-falls F>c (kritischer Wert), dann mind. ein β_j zwischen Ost und West unterschiedlich

f) Würden Sie Modell 1 oder Modell 2 den Vorzug geben? Begründen Sie. (1 Punkt)

Angepasstes $\overline{R}^2$ im Modell mit educ höher, somit Modell 2 besser.

Aufgabe 4: **[25 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 in dieser Aufgabe kann nicht negativ werden.

W	Nach einer KQ-Schätzung des Modells $y = \beta_0 + \beta_1 x_1 + u$ liegt der Punkt $(x_i, \hat{y}_i)$ auf der Regressionsgerade.
W	Die vorhergesagten Werte einer KQ-Schätzung und die Residuen sind unkorreliert.
F	Das Modell $y = \beta_0 + \frac{1}{\beta_1} x + u$ ist in Parametern linear.
W	Die Abbildung konvexer Zusammenhänge ist im linearen Regressionsmodell möglich.
F	Im Regressionsmodell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ lassen sich vier verschiedene Bedingungen erster Ordnung ableiten.
F	Je höher die gesamte Variation in der abhängigen Variable, desto höher das R^2.
F	Das R^2 fällt, wenn in die Schätzung eine irrelevante Variable zusätzlich aufgenommen wird.
F	Schätzt man fälschlicherweise mit Konstante, so können alle Steigungsparameter verzerrt werden.
W	In einer KQ-Schätzung können Funktionen der erklärenden Variablen als Regressoren verwendet werden.
F	Bei Multikollinearität ist eine KQ-Schätzung nicht durchführbar.
W	Die Gauss-Markov Annahmen werden verletzt, wenn die funktionale Form falsch spezifiziert wurde.
F	Unter Heteroskedastie sind die Parameterschätzer verzerrt.
W	Bei einem R^2 von 0,9 sind präzise Parameterschätzungen möglich.
F	Eine mit den Werten 1 und 2 kodierte binäre Variable kann in einer KQ-Schätzung nicht berücksichtigt werden.
F	Die Parameterschätzer von Dummyvariablen geben an, um wie viel sich die Steigungsparameter bei Teilgruppen unterscheiden.
W	Interaktionen zwischen zwei erklärenden Variablen sind im linearen Regressionsmodell zulässig.
F	Um einen gruppenspezifischen Unterschied in einem Steigungsparameter zu testen, muss das Modell vollständig interagiert sein.
F	Das angepasste Bestimmtheitsmaß schwankt mit der Zahl der abhängigen Variablen.
F	Der Korrelationskoeffizient gibt die Stärke des nicht-linearen Zusammenhangs zwischen zwei Variablen wider.
F	Die F-Verteilung ist linksschief.
F	In Querschnittsdaten sind Daten einer Beobachtungseinheit zu verschiedenen Zeitpunkten enthalten.
F	Die ceteris paribus Interpretation der Steigungsparameter ist abhängig vom Wert der Regressionskonstanten.
F	Eine Bernoulli Zufallsvariable nimmt Werte von -1 oder 1 an.
F	Eine Variable X ist eine stetige Zufallsvariable, wenn sie jeden einzelnen Wert mit der Wahrscheinlichkeit 1 annehmen kann.
W	Die Varianz einer Konstanten ist 0.
W	Wenn der Pearsonsche Korrelationskoeffizient den Wert +1 bzw. -1 annimmt, impliziert dies einen perfekten positiven bzw. negativen linearen Zusammenhang.
W	Die Standardnormalverteilung hat den Mittelwert 0 und eine Varianz von 1.

F	Die F-Verteilung nähert sich mit steigender Stichprobengröße der t-Verteilung an.
W	Die Stichproben-Regressionsfunktion ergibt eine von mehreren möglichen geschätzten Versionen der Bevölkerungs-Regressionsfunktion.
W	Der geschätzte Koeffizient $\hat{\beta}$ lässt sich aus dem Produkt aus t-Wert und Standardfehler von $\hat{\beta}$ berechnen.
W	Eine Modellspezifikation mit Polynomen 3.Ordnung ist flexibler als eine quadratische Modellspezifikation.
F	Je größer der Vorhersagefehler für y, desto kleiner die Varianz des geschätzten Koeffizienten $\hat{\beta}_k$.
F	Ist die Varianz des Fehlerterms u für alle Beobachtungen identisch, so bezeichnet man ihn als heteroskedastisch.
W	Das angepasste Bestimmtheitsmaß ($\bar{R}^2$) kann nicht unmittelbar zur Berechnung des F-Tests genutzt werden.
F	Bei kleinem R^2 lassen sich keine präzisen und unverzerrten Koeffizienten schätzen.
W	Das R^2 wird durch eine Umskalierung der erklärenden Variablen nicht beeinflusst.
F	Ein zweiseitiger t-Test auf dem 10% Niveau hat eine größere Irrtumswahrscheinlichkeit als ein einseitiger t-Test auf dem 10% Niveau.
F	Das 90%-Konfidenzintervall ist breiter als das 95%-Konfidenzintervall.
F	Gilt Annahme MLR.6 nicht, ist der F-Test, aber nicht der t-Test, für $n \rightarrow \infty$ approximativ gültig.
F	Ist die Stichprobengröße zu klein, so sind die Parameter verzerrt.
F	Bei immer kleiner werdenden Stichproben tendiert die Varianz konsistenter Schätzer gegen Null.
W	$\frac{1}{n+1}\sum x_i$ ist ein verzerrter aber konsistenter Schätzer für den Mittelwert.
W	Die BLUE-Eigenschaft eines Schätzers impliziert seine Unverzerrtheit.
F	Im Modell $\mathrm{wage_i} = \beta_0 + \beta_1\mathrm{educ_i} + \beta_2\mathrm{age_i} + \beta_3(\mathrm{educ_i} + \mathrm{age_i}) + \mathrm{u_i}$ misst β_3 den Effekt der Interaktion von educ und age.
F	Das Modell $\mathrm{wage_i} = \beta_1\mathrm{female_i} + \beta_2\mathrm{male_i} + \mathrm{u_i}$ kann nicht geschätzt werden.
F	Die Teststatistik beim t-Test kann nicht negativ werden.
W	Die Teststatistik beim F-Test kann nicht negativ werden.
W	Im Modell $y_i = \beta_0 + u_i$ beträgt das R^2=0.
W	Wird die abhängige Variable umskaliert, ändern sich sowohl die Koeffizienten als auch deren Standardfehler.
W	Im Modell $y_i = \beta_0 + u_i$ gibt der Koeffizient β_0 den Mittelwert der abhängigen Variable an.

Aufgabe 5: **[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 in dieser Aufgabe keine Punkte abgezogen.

1.	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

2.	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 c

3.	Unter den Gauss-Markov Annahmen (MLR.1–MLR.5) ist der KQ-Schätzer	
a	❑	unverzerrt.
b	❑	konsistent.

c	❑	effizient.
d	❑	asymptotisch effizient.
e	❑	a und b
f	☑	alle genannten Antworten

4.	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

5.	Im Modell Eisnachfrage $= \beta_0 + \beta_1$Frühling $+ \beta_2$Sommer $+ \beta_3$Herbst $+ \beta_4$Winter $+ u$	
a	❑	sind die Regressoren perfekt multikollinear.
b	❑	liegt das Problem der dummy variable trap vor.
c	❑	tritt der Typ I Fehler ein.
d	❑	führt eine KQ-Schätzung zu verzerrten Parametern.
e	☑	a und b
f	❑	b und d

6.	Die zwei Modelle $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ und $y = \beta_0 + \beta_1 x_1 + u$,	
a	❑	liefern identische Koeffizienten für x_1, wenn x_1 und x_2 unkorreliert sind.
b	❑	liefern identische Koeffizienten für x_1, wenn x_2 keinen Effekt auf y hat.
c	❑	liefern beide unter den Gauss-Markov Annahmen unverzerrte Schätzer, wenn $\beta_2 = 0$.
d	❑	sind genestet.
e	❑	c und d
f	☑	alle der genannten Antworten

7.	Interaktionsterme mit binären Variablen werden verwendet,	
a	❑	Um 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

8.	Die SST	
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

9.	Im einfachen linearen Modell $y_i = \beta_0 + \beta_1 x_{1i} + u_i$	
a	❑	wird die Zielfunktion $\sum_{i=1}^{n}(y_i - \widehat{\beta_0} + \widehat{\beta_1} x_{1i})^2$ maximiert.
b	❑	wird die Zielfunktion $\sum_{i=1}^{n} u_i^2$ maximiert.
c	❑	wird die Zielfunktion $\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$ maximiert.
d	☑	ergeben sich zwei Bedingungen erster Ordnung.
e	❑	alle der genannten Antworten
f	❑	keine der genannten Antworten

10.	Unter den Annahmen MLR.1–MLR.6	
a	❑	ist der KQ-Schätzer effizient.
b	❑	hat der KQ-Schätzer unter allen unverzerrten linearen Schätzern die kleinste Varianz.
c	❑	Sind die erklärenden Variablen und der Störterm unabhängig.
d	❑	folgt die t-Teststatistik unter der Nullhypothese der t-Verteilung.

e	❑	a und c
f	☑	alle Antworten

11.	Der zentrale Grenzwertsatz besagt, dass für eine Zufallsvariable Y mit Mittelwert μ und Varianz σ^2, gilt:	
a	❑	$\frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma}$ folgt asymptotisch der log-Normalverteilung
b	☑	$\frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma}$ folgt asymptotisch der Normalverteilung.
c	❑	$\frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma}$ folgt asymptotisch der t-Verteilung
d	❑	$\frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma}$ folgt der F-Verteilung.
e	❑	b und c
f	❑	a und c

12.	Beim Signifikanzniveau α gilt für den 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

13.	Bei konsistenten Schätzverfahren	
a	❑	ist der Schätzer immer unverzerrt.
b	❑	sinkt die Varianz des Schätzers mit steigender Beobachtungszahl n.
c	❑	kann die Konstante nicht interpretiert werden.
d	❑	liegt der Erwartungswert des Schätzers umso näher am wahren Wert, desto größer die Stichprobengröße n ist.
e	❑	a und c
f	☑	b und d

14.	F-Tests	
a	❑	können nur verwendet werden, um Hypothesen betreffend mehrerer Parameter zu testen.
b	❑	können nicht in kleinen Stichproben angewendet werden.
c	❑	sind asymptotisch effizient.
d	☑	werden unter Berücksichtigung von Zähler- und Nennerfreiheitsgraden durchgeführt.
e	❑	a und d
f	❑	b und c