Aufgabe 1 **[14 Punkte]**

Sie möchten untersuchen, wovon die Abwesenheit der Studierenden in den Vorlesungen an einer Universität abhängt. Sie verfügen über einen Datensatz zu 282 Studierenden mit folgenden Variablen:

skipped	durchschnittliche Anzahl von verpassten Vorlesungen pro Woche
male	Dummy-Variable: =1 wenn männlich, sonst =0
job	Dummy-Variable: =1 wenn Student jobbt, sonst =0
alcohol	durchschnittliche Anzahl von Tagen pro Woche an denen Person Alkohol trinkt
trans	Genutztes Verkehrsmittel auf dem Weg zur Uni (nur eins der drei möglich): 1-zu Fuß, bzw. öffentliche Verkehrsmittel, 2-mit dem Fahrrad, 3-mit dem Auto

a) Zunächst betrachten Sie das folgende Modell:

$$skipped = \beta_0 + \beta_1 male + \beta_2 job + \beta_3 alcohol + u$$

Bei einer Schätzung mit SPSS erhalten Sie folgenden Output:

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten	
	Regressions-koeffizientB	Standardfehler
(Konstante)	0,411	0,155
male	0,212	0,13
job	0,119	0,131
alcohol	0,255	0,05

a. Abhängige Variable: skipped

i) Interpretieren Sie inhaltlich die Größe der geschätzten Parameter $\hat{\beta}_1$ und $\hat{\beta}_3$. (2 Punkte)

β_1 - männliche Studierende verpassen c.p. durchschnittlich 0,212 Vorlesungen pro Woche mehr als Frauen.

β_3 – durchschnittlich erhöht jeder Tag pro Woche mehr, an dem Alkohol getrunken wird, die durchschnittliche Anzahl der ausgelassenen Vorlesungen c.p. um 0,255.

ii) Testen Sie am 1% Signifikanzniveau, ob der Koeffizient von *alcohol* signifikant größer als Null ist. Geben Sie an: welchen Test Sie durchführen, die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert, sowie die Testentscheidung. (4 Punkte)

t-Test

$$H_0: \beta_3 \le 0, H_1: \beta_3 > 0$$

$$t_{\hat{\beta}_3} = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0{,}255}{0{,}050} = 5{,}1$$

$$t_c = t_{\alpha, n-k-1} = t_{0{,}01,278} = 2{,}326$$

Der Koeffizient ist am 1% Signifikanzniveau statistisch signifikant größer als Null.

b) Sie wollen prüfen, b1) ob die genutzten Verkehrsmittel unterschiedliche Einflüsse auf die Vorlesungsabwesenheit haben und b2) ob sich der Effekt der Häufigkeit des Alkoholkonsums auf die

Vorlesungsabwesenheit danach unterscheidet, auf welches Verkehrsmittel man auf dem Weg zur Uni angewiesen ist.

i) Wie sieht die Regressionsgleichung aus, mit der Sie beide Vermutungen in nur einem Modell überprüfen können? Definieren Sie gegebenenfalls neu definierte Variablen. (3 Punkte)

Dummy-Variablen für 2 Ausprägungen der Variable trans und Interaktionen zwischen den 2 Dummies und der Variable alcohol ins Modell einfügen:

$$skipped = \beta_0 + \beta_1 male + \beta_2 job + \beta_3 alcohol + \beta_4 trans2 + \beta_5 trans3 + \\ + \beta_6 trans2 * alcohol + \beta_7 trans3 * alcohol + u,$$

wobei
trans2, Dummy-Variable:= 1, wenn Fahrrad, sonst =0
trans3, Dummy-Variable:= 1, wenn Auto, sonst =0

ii) Erläutern Sie kurz, wie sie b2) testen können. Geben Sie das Testverfahren und die Nullhypothese an. (2 Punkte)

Hier kann ein F-Test auf gemeinsame Signifikanz der Interaktionsterme aus dem Modell i) durchgeführt werden, d.h. es wird getestet, ob die Interaktionsterme gemeinsam einen signifikanten Erklärungsbeitrag leisten: $\beta_6 = \beta_7 = 0$.

c) Anschließend schätzen Sie das folgende Modell:

$$skipped = \beta_0 + \beta_1 male + \beta_2 job + \beta_3 alcohol + \beta_4 male * alcohol + u$$

i) Bestimmen Sie formal den marginalen Effekt der Häufigkeit des Alkoholkonsums für männliche Studierende. (1 Punkt)

$$\left.\frac{\partial skipped}{\partial alcohol}\right|_{male=1} = \beta_3 + \beta_4$$

ii) Bestimmen Sie formal den marginalen Effekt der Häufigkeit des Alkoholkonsums für weibliche Studierende. (1 Punkt)

$$\left.\frac{\partial skipped}{\partial alcohol}\right|_{male=0} = \beta_3$$

iii) Welches Vorzeichen erwarten Sie für den relevanten Parameter, wenn sich die Häufigkeit des Alkoholkonsums auf die Vorlesungsabwesenheit stärker bei weiblichen als bei männlichen Studierenden niederschlägt. (1 Punkt)

$\beta_3 + \beta_4 < \beta_3$

$\beta_4 < 0$

Aufgabe 2 **[4 Punkte]**

Gegeben sei das wahre Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$.

a) Erläutern Sie *kurz* die Auswirkungen bezüglich der Unverzerrtheit des KQ-Schätzers $\hat{\beta}_1$, wenn in einer Schätzung die relevante erklärende Variable x_2 nicht berücksichtigt wird. (2 Punkte)

KQ- Schätzer $\hat{\beta}_1$ *nicht mehr unverzerrt, d.h.* $E\{\hat{\beta}_1\} \neq \beta_1$, *wenn die ausgelassene relevante Größe* x_2 *mit* x_1 *korreliert.*

b) Welche Auswirkungen ergeben sich für $\hat{\beta}_1$, wenn Sie die irrelevante Variable x_3 zusätzlich in einer Schätzung aufnehmen? (2 Punkte)

KQ-Schätzer weiterhin unverzerrt, d.h. $E\{\hat{\beta}_1\} = \beta_1$; *die Varianz des Schätzers steigt aber, wenn* x_3 *mit* x_1 *korreliert ist, d.h. seine Effizienz sinkt.*

Aufgabe 3 **[8,5 Punkte]**

Sie schätzen mit Daten des Sozio-ökonomischen Panels (SOEP) für das Jahr 2008 den Zusammenhang zwischen (logarithmierten) Stundenlöhnen und einer Reihe von erklärenden Größen:

ln_hwage	logarithmierter Stundenlohn
exper	(Vollzeit) Arbeitsmarkterfahrung (in Jahren)
exper2	quadrierte (vollzeit) Arbeitsmarkterfahrung (in Jahren)
male	Dummy-Variable: =1 wenn Person männlich, sonst =0
married	Dummy-Variable: =1 wenn Person verheiratet, sonst =0
high_skill	Dummy-Variablen: =1 wenn Person hochqualifiziert ist, sonst =0
medium_skill	*- Referenz -*
low_skill	Dummy-Variablen: =1 wenn Person geringqualifiziert ist, sonst =0
sued	Dummy-Variable: =1 wenn Person in Süddeutschland wohnt, sonst =0

Sie schätzen das Modell zunächst ohne den Regionen-Dummy (sued) und erhalten folgendes Ergebnis:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	1,8494	0,0163		113,532	,000
	exper	,0403	0,0021	,750	19,594	,000
	exper2	-,0008	0,0001	-,573	-15,290	,000
	male	,1326	0,0136	,107	9,720	,000
	married	,0591	0,0138	,046	4,273	,000
	high_skill	,3952	0,0137	,302	28,897	,000
	low_skill	-,2525	0,0211	-,125	-11,958	,000

a. Abhängige Variable: ln_hwage

a) Berechnen und interpretieren Sie inhaltlich den marginalen Effekt der Arbeitsmarkterfahrung. (2 Punkte)

Marginaler Effekt von exper: $\frac{\partial \ln\_hwage}{\partial exper} = (0.0403 - 2 * 0.0008 \cdot exper) * 100\%$. *Mit einem zus. Jahr Berufserfahrung steigt der erwartete Lohn c.p. im Mittel um (4.03-2*0.08*exper) Prozent. Konkaver Verlauf, da* $\beta_{exper} > 0$ *und* $\beta_{exper2} < 0$.

b) Bestimmen Sie nun den vorhergesagten logarithmierten Stundenlohn einer hochqualifizierten, verheirateten Frau mit 20 Jahren Arbeitsmarkterfahrung. (1,5 Punkte)

*Die Frau hat einen vorhergesagten logarithmierten Stundenlohn von 1,8494+20*0,0403-20*20*0,0008+ 0,0591+0,3952 =2,7897.*

Sie schätzen erneut, diesmal unter Verwendung des Regionen-Dummies:

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressions-koeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	1,8320	0,0185		98,900	,000
	exper	,0404	?	,751	19,647	,000
	exper2	-,0008	0,0001	-,575	-15,352	,000
	male	,1322	0,0136	,107	9,705	,000
	married	?	0,0138	,046	4,244	,000
	high_skill	,3950	0,0137	,302	28,918	,000
	low_skill	-,2531	0,0211	-,125	-11,999	,000
	sued	,0438	0,0139	,035	?	,002

a. Abhängige Variable: ln_hwage

c) Berechnen Sie die 3 fehlenden Werte der Tabelle. (1,5 Punkte)

$se(\beta_1) = \frac{0{,}0404}{19{,}647} = 0{,}0021$, β_6 *= 4,244*0,0138 = 0,0586* , $t(\beta_{10}) = \frac{0{,}0438}{0{,}0139} = 3{,}151$

d) Sie stellen fest, dass sich die geschätzten exper-Koeffizienten und Standardfehler verglichen mit denen aus Teilaufgabe c) nicht ändern. Erläutern Sie *kurz*, warum dies der Fall sein könnte. (2 Punkte)

Dies ist wahrscheinlich darauf zurückzuführen, dass der Regionen-Dummy nicht mit den anderen erklärenden Variablen korreliert (Parameter einer Hilfsregression wäre $\delta_j = 0$*).*

e) Ist „sued“ eine irrelevante erklärende Variable? Argumentieren Sie statistisch und inhaltlich. (1,5 Punkte)

Nichtsdestotrotz ist „sued“ eine relevante erklärende Variable, da der geschätzte Koeffizient statistisch hochsignifikant von Null verschieden ist (am 1%-Niveau) und mit einer Lohnwirkung von 4,4 Prozent auch ökonomisch relevant ist.

Aufgabe 4: **[23,5 Punkte]**

Sie interessieren sich für die Determinanten der subjektiven Lebenszufriedenheit.
Ihr Datensatz beinhaltet folgende Variablen:

satlife	Subjektive Lebenszufriedenheit mit den Ausprägungen 0 (sehr unzufrieden) bis 10 (sehr zufrieden)
age	Alter einer Person in Jahren
sqage	Alter^2 /100
educ	Anzahl der Bildungsjahre
sqeduc	Bildungsjahre^2 /100
west	Dummy-Variable: =1, wenn Person in den alten Bundesländern lebt, =0 sonst
male	Dummy-Variable: =1, wenn Person männlich, =0 sonst
unempl	Dummy-Variable: =1, wenn Person arbeitslos, =0 sonst
child	Anzahl der Kinder im Haushalt
log(incnet)	logarithmiertes Nettomonatseinkommen einer Person in €

Sie schätzen zunächst folgendes **Modell 1**:

$$satlife_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sqage}_i + \beta_3 \text{educ}_i + \beta_4 \text{sqeduc}_i + \beta_5 \text{west}_i + \beta_6 \text{male}_i + \beta_7 \text{unempl}_i + \beta_8 \text{child}_i + u_i$$

	Modell 1				**Modell 2**			
	Koeffizient	Std. fehler	T	Sig.	Koeffizient	Std. fehler	T	Sig.
age	-,0327	,0018	-18,03	0,000	-,1153	,0040	-28,69	0,000
sqage	,0282	,0018	15,71	0,000	,1282	,0047	27,18	0,000
educ	,2582	,0183	14,07	0,000	,1086	,0251	4,33	0,000
sqeduc	-,6123	,0679	-9,01	0,000	-,1838	,0909	-2,02	0,043
west	,6100	,0128	47,39	0,000	,5087	,0166	30,61	0,000
male	-,0297	,0108	-2,75	0,006	-,1625	,0154	-10,53	0,000
unempl	-1,2072	,0213	-56,46	0,000	-,7218	,0622	-11,59	0,000
child	,0496	,0065	7,56	0,000	,0905	,0078	11,55	0,000
log(incnet)	---	---	---	---	,2477	,0112	22,04	0,000
Konstante	5,1702	,1279	40,41	0,000	6,3024	,1819	34,63	0,000
SSE	27.646				8.788			
SSR	306.288				137.742			
SST	333.934				146.530			
N	102.575				54.615			

Abhängige Variable: satlife

a) Interpretieren Sie den geschätzten Parameter $\hat{\beta}_7$ statistisch unter Verwendung des p-Wertes und inhaltlich. Beurteilen Sie die Stärke des Effektes. (3 Punkte)

-statistisch signifikant auf 1% -Niveau, da p-Wert<0,01.
-ökonomisch: Arbeitslosigkeit ist korreliert mit einem Zufriedenheitsverlust von 1,2 Punkten
-sehr großer Einfluss, größter Parameter im Modell
-vor dem Hintergrund einer Gesamtstreuung der abh. Variable von 0-10 ist 1,2 ein relativ großer Effekt

b) Unterstellen Sie, dass die Variable sqage nicht durch 100 geteilt worden wäre. Welche Werte erwarten Sie nun für die Schätzwerte des Regressionskoeffizienten, des Standardfehlers, des t-Wertes und des Signifikanzniveaus des Parameters β_2? (3 Punkte)

-Koeff: 0,000282, SE: 0,000018, t-Wert und Signifikanz identisch.

c) Bei welchem Alter ist die durchschnittliche Zufriedenheit minimal? (2,5 Punkte)

$$\frac{\partial satlife}{\partial age} = \beta_1 + 2\beta_2 age$$
$$\beta_1 + 2\beta_2 age = 0$$
$$age = \frac{-\beta_1}{2\beta_2} = \frac{0{,}0327}{2\cdot 0{,}000282} = 55{,}05\ Jahre$$

d) Testen Sie die gemeinsame Signifikanz von educ und sqeduc auf dem 1%-Niveau. ($\mathrm{SSR_r} = 306.396$). Geben Sie Testverfahren, Teststatistik, Nullhypothese, kritischen Wert und Ihre Testentscheidung an. (4,5 Punkte)

F-Test
$H_0{:}\, \beta_3 = \beta_4 = 0$ *;* $H_1{:}\, \beta_3 \neq 0$ *oder* $\beta_4 \neq 0$ *oder beides.*
$F = \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)} = \frac{(306.396-306.288)/2}{306.288/(102.575-8-1)}$ =18,08
$F_c = F_{\alpha;q;n-k-1} = F_{1\%;2;102566} = F_{1\%;2;\infty} = 4{,}61$
Da $F > F_c$ wird H_0 am 1%-Niveau verworfen.

Sie schätzen nun folgendes **Modell 2**:

$$satlife_i = \beta_0 + \beta_1 \mathrm{age}_i + \beta_2 \mathrm{sqage}_i + \beta_3 \mathrm{educ}_i + \beta_4 \mathrm{sqeduc}_i + \beta_5 \mathrm{west}_i + \beta_6 \mathrm{male}_i + \beta_7 \mathrm{unempl}_i + \beta_8 \mathrm{child}_i + \boldsymbol{\beta_9 \log(incnet)} + u_i$$

e) Interpretieren Sie auf Basis der Ergebnisse in obiger Tabelle $\hat{\beta}_9$ inhaltlich. (1 Punkt)

-eine Erhöhung des Einkommens um 1% ist c.p. korreliert mit einer Erhöhung der Zufriedenheit um 0,002 Punkte.

f) Berechnen Sie, wie sich das Einkommen einer Person ohne Kind ändern müsste, damit das gleiche Zufriedenheitsniveau erreicht wird wie bei einer Person mit einem Kind. (4 Punkte)

- $\Delta\beta_8 = \Delta\beta_9 \log(incnet)$
- $\Delta\, 0{,}0905 = \Delta\, 0{,}2477 \log(incnet)$
- $\frac{0{,}0905}{0{,}2477} = \log(incnet) = 0{,}3654$
- $(e^{0{,}3654}) - 1 = 0{,}4411$
- *Das Einkommen müsste bei einer Person ohne Kind um 44,11% ansteigen, um c.p. das gleiche Zufriedenheitsniveau wie eine Person mit Kind zu erhalten.*

g) Berechnen Sie das R^2 für Modell 2. Interpretieren Sie den von Ihnen berechneten Wert für R^2. (3 Punkte)

Modell 2: R^2*=1-SSR/SST=1-137742/146530=0,06*

Das Bestimmtheitsmaß gibt an, welcher Anteil der gesamten Variation in y durch die erklärenden Größen erklärt werden kann. Hier beträgt dieser Anteil 6%.

h) Diskutieren Sie unter welchen Umständen die Gauss-Markov-Annahme MLR.4 im Modell 2 verletzt ist. (2,5 Punkte)

MLR.4 ist verletzt, wenn $E\{u|x\} \neq 0$, d.h. wenn der auf x bedingte Erwartungswert von u ungleich Null ist. Dies kann der Fall sein, wenn es beispielsweise zusätzliche Variablen gibt, die sowohl mit den Regressoren des Modells als auch das Zufriedenheitsniveau korrelieren. Hier z.B. persönliche Charakteristika denkbar, die sowohl das Einkommen als auch die Zufriedenheit beeinflussen, z.B. Gesundheitszustand.

Aufgabe 5: **[24 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 F-Verteilung ist eine symmetrische Verteilungsfunktion.
w	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$.
f	Eine Annahme des einfachen linearen Modells ist $E(u\|x) = 1$.
f	Große p-Werte sind Evidenz gegen H0.
w	Der geschätzte Koeffizient $\hat{\beta}$ lässt sich aus dem Produkt aus t-Wert und Standardfehler von $\hat{\beta}$ berechnen.
f	Das angepasste R² kann nicht negativ werden.
w	Im einfachen linearen Regressionsmodell liegt der Punkt $(\bar{x}, \bar{y})$ auf der Regressionsgeraden.
w	Wenn es gelingt, alle relevanten unabhängigen Variablen konstant zu halten, lässt sich der ceteris paribus Effekt als kausale Wirkung interpretieren.
f	Der Parameter β_1 im Modell $\log(y) = \beta_0 + \beta_1 \log(x_1) + u$ wird als Semielastizität interpretiert.
w	$Var(X - 2Y) = Var(X) + 4Var(Y) - 4Cov(X, Y)$.
f	Multikollinearität führt zu verzerrten Schätzern.
w	Das Gauss-Markov-Theorem besagt, dass unter bestimmten Bedingungen das KQ-Verfahren das effizienteste der linearen unverzerrten Schätzverfahren ist.
f	Für die Unverzerrtheit des KQ-Schätzers spielt $E\{u\|x\} = 0$ keine Rolle.
w	Man kann Dummy-Variablen aus stetigen Variablen bilden.
f	Im multiplen linearen Modell ist das R^2 das geeignete Maß, um über die Aufnahme zusätzlicher erklärender Größen zu entscheiden.
w	Eine standardisierte stetige Variable hat den Erwartungswert von 0 und eine Standardabweichung von 1.
f	Bei unverzerrten Schätzern ist jeder Schätzwert mit dem wahren Wert identisch.
f	Ein Schätzer ist unverzerrt, wenn mit größer werdender Stichprobe die Varianz sinkt.
w	Der F-Wert ist nie negativ.
w	Für große Stichproben gelten t-Tests asymptotisch auch ohne normalverteilte Störterme.
f	Die Varianz einer Konstanten ist gleich dem Wert dieser Konstante.
w	Durch Verwendung von Polynomen der erklärenden Variablen können in einer Regression nicht-lineare Effekte dieser Variablen abgebildet werden.
f	Das lineare Regressionsmodell heißt linear, weil alle erklärenden Größen nur in erster Potenz vorkommen dürfen.
w	Der Lagrange-Multiplier Test verwendet im 2. Schritt die Residuen des restringierten Modells als abhängige Variable.
w	Von zwei unverzerrten Schätzern $\hat{\beta}_1$ und $\tilde{\beta}_1$ für einen unbekannten Bevölkerungsparameter β_1 ist $\tilde{\beta}_1$ effizienter, wenn $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$.
f	Im einfachen linearen Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ bezeichnet β_1 den Achsenabschnitts-parameter.
f	Benutzt man 15 Bundesland-Dummies (einer pro Land), um für Unterschiede zwischen den 16 Bundesländern zu kontrollieren, so kann über das ausgelassene Bundesland keine Aussage getroffen werden.
f	Over-controlling führt zu hohen Standardfehlern und zu verzerrten Parametern.
w	Selbst wenn es im einfachen linearen Modell keine Variation in der erklärenden Variable gibt, kann der Effekt der Variable mit KQ geschätzt werden.
w	Genestete Modelle können mittels R² verglichen werden.
w	Eine Modellspezifikation mit Polynomen 3.Ordnung ist flexibler als eine quadratische Modellspezifikation.
f	Schätzt man im multiplen linearen Modell ohne Achsenabschnitt, dann geht man davon aus, dass $\beta_0 = 1$.
w	Die Störtermvarianz ändert sich bei Umskalierung der abhängigen Variable.

w	Bei einer normalverteilten Zufallsvariablen ist in der Grundgesamtheit der Medianwert gleich dem arithmetischen Mittel.
w	Das multiple Regressionsmodell erlaubt die Berechnung von Partialeffekten.
w	Durch die Aufnahme mehrerer Kontrollvariablen im multiplen Regressionsmodell können kausale Interpretationen plausibler werden.
w	Folgt log(Y) einer Normalverteilung, dann ist Y lognormal verteilt.
w	Hat der Fehlerterm u für alle Beobachtungen die gleiche Varianz, so bezeichnet man ihn als homoskedastisch.
f	Auch wenn gilt $E\{u\|x\} \neq 0$ kann im Rahmen einer linearen Schätzung durch eine binäre Variable der Effekt einer Politikmaßnahme evaluiert werden.
f	Bei der „dummy variable trap“ sind die Regressionskoeffizienten eindeutig bestimmbar aber nicht effizient
f	Mehrere Dummy-Variablen in einer Schätzung führen zu „dummy variable trap“.
f	Werden irrelevante Größen in ein Schätzmodell aufgenommen so führt dies zu „omitted variable bias“.
f	Wird im Modell $y = \beta_0 + \delta_0 Geschlecht$ die Variable *Geschlecht* mit den Werten 1-männlich, 2-weiblich kodiert, so ist δ_0 verzerrt.
w	Die Zählerfreiheitsgrade beim F-Test geben die Anzahl der getesteten Restriktionen an.
f	Im multiplen linearen Regressionsmodell mit k erklärenden Größen ergeben sich bei der Herleitung des KQ-Schätzers genau k Bedingungen 1. Ordnung.
f	Im Modell $y = \beta_0 + \beta_1 \log(x_1) + u$ gibt β_1 an, um wie viel Prozent sich die abhängige Variable ändert, wenn x_1 um eine Einheit steigt.
w	Die Varianz des Fehlerterms kann im einfachen linearen Modell mittels $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$ unverzerrt geschätzt werden.
w	Die SSR im restringierten Modell ist immer mindestens genauso groß ist wie die SSR im unrestringierten Modell.

Aufgabe 6: **[16 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.	Wenn der p-Wert für den Signifikanztest eines Parameters den Wert 0,06 annimmt, dann	
a	❑	wird H_0 am 5% Niveau verworfen.
b	❑	wird H_0 am 10% Niveau verworfen.
c	❑	beträgt der Typ I Fehler bei Verwerfen von H_0 6%.
d	❑	beträgt der Typ II Fehler bei Verwerfen von H_0 6%.
e	❑	a und c
f	☑	b und c

2.	Beim Testen gegen die Alternative H_1: $\beta_j < 0$, wird H_0 verworfen, wenn	
a	❑	$\|t\| < -c$ (c=kritischer Wert)
b	❑	$\|t\| < c$
c	☑	$t < -c$
d	❑	$t > -c$
e	❑	b und c
f	❑	b und d

3.	Ein 99% Konfidenzintervall für einen Parameter β_j	
a	❑	kann nicht mit dem Konfidenzintervall für einen Parameter β_k überlappen.
b	❑	schließt bei signifikantem Parameterschätzer den Wert Null ein.
c	☑	ist abhängig von der Stichprobengröße.
d	❑	enthält mit Wahrscheinlichkeit 99% den wahren Wert.
e	❑	a, b und c
f	❑	b und d

4.	In einer quadratischen Spezifikation $y = \beta_1 x + \beta_2 x^2 + u$ ergibt sich ein umgekehrt u-förmiger Verlauf der Regressionsgeraden, wenn	
a	❑	$\beta_1 < 0, \beta_2 < 0$
b	☑	$\beta_1 > 0, \beta_2 < 0$
c	❑	$\beta_1 > 0, \beta_2 > 0$
d	❑	$\beta_1 = 0, \beta_2 < 0$
e	❑	$\beta_1 < 0, \beta_2 > 0$
f	❑	keine der Antworten

5.	Welche Aussage beschreibt keine Eigenschaft des KQ-Schätzers?	
a	❑	Jeder vorhergesagte Wert $\hat{y}$ liegt auf der Regressionsgeraden.
b	❑	Die Stichprobenkovarianz zwischen x_i und u_i ist 1.
c	❑	y_i lässt sich über $\hat{y}_i$ und $\hat{u}_i$ berechnen.
d	❑	$\sum_{i=1}^{n} \hat{u}_i^2 = 0$.
e	❑	a und c
f	☑	b und d

6.	Wenn die Kovarianz von zwei Zufallsvariablen Null ist (d.h. $Cov(X, Y) = 0$), dann	
a	❑	sind die zwei Variablen unabhängig.
b	☑	sind die zwei Variablen unkorreliert.
c	❑	besteht kein Zusammenhang zwischen den Variablen.
d	❑	kann Skalieren der Variablen die Kovarianz ändern.
e	❑	a, b und c
f	❑	b, c und d

7.	Folgende Annahmen müssen zutreffen, damit mit dem KQ-Verfahren im multiplen Modell ohne Konstante mit k erklärenden Größen unverzerrt geschätzt werden kann:	
a	❑	Eine betrachtete Zufallsstichprobe folgt dem Bevölkerungsmodell.
b	❑	Keine unabhängige Variable ist konstant.
c	❑	$E(u\|x_1 \dots x_k) = 0$.
d	❑	Der Fehlerterm u ist homoskedastisch.
e	☑	a, b und c
f	❑	alle Antworten

8.	Im einfachen linearen Modell $y_i = \beta_0 + \beta_1 x_{1i} + u_i$ erhalten wir den KQ-Schätzer $\hat{\beta}_1$ durch die Minimierung folgender Zielfunktion S:	
a	❑	$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i})^2$
b	❑	$\sum_{i=1}^{n}(y_i - \bar{y})^2$
c	❑	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
d	❑	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
e	❑	a und b
f	☑	a und d

9.	Ein Chow-Test	
a	❑	kann nur durch Verwendung von Interaktionstermen durchgeführt werden.
b	❑	basiert auf einer Chi-Quadrat verteilten Teststatistik.
c	❑	prüft, ob Regressionskoeffizienten nach Gruppen unterschiedlich ausfallen.
d	❑	kann durch Schätzung getrennter Regressionen für zwei Gruppen durchgeführt werden.
e	❑	b und c
f	☑	c und d

10.	Bei einer logarithmierten abhängigen Variable wird der marginale Effekt einer erklärenden Dummy-Variable:	
a	❑	als $(exp\,(\beta) - 1) \cdot 100$ berechnet.
b	❑	approximativ als $100 \cdot \beta$ berechnet.
c	❑	als Prozentgröße interpretiert.
d	❑	relativ zur Referenzgruppe interpretiert.
e	❑	a, c und d
f	☑	alle Antworten

11.	Ein Schätzer eines unbekannten Parameters	
a	❑	ist unverzerrt, wenn sein Erwartungswert mit dem wahren Wert identisch ist.
b	❑	ist verzerrt wenn sein Wert von Stichprobe zu Stichprobe variiert.
c	❑	ist effizient nur wenn er gleichzeitig symmetrisch ist.
d	❑	ist eine Zufallsvariable.
e	☑	a und d
f	❑	a, b und d

12.	Im Modell $y = \beta_0 + \beta_1 x_1 + u$ kommt es aufgrund einer ausgelassenen Variablen x_2 zu einer positiver Verzerrung von $\hat{\beta}_1$, wenn: (Hilfsregression: $x_1 = \delta_0 + \delta_1 x_2 + \varepsilon$)	
a	❑	$\beta_2 = 0$ und $\tilde{\delta}_1 > 0$
b	☑	$\beta_2 < 0$ und $\tilde{\delta}_1 < 0$
c	❑	$\beta_2 > 0$ und $\tilde{\delta}_1 = 0$
d	❑	$\beta_2 = 0$ und $\tilde{\delta}_1 = 0$
e	❑	a und c
f	❑	keine der Antworten

13.	In einer geschätzten Lohnregression $\log(wage) = \hat{\beta}_0 + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 exper^2$	
a	❑	ist $\hat{\beta}_1$ als Semi-Elastizität interpretierbar.
b	❑	würde eine Steigerung in educ um eine Einheit eine approximative Änderung in wage in Höhe von $100 \cdot \hat{\beta}_1$ Prozentpunkten verursachen.
c	❑	wird der marginale Effekt von exper als partielle Ableitung von log(wage) nach exper berechnet.
d	❑	ist die Lohnwirkung einer zusätzlichen Einheit exper sowohl von educ als auch vom Ausgangsniveau in exper abhängig.
e	❑	b, c und d
f	☑	a und c

14.	Von Heteroskedastie spricht man, wenn	
a	❑	irrelevante erklärende Variablen in einem Regressionsmodell aufgenommen werden.
b	❑	eine erklärende Variable x_j mit dem Fehlerterm u korreliert ist.
c	❑	Parameter in nicht-linearer Weise in das Regressionsmodell aufgenommen werden.
d	☑	die Varianz des Fehlerterms von der Ausprägung der erklärenden Größen abhängt.
e	❑	a und d
f	❑	b und d

15.	Wird im Modell $y = \beta_0 + \beta_1 x_1 + u$ die unabhängige Variable mit 100 multipliziert, so	
a	☑	wird β_1 durch 100 geteilt.
b	❑	werden β_0 und β_1 durch 100 geteilt.
c	❑	muss y durch 100 geteilt werden.
d	❑	muss y mit 100 multipliziert werden.
e	❑	a und d
f	❑	b und c

16.	Der Lagrange Multiplier Test	
a	❑	verwendet eine χ^2-verteilte Teststatistik.
b	❑	wird u.a. zum Testen mehrerer linearer Restriktionen verwendet.
c	❑	ist mit dem Chow-Test identisch wenn q=2.
d	❑	schätzt zwei Regressionsmodelle.
e	☑	a, b und d
f	❑	a und b