Bachelorprüfung SS 2013 - MUSTERLÖSUNG

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen. Es wird nur der Lösungsbogen eingesammelt.

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: **[15 Punkte]**

Sie haben Einkommensdaten für 3000 in Vollzeit beschäftigte Personen und interessieren sich für den Zusammenhang zwischen dem logarithmierten Stundenlohn und den persönlichen Eigenschaften der betrachteten Personen. Folgende Informationen stehen Ihnen zur Verfügung:

ln_hwage	logarithmierter Stundenlohn
male	= 1 wenn Person männlich, = 0 wenn Person weiblich
married	= 1 wenn Person verheiratet, = 0 sonst
exper	Berufserfahrung in Jahren
public	= 1 wenn Person im öffentlichen Dienst arbeitet, = 0 sonst

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

$$ln\_hwage_i = \beta_0 + \beta_1 male_i + \beta_2 married_i + \beta_3 exper_i + \beta_4 public_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(*Konstante*)	1,575	**???**	8,750	0.000
male	**???**	0,056	2,107	0,036
married	0,137	0,069	1,986	0,047
exper	0,052	0,022	**???**	0,018
public	0,143	0,097	1,474	0,141

a. Abhängige Variable: *ln_hwage*

Hinweis: Das R^2 dieser Schätzung beträgt $0,341$.

a) Berechnen Sie $se(\hat{\beta}_0)$, $\hat{\beta}_1$ und $t(\hat{\beta}_3)$ (**3 Nachkommastellen**). Ist der geschätzte Koeffizient für β_4 am 5%-Niveau statistisch signifikant? Begründen Sie kurz. (4 Punkte)

- $se(\hat{\beta}_0) = \frac{\hat{\beta}_0}{t(\hat{\beta}_0)} = \frac{1{,}575}{8{,}750} = 0,180$ [1P]
- $\hat{\beta}_1 = t(\hat{\beta}_1) \cdot se(\hat{\beta}_1) = 2,107 \cdot 0,056 = 0,118$ [1P]
- $t(\hat{\beta}_3) = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0{,}052}{0{,}022} = 2,364$ [1P]
- Da $t(\hat{\beta}_4) = 1{,}474 < 1{,}96$ (kritischer Wert der t-Verteilung bei 2995 Freiheitsgraden) ist der geschätzte Koeffizient nicht statistisch signifikant am 5%-Niveau (Alternative Antwort möglich). [1P]

b) Berechnen und interpretieren Sie den genauen (!) Effekt der Variable *married* auf den Stundenlohn (**3 Nachkommastellen**). (2 Punkte)

- $\hat{\beta}_2 = 0,137$
- Der genaue Effekt beträgt $e^{\hat{\beta}_2} - 1 = 0,147$. [1P]
- Eine verheiratete Person verdient c.p. im Mittel $14,7$ Prozent mehr pro Stunde als eine nicht verheiratete Person. [1P]

c) Wie gehen Sie vor um zu entscheiden, ob sich der Lohneffekt des öffentlichen Sektors für verheiratete und nicht verheiratete Personen unterscheidet ohne getrennt für beide Gruppen zu schätzen? (2 Punkte)

- Im Modell müsste ein Interaktionsterm von *public* und *married* aufgenommen werden: [1P]
- $ln\_hwage_i = \beta_0 + \beta_1 male_i + \beta_2 married_i + \beta_3 exper_i + \beta_4 public_i + \beta_5 (married_i \cdot public_i) + u_i$
- Der Koeffizient des Interaktionsterms β_5 muss auf statistische Signifikanz getestet werden. [1P]

d) Sie interessieren sich dafür, ob der Einfluss der Variable *exper* für alle beobachteten Ausprägungen dieser Variable konstant ist und schätzen mit SPSS das Modell:

$$ln\_hwage_i = \beta_0 + \beta_1 male_i + \beta_2 married_i + \beta_3 exper_i + \beta_4 exper_i^2 + \beta_5 public_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	1,582	0,184	8,598	0,000
male	0,121	0,055	2,200	0,028
married	0,135	0,067	2,015	0,045
exper	0,046	0,020	2,300	0,022
$exper^2$	-0,006	**???**	**???**	**???**
public	0,145	0,096	1,510	0,131

a. Abhängige Variable: *ln_hwage*

Hinweis: Das R^2 dieser Schätzung beträgt $0,362$.

Beim Übernehmen der Ergebnisse aus SPSS sind Ihnen leider die Informationen zu Standardfehler, t-Wert und Signifikanz des Koeffizienten von $exper^2$ verloren gegangen. Wie können Sie trotzdem bestimmen, ob die neu aufgenomme Variable einen signifikanten Erklärungsbeitrag in Ihrem Modell liefert? Führen Sie einen entsprechenden Test am 5%-Niveau durch, indem Sie Nullhypothese, Alternativhypothese, Freiheitsgrade, Teststatistik, kritischen Wert und Testentscheidung angeben. Sollten Sie auf Grundlage Ihrer Ergebnisse die Variable $exper^2$ in Ihr Modell aufnehmen? (7 Punkte)

1. **Nullhypothese:** $H_0 : \beta_4 = 0$ [0,5P]
2. **Alternativhypothese:** $H_1 : \beta_4 \neq 0$ [0,5P]
3. **Freiheitsgrade (aus unrestringiertem Modell):** $n-k-1 = 3000-5-1 = 2994$ [1P] und $q = 1$ [1P]
4. **Teststatistik:** $F_{empirisch} = \frac{(R_u^2 - R_R^2)/q}{(1-R_u^2)/(n-k-1)} = \frac{(0{,}362-0{,}341)/1}{(1-0{,}362)/(3000-5-1)} = \frac{0{,}021}{0{,}00021309} = 98{,}55$ [1P]
5. **Kritischer Wert:** $F_{kritisch} = F_{0{,}05;1{,}2994} = 3{,}84$ [1P]
6. **Testentscheidung:** Da $F_{empirisch} > F_{kritisch}$ wird die Nullhypothese auf dem 5% Niveau verworfen. [1P] Auf Grundlage der Teststatistik hat die Variable einen signifikant von Null verschiedenen Einfluss und sollte in das Modell aufgenommen werden. [1P]

Aufgabe 2: **[15 Punkte]**

Sie möchten untersuchen, wovon die Computernutzung von Grundschülern der 3. Klasse abhängt. Der Ihnen zur Verfügung stehende Datensatz enthält folgende Informationen zu 100 Schülern:

$Comp$	Computernutzung gemessen in Minuten pro Tag
$Male$	$= 1$ wenn Schüler männlich, $= 0$ wenn Schüler weiblich
$Work$	Arbeitszeit der Mutter gemessen in Stunden pro Tag
$Work^2$	Quadrierte Arbeitszeit der Mutter ($Work^2 = Work \cdot Work$)
$Sport$	Dauer sportlicher Aktivitäten gemessen in Stunden pro Woche
$Autumn$	$= 1$ wenn Befragung im Herbst, sonst $= 0$
$Winter$	$= 1$ wenn Befragung im Winter, sonst $= 0$
$Spring$	$= 1$ wenn Befragung im Frühling, sonst $= 0$

Sie schätzen das folgende Modell:

$$Comp_i = \beta_0 + \beta_1 Male_i + \beta_2 Work_i + \beta_3 Work_i^2 + \beta_4 Sport_i + \beta_5 Autumn_i + \beta_6 Winter_i + \beta_7 Spring_i + u_i$$

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
$(Konstante)$	42,820	12,778	3,351	0,000
$Male$	12,648	5,547	2,280	0,024
$Work$	15,041	4,003	3,757	0,000
$Work^2$	0,473	0,189	2,503	0,000
$Sport$	-23,991	5,302	-4,525	0,000
$Autumn$	17,488	8,305	2,106	0,037
$Winter$	40,720	13,085	3,112	0,000
$Spring$	8,306	5,431	1,529	0,129

a. Abhängige Variable: *Comp*

a) Interpretieren Sie die geschätzte Konstante inhaltlich und statistisch. (2 Punkte)

- Eine Grundschülerin der 3. Klasse, welche keine sportlichen Aktivitäten betreibt, im Sommer befragt wurde und deren Mutter nicht arbeitet, nutzt den Computer c.p. im Mittel 42,82 Minuten pro Tag. [1P]
- Der geschätzte Koeffizient ist statistisch signifikant von Null verschieden auf dem 1% Niveau. [1P]

b) Testen Sie auf dem 10% Signifikanzniveau, ob Jungen im Vergleich zu Mädchen mit gleichen Charakteristika den Computer mehr als 15 Minuten pro Tag länger nutzen. Geben Sie Testverfahren, Null- und Alternativhypothese, Teststatistik, kritischen Wert und Ihre Testentscheidung an. Welche inhaltliche Aussage können Sie nun zum Unterschied in der Computernutzung treffen? (5,5 Punkte)

1. **Testverfahren:** einseitiger (rechtsseitiger) t-Test [0,5P]
2. **Hypothesen:** $H_0 : \beta_1 \leq 15$; $H_1 : \beta_1 > 15$ [1P]
3. **Teststatistik:** $t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - 15}{se(\hat{\beta}_1)} = \frac{12{,}648 - 15}{5{,}547} = \frac{-2{,}352}{5{,}547} = -0{,}424$ [1P]
4. **Kritischer Wert c:** $c = t_{\alpha, n-k-1} = t_{10\%, 100-7-1} = t_{10\%, 92} = 1{,}291$ [1P]
5. **Testentscheidung:** Da $t_{\hat{\beta}_1} = -0{,}424 < 1{,}291 = c$ kann die Nullhypothese auf dem 10% Signifikanzniveau nicht verworfen werden. [1P] Somit ist der Unterschied in der täglichen Computernutzung zwischen Mädchen und Jungen (mit gleichen Charakteristika) nicht signifikant größer als 15 Minuten. [1P]

c) Tim und Lisa aus Nürnberg wurden im Sommer befragt und treiben beide gleich viel Sport. Tims Mutter arbeitet 6 Stunden pro Tag, Lisas Mutter arbeitet 4 Stunden pro Tag. Wie viele Stunden pro Woche müsste Tim nach obiger Schätzung mehr Sport treiben, damit die tägliche Computernutzung der beiden Kinder exakt übereinstimmt? (4,5 Punkte)

- Unterschied in Computernutzung zwischen Lisa und Tim:
  $\Delta Comp = (15{,}041 \cdot 4 + 0{,}473 \cdot 4^2) - (+12{,}648 + 15{,}041 \cdot 6 + 0{,}473 \cdot 6^2) = -52{,}19$ [2P]
  → Bei gegebenen Charakteristika nutzt Lisa den Computer c.p. im Mittel 52,19 Minuten pro Tag weniger als Tim.
- Ausgleich durch Sportaktivität von Tim:
  $\Delta Comp = \hat{\beta}_4 \Delta Sport \Leftrightarrow \Delta Sport = \frac{-52{,}19}{-23{,}991} = 2{,}175$ [2P]
  → Tim müsste 2,175 Stunden (2 Stunden und 11 Minuten) pro Woche mehr Sport machen, damit die tägliche Nutzungsdauer der beiden Kinder übereinstimmt. [0,5P]

d) Wie würde sich jeweils $\hat{\beta}_2$ und $\hat{\beta}_3$ ändern, wenn die Arbeitszeit der Mutter nicht in Stunden pro Tag, sondern in Minuten pro Tag gemessen wäre? Beeinflusst diese Transformation den p-Wert von $\hat{\beta}_3$? (3 Punkte)

- Erläuterung: Umskalierung der Variable $Work$ und $Work^2$:

$$\widehat{Comp}_i = \hat{\beta}_0 + \hat{\beta}_1 Male_i + \hat{\beta}_2 Work_i + \hat{\beta}_3 Work_i^2 + \hat{\beta}_4 Sport_i + \hat{\beta}_5 Autumn_i + \hat{\beta}_6 Winter_i + \hat{\beta}_7 Spring_i$$

$$\Leftrightarrow \widehat{Comp}_i = \hat{\beta}_0 + \hat{\beta}_1 Male_i + \hat{\beta}_2 \frac{60 \cdot Work_i}{60} + \hat{\beta}_3 \left(\frac{60 \cdot Work_i}{60}\right)^2 + \hat{\beta}_4 Sport_i + \hat{\beta}_5 Autumn_i + \hat{\beta}_6 Winter_i + \hat{\beta}_7 Spring_i$$

$$\Leftrightarrow \widehat{Comp}_i = \hat{\beta}_0 + \hat{\beta}_1 Male_i + \underbrace{\frac{\hat{\beta}_2}{60}}_{neues\ \tilde{\beta}_2} \cdot 60 \cdot Work_i + \underbrace{\frac{\hat{\beta}_3}{3600}}_{neues\ \tilde{\beta}_3} (60 \cdot Work_i)^2 + \hat{\beta}_4 Sport_i + \hat{\beta}_5 Autumn_i + \hat{\beta}_6 Winter_i + \hat{\beta}_7 Spring_i$$

- Neues $\tilde{\beta}_2 = \hat{\beta}_2/60 = 15{,}041/60 = 0{,}251$ [1P]
- Neues $\tilde{\beta}_3 = \hat{\beta}_3/3600 = 0{,}473/3600 = 0{,}00013$ [1P]
- Der p-Wert von $\hat{\beta}_3$ wird durch die Transformation nicht beeinflusst. [1P]
  Begründung: Die Teststatistik $\left(t = \frac{\hat{\beta}_3/3600}{se(\hat{\beta}_3)/3600}\right)$ bleibt gleich.

Aufgabe 3: **[12,5 Punkte]**

Sie interessieren sich für die Determinanten der Lehrveranstaltungsevaluation von Dozenten. Für Ihre Analyse haben Sie zufällig 310 Lehrveranstaltungen ausgewählt, zu welchen Sie folgende Informationen haben:

Course_eval	Note der Lehrveranstaltung auf einer Skala stetig von 1 bis 10 kodiert ($Course\_eval = 1$: sehr schlecht; $Course\_eval = 10$: sehr gut)
Female	$= 1$ wenn Dozent weiblich, $= 0$ wenn Dozent männlich
Participation	Prozentualer Anteil der Note für mündliche Mitarbeit an der Endnote (von 0% bis 100%)
Beauty	Schönheit des Dozenten auf einer Skala stetig von 0 bis 5 kodiert ($Beauty = 1$: nicht schön; $Beauty = 5$: sehr schön)

Sie schätzen das folgende Modell mit SPSS:

$$Course\_eval_i = \beta_0 + \beta_1 Female_i + \beta_2 Participation_i + \beta_3 Beauty_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	5,172	0,135	30,904	0,000
Female	-0,174	0,052	-3,346	0,001
Participation	-0,036	0,085	-0,426	0,670
Beauty	0,181	0,031	5,839	0,000

a. Abhängige Variable: *Course_eval*

a) Interpretieren Sie den geschätzten Koeffizienten $\hat{\beta}_2$ inhaltlich und statistisch. (2 Punkte)

- $\hat{\beta}_2 = -0,036$: Wenn der Anteil der Note für mündliche Mitarbeit an der Endnote um einen Prozentpunkt steigt, so wird eine Lehrveranstaltung c.p. im Mittel um 0,036 Punkte schlechter bewertet. [1P]
- Der Koeffizient ist statistisch nicht signifikant von Null verschieden auf dem 10% Niveau. [1P]

b) SPSS berechnet Ihnen $\overline{R^2} = 0,44$ und $\sum_{i=1}^{310} \hat{u}_i^2 = 92,73$. Wie hoch ist die Gesamtvariation in *Course_eval* (SST)? Zeigen Sie Ihren Rechenweg. (3 Punkte)

- $\overline{R^2} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$
- Mit $\overline{R^2} = 0,44$, $SSR = 92,73$, $n = 310$ und $k = 3$ [2P]
- $0,44 = 1 - \frac{92,73/(310-3-1)}{SST/(310-1)} \Leftrightarrow SST = \frac{SSR/(n-k-1)}{-\overline{R^2}+1} \cdot (n-1) = \frac{92,73/(310-3-1)}{-0,44+1} \cdot (310-1) = 167,213$ [1P]
- Die Gesamtvariation in *Course_eval* beträgt 167,213.

c) Sie nehmen die Variable *Age* (Alter des Dozenten gemessen in Jahren) als vierte erklärende Variable in Ihr Modell auf und erhalten $\hat{\beta}_4 > 0$. Betrachten Sie folgende Korrelationen:

Korrelationen nach Pearson

	Female	*Participation*	*Beauty*	*Age*
Female	1	0,243	0,117	-0,149
Participation	0,243	1	0,052	-0,081
Beauty	0,117	0,052	1	-0,269
Age	-0,149	-0,081	-0,269	1

In welche Richtung würde sich der Effekt der Schönheit des Dozenten auf die Lehrveranstaltungsevaluation ändern, wenn Sie das neue Modell schätzen? Begründen Sie Ihre Antwort. (3 Punkte)

- Modell I: $Course\_eval_i = \beta_0 + \beta_1 Female_i + \beta_2 Participation_i + \beta_3 Beauty_i + u_i$
- Modell II: $Course\_eval_i = \tilde{\beta}_0 + \tilde{\beta}_1 Female_i + \tilde{\beta}_2 Participation_i + \tilde{\beta}_3 Beauty_i + \tilde{\beta}_4 Age_i + \nu_i$

- Der Effekt der Schönheit des Dozenten auf die Lehrveranstaltungsevaluation β_3 in Modell I ist verzerrt geschätzt, wenn $\beta_4 \neq 0$ und $Corr(Beauty, Age) \neq 0$
- Richtung der Verzerrung: Vorzeichen β_4 · Vorzeichen $Corr(Beauty, Age)$
- $\hat{\beta}_4 > 0$ und $Corr(Beauty, Age) = -0,269 < 0$ [1P]
- $\rightarrow \oplus \cdot \ominus = \ominus$: $\hat{\beta}_3$ in Modell I ist nach unten verzerrt [1P], denn der positive Effekt des Alters auf die Lehrveranstaltungsevaluation wurde durch $\hat{\beta}_3$ mitgemessen.
- Im Modell II sollte der Effekt der Schönheit größer sein als im Modell I. [1P]

Sie schätzen nun folgendes Modell:

$$Course\_eval_i = \beta_0 + \beta_1 Female_i + \beta_2 Participation_i + \beta_3 Beauty_i + \beta_4 Age_i + \beta_5 (Female_i \cdot Age_i) + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	4,573	0,182	25,126	0,000
Female	-0,155	0,043	-3,605	0,000
Participation	-0,035	**???**	**???**	**???**
Beauty	0,198	0,044	4,500	0,000
Age	0,034	0,011	3,091	0,002
Female · Age	0,017	0,005	3,009	0,003

a. Abhängige Variable: *Course_eval*

d) Interpretieren Sie den Effekt des Alters auf die Lehrveranstaltungsevaluation für Frauen inhaltlich. (2 Punkte)

- $\frac{\Delta \widehat{Course\_eval}}{\Delta Age}\Big|_{Female=1} = \hat{\beta}_4 + \hat{\beta}_5 = 0,034 + 0,017 = 0,051$ [1P]
- Eine Lehrverantaltung einer Dozentin wird mit jedem zusätzlichen Altersjahr der Dozentin c.p. im Mittel um $0,051$ Punkte besser bewertet. [1P]

e) Ihr Kollege merkt an, dass das 95%-Konfidenzintervall von $\hat{\beta}_2$ den Wert Null einschließt. Welche statistische Interpretation von $\hat{\beta}_2$ lässt sich daraus ableiten? (1,5 Punkte)

- $\hat{\beta}_2$ ist statistisch nicht signifikant von Null verschieden [1P] auf dem 5%-Niveau [0,5P].

f) Welchen Wert würden Sie für $\hat{\beta}_2$ erhalten, wenn *Course_eval* in umgekehrter Reihenfolge kodiert wäre? (*Course_eval* $= 10$: sehr schlecht; *Course_eval* $= 1$: sehr gut) (1 Punkt)

- $\hat{\beta}_2 = 0,035$ (Vorzeichen dreht sich um) [1P]

Aufgabe 4: **[7,5 Punkte]**

Im Rahmen Ihrer Forschung interessieren Sie sich für die Determinanten des Rauchens während des Studiums. Folgende Variablen stehen Ihnen zur Verfügung:

smoke	$= 1$ wenn Person Raucher ist, $= 0$ sonst
income	Monatseinkommen in Euro
party	Anzahl der pro Monat besuchten Partys
lastsem	$= 1$ wenn Person im letzten Semester studiert, $= 0$ sonst

Sie schätzen folgendes lineares Wahrscheinlichkeitsmodell in SPSS:

$$smoke_i = \beta_0 + \beta_1 income_i + \beta_2 party_i + \beta_3 lastsem_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	0,034	0,008	4,250	0,000
income	0,001	0,001	1,501	0,134
party	0,073	0,037	1,973	0,049
lastsem	0,119	0,041	2,902	0,004

a. Abhängige Variable: *smoke*

a) Nennen Sie zwei Schwächen des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

Mögliche Antworten [jeweils 1P]

- Es ist möglich, dass die vorhergesagten Werte außerhalb des (0,1)-Intervalls liegen.
- Es ist oft unplausibel, dass einzelne Variablen über ihren gesamten Wertebereich linear mit der abhängigen Variablen zusammenhängen.
- Die Störtermvarianz ist heteroskedastisch.

b) Interpretieren Sie die geschätzten Koeffizienten $\hat{\beta}_2$ und $\hat{\beta}_3$ inhaltlich und statistisch. (4 Punkte)

- $\hat{\beta}_2 = 0,073$
  → Die Wahrscheinlichkeit, Raucher zu sein, steigt c.p. im Mittel um $7,3$ Prozentpunkte pro zusätzlichem monatlichen Partybesuch. [1P]
  → Der geschätzte Koeffizient ist am 5%-Niveau statistisch signifikant von Null verschieden. [1P]
- $\hat{\beta}_3 = 0,119$
  → Die Wahrscheinlichkeit, Raucher zu sein, ist c.p. im Mittel für Stundierende im letzten Semester um $11,9$ Prozentpunkte höher verglichen mit Studierenden, die nicht im letzten Semester studieren. [1P]
  → Der geschätzte Koeffizient ist am 1%-Niveau statistisch signifikant von Null verschieden. [1P]

c) Wie hoch ist für einen Studierenden im letzten Semester, der keine Partys besucht und ein Monatseinkommen von 500 Euro hat, die Wahrscheinlichkeit, Raucher zu sein? (1,5 Punkte)

- Gesucht: $\widehat{smoke}_i$ für eine Person mit $income_i = 500$, $party_i = 0$ und $lastsem_i = 1$
- $\widehat{smoke}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot 500 + \hat{\beta}_2 \cdot 0 + \hat{\beta}_3 \cdot 1 = 0,034 + 0,001 \cdot 500 + 0,119 = 0,653$ [1P]
- Die Wahrscheinlichkeit, dass eine Person mit den gegebenen Eigenschaften Raucher ist, beträgt im Mittel $65,3$ Prozent. [0,5P]

Aufgabe 5 - MC Fragen **[40 Punkte]**

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. Zu jeder Frage gibt es genau eine richtige Antwort. Für jede korrekt beantwortete Frage erhalten Sie einen Punkt. Falsche Antworten führen nicht zu Punktabzug. Bei mehr oder weniger als einer markierten Antwort auf eine Frage gilt diese als nicht beantwortet. **Angaben auf dem Aufgabenblatt werden nicht gewertet.**

1.	Wird die Nullhypothese eines Strukturbruchtests abgelehnt, so...
a	**X** liegen signifikante Unterschiede in den Steigungsparametern verschiedener Gruppen vor.
b	liegt Heteroskedastie vor.
c	liegt perfekte Multikollinearität vor.
d	liegt ein Endogenitätsproblem vor.

2.	Gegeben sei $y_i = \beta_0 + \beta_1 D_i + u_i$, wobei D_i eine binäre Variable darstellt. Welche Schätzergebnisse ändern sich, wenn Sie in dem Modell D_i durch $(1 - D_i)$ ersetzen?
a	$\hat{\beta}_1$ und SST
b	$\hat{\beta}_0$ und R^2
c	$\hat{\beta}_1$ und R^2
d	**X** $\hat{\beta}_0$ und $\hat{\beta}_1$

3.	Für das Modell $Ln(Lohn_i) = \beta_0 + \beta_1 Alter_i + \beta_2 Alter_i^2 + u_i$ schätzen Sie folgende statistisch signifikante Koeffizienten: $\hat{\beta}_0 = 8,7$, $\hat{\beta}_1 = 0,4$, $\hat{\beta}_2 = -0,005$. Der marginale Effekt des Alters auf den logarithmierten Stundenlohn...
a	beträgt für jedes Alter 0,4.
b	ist negativ für $Alter_i = 35$.
c	**X** ist negativ für $Alter_i = 45$.
d	ist positiv für $Alter_i = 55$.

4.	Der kritische Wert eines t-Tests in dem Modell $ln(y_i) = \beta_0 + \beta_1 ln(x_i) + u_i$ mit der Alternativhypothese $H_1 : \beta_1 > 0,5$, mit 30 Beobachtungen und einer Irrtumswahrscheinlichkeit von $\alpha = 0,01$ beträgt...
a	$2,326$.
b	$1,701$.
c	$2,048$.
d	**X** $2,467$.

5.	Ein Konfidenzintervall...
a	für einen positiven Punktschätzer hat stets positive Intervallgrenzen.
b	**X** wird bei sinkender Stichprobengröße breiter.
c	kann für die Differenz zweier Punktschätzer nicht berechnet werden.
d	umfasst stets den Wert des R^2.

6.	Die F-Statistik...
a	**X** kann mittels SSR oder R^2 berechnet werden.
b	kann nicht dafür genutzt werden, die Signifikanz eines einzelnen Parameters zu testen.
c	ist asymptotisch normalverteilt.
d	nimmt stets Werte zwischen 0 und 1 an.

7.	Der geschätzte Koeffizient $\hat{\beta}_1$ des Modells $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ ist nach oben verzerrt, falls eine ausgelassene Variable z_i...
a	y_i positiv beeinflusst und $Cov(x_{i1}, z_i) < 0$.
b	**X** y_i positiv beeinflusst und $Cov(x_{i1}, z_i) > 0$.
c	y_i negativ beeinflusst und $Cov(x_{i1}, x_{i2}) < 0$.
d	y_i positiv beeinflusst und $Cov(x_{i2}, z_i) > 0$.

8.	Der marginale Effekt von x_1 im Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 \cdot ln(x_2) + u_i$ lautet...
a	$\beta_1 + ln(x_2)\beta_3$.
b	$\beta_1 + \beta_3$.
c	$\beta_1 + ln(x_2)\beta_3 x^2$.
d	**X** Keine der Antworten ist korrekt.

9.	Für die Schätzgleichung $\hat{y}_i = 2,4 - 0,8 \cdot x_i$ und die Beobachtung $(y_1, x_1) = (3,5)$ beträgt das Residuum $\hat{u}_1$...
a	**X** $4,6$.
b	$3,4$.
c	0.
d	$-1,6$.

10.	Sie regressieren in einem einfachen linearen Modell die Schlafzeit in Stunden auf eine Konstante und die Arbeitszeit in Stunden. Welche Schätzergebnisse würden sich ändern, wenn sowohl die Schlafzeit als auch die Arbeitszeit in Minuten gemessen wären?
a	$\hat{\beta}_1$.
b	**X** $\hat{\beta}_0$.
c	$\hat{\beta}_0$ und $\hat{\beta}_1$.
d	$\hat{\beta}_1$ und $se(\hat{\beta}_1)$.

11.	Für den KQ-Schätzer gilt:
a	Jeder vorhergesagte Wert der Residuen liegt auf der Regressionsgeraden.
b	**X** y_i lässt sich als Summe von Vorhersage und Residuum abbilden.
c	Modus und Median der KQ-Residuen sind Null.
d	$\frac{(SSR+SSE)}{SST} - 1 = 1$.

12.	Das KQ-Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ ...
a	**X** ist linear in den Parametern.
b	enthält drei Steigungsparameter.
c	beschreibt den nicht-linearen Zusammenhang zwischen y und x_1.
d	unterstellt bei Gültigkeit der Annahme $E(u) = 0$, dass alle aufgenommenen Variablen in der Grundgesamtheit einen Mittelwert von Null haben.

13.	Für die Varianz eines mittels KQ-Schätzer geschätzten Parameters $\hat{\beta}_1$, $var(\hat{\beta}_1)$, gilt c.p.:
a	**X** Je größer $\hat{\sigma}^2$, umso größer $var(\hat{\beta}_1)$.
b	Je größer die Stichprobe, umso größer ist $var(\hat{\beta}_1)$.
c	Je kleiner die Stichprobe, umso kleiner ist $var(\hat{\beta}_1)$.
d	Je größer die geschätzte Konstante, umso größer ist $var(\hat{\beta}_1)$.

14.	Wird im Modell $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ mit $\beta_0 \neq 0$ die Konstante β_0 nicht mitgeschätzt, so...
a	**X** steigt die Zahl der Freiheitsgrade.
b	werden die Steigungsparameter unverzerrt geschätzt.
c	ändert sich der R^2-Wert nicht.
d	geht dies einher mit sinkenden Standardfehlern der geschätzten Steigungsparameter.

15.	Unter den Annahmen MLR.1-5 ist KQ...
a	das praktischste, logischste, unkonsistente Schätzverfahren.
b	das beste, logistische, unorthodoxe Schätzverfahren.
c	das einfachste, latente, unverzerrte Schätzverfahren.
d	**X** das beste, lineare, unverzerrte Schätzverfahren.

16.	Ist ein Schätzer konsistent, so bedeutet dies, dass...
a	die Varianz des Schätzers bei zunehmender Beobachtungszahl steigt.
b	Parameterschätzer auch bei kleinen Stichproben im Mittel stets den wahren Wert angeben.
c	**X** geschätzte Parameter umso näher am wahren Wert liegen, je größer die Beobachtungszahl.
d	der Schätzer unverzerrt ist.

17.	Wenn im einfachen linearen Modell Homoskedastie vorliegt, so...
a	ist der KQ-Schätzer inkonsistent.
b	ist der KQ-Schätzer ineffizient aber erwartungstreu.
c	ist die Varianz des Fehlerterms u für alle Beobachtungen unterschiedlich.
d	**X** ist die Varianz des Fehlerterms u für alle Beobachtungen gleich.

18.	Typ I-Fehlerwahrscheinlichkeiten...
a	**X** sind für F-Tests relevant.
b	werden durch t-Werte beschrieben.
c	beschreiben die Wahrscheinlichkeit, die Alternativhypothese nicht zu verwerfen.
d	unterscheiden sich für positive und negative geschätzte Parameter.

19.	Ein nicht-linearer Zusammenhang zwischen zwei Variablen...
a	wird mit dem Chow-Test gemessen.
b	verursacht in der Regel omitted variable bias.
c	**X** kann in einem multiplen Regressionsmodell abgebildet werden.
d	verursacht Heteroskedastie.

20.	Das Gauss-Markov Theorem...
a	basiert auf der Normalverteilungsannahme.
b	macht eine Aussage über die Konsistenz der KQ-Schätzer.
c	**X** macht eine Aussage über die Effizienz der KQ-Schätzer.
d	basiert auf der Annahme heteroskedastischer Störterme.

21.	Der KQ-Schätzer...
a	maximiert die Summe der quadrierten vertikalen Abweichungen von der Regressionsgeraden.
b	maximiert die Residuenquadratsumme.
c	maximiert den unerklärten Teil der Streuung der abhängigen Variablen.
d	**X** maximiert R^2.

22.	Sei X eine Zufallsvariable mit $E(X) = \mu$ und Varianz σ. Somit gilt, dass...
a	**X** $var(X) = E[(X-\mu)^2] = \sigma^2$
b	$sd(X) = \sqrt[2]{\sigma}$
c	$sd(aX) = a^2\sigma^2$
d	$var(X) < 0$

23.	Ein Schätzer eines unbekannten Parameters auf Basis von Stichprobendaten...
a	ist dann effizient, wenn der unbekannte Parameter einer t-Verteilung folgt.
b	ist verzerrt, wenn sein Erwartungswert mit dem wahren Wert identisch ist.
c	**X** ist eine Zufallsvariable.
d	ist unverzerrt, wenn bei $n \to \infty$ der geschätze Paramter gegen den wahren Wert konvergiert.

24.	Man erhält den KQ-Schätzer im einfachen linearen Modell $y = \beta_0 + \beta_1 x + u$ durch Minimierung folgender Zielfunktion:
a	$\sum_{i=1}^{n}(y_i - \bar{y})$
b	$\sum_{i=1}^{n}(x_i - x_i^2)^{\frac{1}{2}}$
c	**X** $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
d	$\sum_{i=1}^{n}(x_i - \hat{x}_i)^{\sqrt{2}}$

25.	Das R^2-Maß...
a	kann mit zunehmender Anzahl erklärender Variablen sinken.
b	**X** kann zum Modellvergleich herangezogen werden.
c	kann genauso interpretiert werden wie das angepasste R^2.
d	gibt den Anteil der durch das Modell erklärten Variation der unabhängigen Variable an.

26.	Ein R^2-Maß von $0,50$ zeigt an, dass...
a	$SSR > SST$.
b	$SSE < SSR$.
c	**X** $SSR = SSE$.
d	$SSE = SST$.

27.	Der Korrelationskoeffizient ρ für zwei Variablen X und Y...
a	hat das gleiche Vorzeichen wie das Produkt aus σ_X und σ_Y.
b	**X** misst den linearen Zusammenhang zwischen X und Y.
c	hat den Wertebereich $log(min(X)) \leq \rho \leq log(min(Y))$.
d	kann nicht negativ werden.

28.	Für eine normalverteilte Zufallsvariable $X \sim N(\mu, \sigma^2)$ gilt, dass...
a	**X** Modus $=$ Median.
b	Median $\neq$ Mittelwert.
c	der Wert am 10. Perzentil dem Wert am 90. Perzentil entspricht.
d	Die größte Merkmalsausprägung mit der höchsten Wahrscheinlichkeit auftritt.

29.	Die gesamte Quadratsumme SST...
a	wird verwendet, um die Stichprobenvarianz der unabhängigen Variablen zu berechnen.
b	wird mittels eines F-Tests ermittelt.
c	hängt von der Anzahl der Steigungsparameter ab.
d	**X** ist für genestete Modelle identisch.

30.	Ergibt sich in einem einfachen linearen Modell nach einer KQ-Schätzung für die erste Beobachtung ein Residuum von $\hat{u}_1 = 0$, so...
a	kann geschlossen werden, dass es keine Probleme mit ausgelassenen Variablen gibt.
b	ist davon auszugehen, dass $\hat{u}_2 = 0$.
c	**X** liegt y_1 auf der Regressionsgeraden.
d	liegt $\hat{y}_1$ nicht auf der Regressionsgeraden.

31.	Beim Umskalieren der unabhängigen Variable mit einem konstanten multiplikativen Faktor im Modell $y_i = \beta_0 + \beta_1 x_i + u_i$ ändert sich...
a	die Konstante.
b	**X** der Steigungsparameter.
c	das Bestimmtheitsmaß.
d	die Quadratsumme der Residuen.

32.	Zwei Koeffizienten sind gemeinsam signifikant, wenn...
a	die 0 in beiden Konfidenzintervallen enthalten ist.
b	beide Koeffizienten einzeln signifikant sind.
c	mindestens einer der beiden Koeffizienten signifikant ist.
d	**X** Keine der Antworten ist korrekt.

33.	In dem Modell $y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$ mit den zwei unkorrelierten Dummyvariablen D_{1i} und D_{2i} gilt:
a	Der Effekt von D_{1i} hängt davon ab, ob D_{2i} gleich 0 oder 1 ist.
b	β_1 gibt an, um wieviel sich der Erwartungswert von y_i ändert, wenn D_{2i} von 0 auf 1 steigt.
c	β_2 kann nur Werte zwischen 0 und 1 annehmen.
d	**X** β_0 entspricht dem Erwartungswert von y_i, wenn beide Dummyvariablen gleich 0 sind.

34.	In einer Lohnregression $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u_i$
a	wird der marginale Effekt von *exper* als partielle Ableitung von *wage* nach $exper^2$ berechnet.
b	ist β_1 als Elastizität interpretierbar.
c	**X** hängt der Effekt von *educ* auf $log(wage)$ nicht vom Ausgangsniveau von *exper* ab.
d	kann der Effekt von *exper* auf $log(wage)$ nicht negativ werden.

35.	Beim Hinzufügen eines neuen Regressors...
a	fällt das R^2 der Schätzung.
b	kann das $\overline{R^2}$ nicht negativ werden.
c	**X** kann die Fehlerquadratsumme (SSR) nicht steigen.
d	steigt der Standardfehler der Regression. ($\hat{\sigma}_u$)

36.	Von Heteroskedastie spricht man beispielsweise, wenn...
a	die Konstante mit dem Fehlerterm korreliert ist.
b	Parameter in nicht-linearer Weise in das Regressionsmodell aufgenommen werden.
c	ein Problem ausgelassener Variablen vorliegt.
d	**X** die Varianz des Fehlerterms von den Ausprägungen der erklärenden Variablen abhängt.

37.	Perfekte Multikollinearität...
a	tritt nur in Modellen mit binären Variablen auf.
b	**X** lässt sich durch das Weglassen einer erklärenden Variable vermeiden.
c	führt zu kollinearen Parametern.
d	lässt sich vermeiden, wenn man den Stichprobenumfang erhöht.

38.	Der Zusammenhang zwischen zwei Variablen...
a	wird durch den Standardfehler gemessen.
b	wird mit der Standardabweichung gemessen.
c	**X** wird mit dem Korrelationskoeffizienten als linearer Zusammenhang gemessen.
d	wird nur mit experimentellen Daten verlässlich gemessen.

39.	Die F-Verteilung...
a	ist eine linksschiefe Verteilung.
b	**X** ist eine rechtsschiefe Verteilung.
c	ist eine symmetrische Verteilung.
d	Keine der Antworten ist korrekt.

40.	Wenn die Störterme einer multiplen Regression normalverteilt sind, dann...
a	ist der KQ-Schätzer inkonsistent.
b	**X** sind t- und F-Tests auch in kleinen Stichproben gültig.
c	kann man die geschätzten Koeffizienten als kausale Effekte interpretieren.
d	ist der KQ-Schätzer asymptotisch ineffizient.