Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Prof. Regina T. Riphahn, Ph.D.

Bachelorprüfung SS 2020 - MUSTERLÖSUNG

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen. **Es wird nur der Lösungsbogen eingesammelt.** Angaben auf dem Aufgabenzettel werden nicht gewertet.
Bewertung:	Es können maximal 60 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 wollen die Determinanten der wöchentlichen Arbeitsstunden von Angestellten untersuchen. Ihnen liegen Daten eines Unternehmens mit 217 Beobachtungen vor:

$Hours_i$	monatliche Arbeitszeit in Stunden
$Educ_i$	Bildung in Jahren
Age_i	Alter in Jahren
$Age2_i$	quadriertes Alter
$Income_i$	monatliches Einkommen in Euro

Sie schätzen das folgende Modell:

$$Hours_i = \beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Age2_i + \beta_4 Income_i + u_i$$

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	7869	4	2623195	29,76	,000(a)
	Residuen	33478	212	88131		
	Gesamt	41347	216			

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	0,313(a)	?,???	?,???	?,???

Koeffizienten[a]

	Nicht standardisierte Koeffizienten			
Modell	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(*Konstante*)	9,404	6,442	1,460	0.144
Educ	0,058	0,005	10,663	0,000
Age	2,05	0,448	4,574	0,000
Age2	-0,115	0,0025	-4,600	0,000
Income	0,063	0,013	4,846	0,043

a. Abhängige Variable: *Hours*

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

a) Interpretieren Sie den geschätzten Koeffizienten von *Income* inhaltlich. Ist der Effekt statistisch signifikant? (2 Punkte)

- $\hat{\beta}_4 = 0,063$
- `Steigt das monatliche Einkommen um 1 Euro, so steigt die monatliche Arbeitszeit c.p. im Mittel um 0,063 Stunden. [1P]`
- `Der Koeffizient ist statistisch signifikant auf dem 5 %-Niveau. [1P]`

b) Testen Sie auf dem 5%-Niveau, ob *Age* und *Age2* gemeinsam signifikant sind. Geben Sie das Testverfahren, die Nullhypothese, die Alternativhypothese, die Teststatistik, den kritischen Wert und die Testentscheidung an. (Hinweis: Residuenquadratsumme (SSR) der Schätzung ohne *Age* und *Age2*: 34899.) (5,5 Punkte)

- F-Test auf gemeinsame Signifikanz [0,5P]
- $H_0 : \beta_2 = \beta_3 = 0$ [0,5P]
- H_1: mind. ein $\beta_i \neq 0$ für $i = 2,3$ [0,5P]
- Teststatistik: $F^{empirisch} = \frac{(SSR_R - SSR_U)/q}{(SSR_U)/(n-k-1)} = \frac{(34899-33478)/2}{(33478)/(217-4-1)} = 4,499$ [2,5P] je ein halber Punkt auf SSR_U, SSR_R (n-k-1), q und das Ergebnis
- kritischer Wert: $\alpha = 0,05$: $F^{kritisch} = F(0,05;2;\infty) = 3,0$ [0,5P]
- Testentscheidung: $F^{empirisch} = 4,499 > F^{kritisch} = 3,000$. Die Nullhypothese kann abgelehnt werden. Die beiden Variablen haben einen gemeinsamen signifikanten Einfluss auf die wöchentliche Arbeitszeit. [1P.]

c) Bestimmen und interpretieren Sie das 90%-Konfidenzintervall von β_1. (3 Punkte)

- KI $= [\hat{\beta}_1 - t_{\frac{\alpha}{2};n-k-1} \cdot se(\beta_1);\ \hat{\beta}_1 + t_{\frac{\alpha}{2};n-k-1} \cdot se(\beta_1)]$
- KI $= [0,058 - t_{0,05;212} \cdot 0,005; 0,058 + t_{0,05;212} \cdot 0,005;]$ [1P.] jeweils halber Punkt auf 0,058 und 0,005
- KI $= [0,058 - 1,645 \cdot 0,005; 0,058 + 1,645 \cdot 0,005]$ [0,5P.]
- KI $= [0,050; 0,066]$ [0.5P.]
- Für wiederholte Stichproben liegt in 90% der Fälle der wahre Wert β_1 innerhalb der auf diese Weise berechneten Intervallgrenzen. [1P.]

d) Berechnen Sie die geschätzte Arbeitszeit einer Person mit Abitur (12 Jahre) und Bachelorabschluss (3 Jahre), welche 30 Jahre alt ist und 3000 Euro im Monat verdient. (2,5 Punkte)

- $\widehat{Hours_i} = 9,404 + 0,058 \cdot 15 + 2,05 \cdot 30 + (-0,115) \cdot 30^2 + 0,063 \cdot 3000 = 157,274$
- je einen halben Punkt auf $Educ = 15$, $Age = 30$, $Age^2 = 30^2$, $Income = 3000$ und Endergebnis

Aufgabe 2: **[15 Punkte]**

Sie untersuchen, welche Faktoren die erreichte Punktzahl in der PEWI Klausur beeinflussen. Ihnen liegen Daten aus vorherigen Semestern mit 524 Beobachtungen vor:

$Punkte_i$	erreichte Punkte in der Klausur auf einer Skala von 0 bis 60
$Lernzeit_i$	wöchentlicher Lernaufwand in Stunden
$Frau_i$	=1, wenn Frau; =0, wenn Mann
$Party_i$	Anzahl der Partybesuche in den letzten 4 Wochen

Sie schätzen das folgende Modell:

$$Punkte_i = \beta_0 + \beta_1 Lernzeit_i + \beta_2 Frau_i + \beta_3 Party_i + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	0,124(a)	0,078	?,???	0,387

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(*Konstante*)	15,346	6,442	2,382	0,095
Lernzeit	1,367	0,245	**???**	**???**
Frau	5,189	4,587	1,131	0,127
Party	-2,476	1,155	-2,144	0,087

a. Abhängige Variable: *Punkte*

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

a) Interpretieren Sie den geschätzten Koeffizienten von *frau*. Ist der Effekt statistisch signifikant? Begründen Sie Ihre Antwort. (2 Punkte)

- $\hat{\beta}_2 = 5,189$
- Frauen erzielen c.p. im Mittel 5,189 Klausurpunkte mehr als Männer [1P]
- Der Koeffizient ist statistisch nicht signifikant auf dem 10%-Niveau. [1P]

b) Interpretieren Sie das R^2 und berechnen Sie das korrigierte Bestimmtheitsmaß $\bar{R}^2$. (3 Punkte)

- 7,8% der Variation in den Punkten kann durch das Modell erklärt werden [1P]
- $\bar{R}^2 = 1-(1-R^2)\cdot\frac{n-1}{n-k-1} = 1-(1-0,078)\cdot\frac{524-1}{524-3-1} = 0,073$ [2P] 0,5P für das Ergebnis, n und k und R^2

c) Testen Sie auf dem Signifikanzniveau von 5%, ob der wöchentliche Lernaufwand für PEWI einen positiven Einfluss auf die erreichte Punktzahl in der Klausur hat. Geben Sie das Testverfahren, die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert und die Testentscheidung an. (5 Punkte)

- Testverfahren: Einseitig, rechtsseitiger Test [1P]
- Nullhypothese: $H_0 = \hat{\beta}_1 \leq 0$ [0,5P]
- Alternativhypothese: $H_1 = \hat{\beta}_1 > 0$ [0,5P]
- Teststatistik: $t_{empirisch} = \frac{\hat{\beta}_1 - 0}{0,245} = \frac{1,367-0}{0,245} = 5,580$ [1P]
- kritischer Wert c: $c = t_{\alpha;n-k-1} = t_{0,05;524-3-1} = 1,645$ [1P]
- $t_{empirisch} = 5,580 > 1,645 = c$: Die Nullhypothese kann auf dem 5%-Niveau verworfen werden. Der wöchentliche Lernaufwand hat einen signifikant positiven Einfluss auf die erreichte Punktzahl in der Klausur. [1P.]

d) Berechnen und interpretieren Sie inhaltlich den geschätzen Effekt auf die erreichte Punktzahl in der Klausur bei einem Anstieg der Partybesuche pro Woche um 1. (2 Punkte)

- Anstieg um 1 Partybesuch pro Woche $\widehat{=}$ 4 zusätzliche Partybesuche in den letzten 4 Wochen
- $\widehat{punkte}_i = \hat{\beta}_3 \cdot \Delta party_i \Leftrightarrow \widehat{punkte}_i = (-2,476)\cdot 4 = -9,904$ [1P]
- Wenn sich die Anzahl der Partybesuche pro Woche um 1 erhöht, dann sinkt die Anzahl der erreichten Punkte c.p. im Mittel um 9,904 Punkte. [1P]

e) Sie nehmen zusätzlich die Variable $Buch_i$ in das Modell auf ($Buch_i = 1$, wenn der Studierende das Lehrbuch genutzt hat und = 0, wenn nicht). Nehmen Sie an, dass $Buch_i$ und $Lernzeit_i$ stark miteinander korreliert sind. Wie beeinflusst dies die Effizienz der Schätzung? (3 Punkte)

- Eine zusätzlich aufgenommene relevante Variable senkt die Variation des Störterms und erhöht damit die Effizienz der Schätzung [1P.]
- Die starke Korrelation zwischen $Buch_i$ und $Lernzeit_i$ senkt die Effizienz der Schätzung [1P.]
- Insgesamt kann keine eindeutige Aussage getroffen werden. Es kommt darauf an, welche der beiden Wirkungen auf die Effizienz der Schätzung dominiert [1P.]

Aufgabe 3: **[15 Punkte]**

Sie interessieren sich für Siegwahrscheinlichkeiten im Fußball und haben Daten zu 68 Spielen des 1. FC Nürnberg (FCN) aus der 1. Bundesliga mit folgenden Informationen gesammelt:

$Sieg_i$	$= 1$, wenn der FCN das Spiel i gewonnen hat; $= 0$ sonst
$DGehalt_i$	Durchschnittliches monatliches Gehalt in 1000 Euro aller eingesetzten Spieler des FCN bei Spiel i
$DGehalt2_i$	Quadriertes durchschnittliches monatliches Gehalt in 1000 Euro aller eingesetzten Spieler des FCN bei Spiel i
$Fans_i$	Anteil von Fans des FCN an allen Zuschauern im Stadion bei Spiel i (von 0 bis 100 %)
$Heimspiel_i$	$= 1$, wenn Spiel i in Nürnberg stattfindet, $= 0$ sonst

Sie stellen folgendes Regressionsmodell auf und schätzen dieses anschließend mit SPSS:

$$Sieg_i = \beta_0 + \beta_1 DGehalt_i + \beta_2 DGehalt2_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(*Konstante*)	-0,981	0,350	-2,803	0,002
DGehalt	0,147	0,034	4,262	0,000
*DGehalt*2	-0,003	0,001	-2,840	0,003

a. Abhängige Variable: *Sieg*

a) Berechnen Sie den gesamten marginalen Effekt der Variable *DGehalt* auf die Siegwahrscheinlichkeit. Bei welchem Durchschnittsgehalt der Mannschaft wird die Siegwahrscheinlichkeit maximiert? Wie lautet die höchste mit Hilfe des geschätzten Modells vorhergesagte Siegwahrscheinlichkeit? (4 Punkte)

- Berechnung marginaler Effekt: $\frac{\Delta \widehat{Sieg_i}}{\Delta DGehalt_i} = \hat{\beta}_1 + 2 \cdot \hat{\beta}_2 DGehalt_i = 0,147 - 2 \cdot 0,003 \cdot DGehalt_i$ [1P]
- $0,147 - 2 \cdot 0,003 \cdot DGehalt_i = 0 \Leftrightarrow DGehalt_i = 24,5$ [1P]
- Die Siegwahrscheinlichkeit wird maximal für ein Durchschnittsgehalt der Mannschaft von 24500 Euro.
- $\widehat{Sieg}_{max} = -0,981 + 0,147 \cdot 24,5 - 0,003 \cdot (24,5)^2 = 0,82$. [1P]
- Die maximal erwartete Siegwahrscheinlichkeit beträgt 82%. [1P]

b) Welche Werte würden $\hat{\beta}_1$ und $\hat{\beta}_2$ annehmen, wenn das Durchschnittsgehalt der Mannschaft nicht in 1.000 Euro, sondern in 10.000 Euro gemessen worden wäre? (2 Punkte)

- Umskalierung der unabhängigen Variable $Gehalt_i$:
- $\widehat{Sieg_i} = \hat{\beta}_0 + \hat{\beta}_1 DGehalt_i + \hat{\beta}_2 DGehalt2_i \Leftrightarrow \widehat{Sieg_i} = \hat{\beta}_0 + (\hat{\beta}_1 \cdot 10) \cdot \frac{DGehalt_i}{10} + (\hat{\beta}_2 \cdot 100) \cdot \frac{DGehalt2_i}{100}$
- $\tilde{\beta}_1 = \hat{\beta}_1 \cdot 10 = 0,147 \cdot 10 = 1,470$ [1P]
- $\tilde{\beta}_2 = \hat{\beta}_2 \cdot 100 = -0,003 \cdot 100 = -0,300$ [1P]

Sie erweitern das Modell um die Variablen $Fans_i$ und $Heimspiel_i$:

$$Sieg_i = \beta_0 + \beta_1 DGehalt_i + \beta_2 DGehalt2_i + \beta_3 Fans_i + \beta_4 Heimspiel_i + u_i$$

c) SPSS berechnet [0,009;0,013] als 95%-Konfidenzintervall für $\hat{\beta}_3$. Berechnen und interpretieren Sie inhaltlich den Effekt eines Anstiegs des Anteils der Fans des FCN um 2 Prozentpunkte auf die Siegwahrscheinlichkeit der Mannschaft. Ist der geschätzte Parameter der Variable $Fans_i$ statistisch signifikant von Null verschieden auf dem 5%-Signifikanzniveau? (3 Punkte)

- Berechnung $\hat{\beta}_3$: $\frac{0,013+0,009}{2} = 0,011$ [1P]
- Inhaltliche Interpretation: Steigt der Anteil der eigenen Fans im Stadion um 2 Prozentpunkte, so erhöht sich c.p. im Durchschnitt die Siegwahrscheinlichkeit um $0,011 \cdot 2 \cdot 100 = 2,2$ Prozentpunkte. [1P]
- Da das Konfidenzintervall die Null nicht beinhaltet, ist der Effekt statistisch signifikant von Null verschieden auf dem 5%-Niveau. [1P]

d) Stellen Sie eine Modellgleichung auf, mit welcher Sie testen können, ob sich sowohl der Gehaltseffekt als auch der Faneffekt zwischen Heim- und Auswärtsspielen unterscheidet. (3 Punkte)

- $Sieg_i = \beta_0 + \beta_1 DGehalt_i + \beta_2 DGehalt2_i + \beta_3 Fans_i + \beta_4 Heimspiel_i + \delta_1 Heimspiel_i \cdot DGehalt_i + \delta_2 Heimspiel_i \cdot DGehalt2_i + \delta_3 Heimspiel_i \cdot Fans_i + u_i$
- Je 1P pro Interaktionseffekt.

e) Sie vermuten, dass der Anteil der Fans der gegnerischen Mannschaft im Stadion während des Spiels ($Fans\_Gegner_i$) einen Einfluss auf die Siegwahrscheinlichkeit hat und nehmen die Variable in das Modell auf. Welches Problem tritt bei der Schätzung auf? Wie lässt es sich lösen? Begründen Sie Ihre Antwort knapp. (3 Punkte)
Hinweis: Nehmen Sie an, dass nur Fans der beiden Mannschaften anwesend sind.

- Es besteht das Problem der perfekten Multikollinearität. [1P]
- Die Variablen $Fans_i$ und $Fans\_Gegner_i$ sowie die Konstante sind linear abhängig [1P]; es ist nicht möglich, beide Variablen und die Konstante gleichzeitig in die Regression aufzunehmen.
- Lösung: Entweder die Konstante oder $Fans_i$ oder $Fans\_Gegner_i$ aus dem Modell nehmen. (Eine der Möglichkeiten reicht aus) [1P]

Aufgabe 4: **[17 Punkte]**

Sie haben Einkommensdaten für 3000 vollzeitbeschäftigte 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_i	logarithmierter Stundenlohn
$male_i$	$= 1$ wenn Person männlich, $= 0$ wenn Person weiblich
$married_i$	$= 1$ wenn Person verheiratet, $= 0$ sonst
$exper_i$	Berufserfahrung in Jahren
$public_i$	$= 1$ wenn Person im öffentlichen Dienst arbeitet, $= 0$ sonst

Sie 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	0,027	5,296	0,002

a. Abhängige Variable: *ln_hwage*

Hinweis: Das R^2 dieser Schätzung beträgt $0,375$.
Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

a) Berechnen Sie $se(\hat{\beta}_0)$, $\hat{\beta}_1$ und $t(\hat{\beta}_3)$. Ist der geschätzte Koeffizient für β_4 am 5%-Niveau statistisch signifikant? Begründen Sie kurz. (2,5 Punkte)

- $se(\hat{\beta}_0) = \frac{\hat{\beta}_0}{t(\hat{\beta}_0)} = \frac{1,575}{8,750} = 0,180$ [0,5P]
- $\hat{\beta}_1 = t(\hat{\beta}_1) \cdot se(\hat{\beta}_1) = 2,107 \cdot 0,056 = 0,118$ [0,5P]
- $t(\hat{\beta}_3) = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0,052}{0,022} = 2,364$ [0,5P]
- Da $t(\hat{\beta}_4) = 5,296 > 1,96$ (kritischer Wert der t-Verteilung bei 2995 Freiheitsgraden) ist der geschätzte Koeffizient statistisch signifikant am 5%-Niveau (Alternative Antwort möglich). [1P]

b) Berechnen und interpretieren Sie den genauen (!) Effekt der Variable $public_i$ auf den Stundenlohn. (2 Punkte)

- $\hat{\beta}_4 = 0,143$
- Der genaue Effekt beträgt $e^{\hat{\beta}_4} - 1 = 0,154$. [1P]
- Eine im öffentlichen Dienst beschäftigte Person verdient c.p. im Mittel 15,4 Prozent mehr pro Stunde als eine nicht im öffentlichen Dienst beschäftigte Person. [1P]

c) Erklären Sie anhand eines konkreten Beispiels, unter welchen Bedingungen der Koeffizient der Variable $public_i$ verzerrt geschätzt wird. (3 Punkte)

- Es kommt zu einer Verzerrung, wenn es eine ausgelassene Variable gibt, die sich auf den Stundelohn auswirkt und gleichzeitig mit der Beschäftigung im öffentlichen Dienst korreliert ist. [2P]
- Beispiel: Verzerrung, falls Personen mit höherer Schulbildung im öffentlichen Dienst arbeiten und Schulbildung einen Effekt auf den Stundenlohn hat. Andere Antworten möglich. [1P]

d) Sie interessieren sich dafür, ob der Einfluss der Variable $exper_i$ 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 $0,382$.

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 aufgenommene Variable einen signifikanten Erklärungsbeitrag in Ihrem Modell liefert? Mit welchem Test können Sie die Signifikanz von $exper^2$ hier testen? 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. F-Test [1P]
2. Nullhypothese: $H_0 : \beta_4 = 0$ [0,5P]
3. Alternativhypothese: $H_1 : \beta_4 \neq 0$ [0,5P]
4. Freiheitsgrade (aus unrestringiertem Modell): $n-k-1 = 3000-5-1 = 2994$ [0.5P] und $q = 1$ [0.5P]
5. Teststatistik: $F_{empirisch} = \frac{(R_u^2 - R_R^2)/q}{(1-R_u^2)/(n-k-1)} = \frac{(0{,}382-0{,}375)/1}{(1-0{,}382)/(3000-5-1)} = \frac{0{,}007}{0{,}00020641} = 33{,}91$ [1P]
6. Kritischer Wert: $F_{kritisch} = F_{0{,}05;1;2994} = 3{,}84$ [1P]
7. 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]

e) In welcher Situation spricht man von heteroskedastischen Störtermen? Und welchen Einfluss hat Heteroskedastie auf die Unverzerrtheit des KQ-Schätzers? (2,5 Punkte)

- Heteroskedastie bedeutet, dass die Varianz der Störterme über die Beobachtungen i variiert. [1,5P]
- Heteroskedastie beeinflusst die Unverzerrtheit des KQ-Schätzers nicht. [1P]

Formelsammlung – Praxis der empirischen Wirtschaftsforschung

Kapitel 1:

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_i(y_i-\bar{y})$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})y_i$$

$$=\sum_{i=1}^{n}x_iy_i-n\bar{x}\cdot\bar{y}$$

$$E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i)$$

$$E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)$$

$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y) + 2ab\,Cov(X,Y)$

Für identisch und unabhängig verteilte Zufallsvariablen Y_i:

$$\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_i\right)^2$$

Kapitel 2:

$y_i = \beta_0 + \beta_1 x_i + u_i$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\sum_{i=1}^{n}\hat{u}_i=0 \qquad \sum_{i=1}^{n}x_i\hat{u}_i=0$$

$$SST\equiv\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$SSE\equiv\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$SSR\equiv\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$$

$$R^2=\frac{SSE}{SST}=1-\frac{SSR}{SST}\;,\ 0\le R^2\le 1$$

$$E\left(\hat{\beta}_0\right)=\beta_0 \qquad E\left(\hat{\beta}_1\right)=\beta_1$$

$$Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sigma^2}{SST_x}$$

$$Var(\hat{\beta}_0)=\frac{\sigma^2\frac{1}{n}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{\sigma}^2=\frac{1}{(n-2)}\sum_{i=1}^{n}\hat{u}_i^2=\frac{SSR}{(n-2)}$$

Regression durch den Ursprung:

$$\tilde{\beta}_1=\frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}x_i^2}$$

Kapitel 3:

$$R^2=\frac{SSE}{SST}=\frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2\right)}$$

Wenn $y=\beta_0+\beta_1x_1+\beta_2x_2+u$

und $\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$

dann $\tilde{\beta}_1=\hat{\beta}_1+\hat{\beta}_2\tilde{\delta}_1$ mit $x_2=\tilde{\delta}_0+\tilde{\delta}_1x_1$

Allgemein für j = 1, 2, …, k:

$$Var\left(\hat{\beta}_j\right)=\frac{\sigma^2}{SST_j\left(1-R_j^2\right)}$$

$$SST_j=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-k-1}=\frac{SSR}{n-k-1}$$

$$se(\hat{\beta}_j)=\frac{\hat{\sigma}}{\left[SST_j(1-R_j^2)\right]^{\frac{1}{2}}}$$

MLR.1: Modell der Grundgesamtheit
$y = \beta_0 + \beta_1x_i + \beta_2x_2 + \ldots + \beta_kx_k + u$

MLR.2: Zufallsstichprobe der Größe n folgt dem Bevölkerungsmodell.

MLR.3: Keine unabhängige Variable ist konstant. Keine perfekte Kollinearität.

MLR.4: $E\left(u\,|\,x_1,x_2,\ldots,x_k\right)=0$

MLR.5: $Var\left(u\,|\,x_1,x_2,\ldots,x_k\right)=\sigma^2$

MLR.6: u ist von x_1, x_2, …, x_k unabhängig und u ~ Normal (0, σ^2).

Kapitel 4:

$$(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) \sim t_{n-k-1}$$

$$-t_{\frac{\alpha}{2},n-k-1} \leq \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \leq t_{\frac{\alpha}{2},n-k-1}$$

$$\hat{\beta}_j - c \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c \cdot se(\hat{\beta}_j)$$

$$F \equiv \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)}$$

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

Kapitel 5:

$$\lim_{n\to\infty} \quad P(\left|\hat{\beta}_1 - \beta_1\right| > \varepsilon) \to 0$$

$$\text{plim}\left(\hat{\beta}_1\right) = \beta_1$$

Kapitel 6:

Standardisierung:

$$\frac{y_i - \bar{y}}{\hat{\sigma}_y} = \hat{\beta}_1 \left(\frac{\hat{\sigma}_1}{\hat{\sigma}_y}\right)\left(\frac{x_{i1} - \bar{x}_1}{\hat{\sigma}_1}\right) + \hat{\beta}_2 \left(\frac{\hat{\sigma}_2}{\hat{\sigma}_y}\right)\left(\frac{x_{i2} - \bar{x}_2}{\hat{\sigma}_2}\right)$$
$$+ \ldots + \hat{\beta}_k \left(\frac{\hat{\sigma}_k}{\hat{\sigma}_y}\right)\left(\frac{x_{ik} - \bar{x}_k}{\hat{\sigma}_k}\right) + \frac{\hat{u}_i}{\hat{\sigma}_y}$$

Semielastizität:

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_j \Delta x_j) - 1]$$

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$$

$$\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

$$P[\hat{y}^0 - t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0) \leq y^0 \leq \hat{y}^0 + t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0)] = 1\text{-}\alpha$$

$$\widehat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \ldots + \hat{\beta}_k x_k$$

$$E(y|\mathbf{x}) = \exp(\sigma^2/2) \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Kapitel 7:

Regression nach Gruppen
- Modell gepoolt: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$

Chow-Test (mit $SSR_P = SSR_{\text{gepooltes Modell}}$):

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

TABLE G.2

Critical Values of the *t* Distribution

		Significance Level				
1-Tailed:		**.10**	**.05**	**.025**	**.01**	**.005**
2-Tailed:		**.20**	**.10**	**.05**	**.02**	**.01**
Degrees of Freedom	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical value for a two-tailed test with large (> 120) *df* is 1.96.

Source: This table was generated using the Stata® function invttail.

TABLE G.3b

5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator df (∞) is 2.37.

Source: This table was generated using the Stata® function invFtail.