Aufgabe 1: [26]

Sie möchten die Bestimmungsgründe für den Preis eines Gebrauchtwagens herausfinden und haben zu diesem Zweck Daten aus 95 Zeitungsannoncen t gesammelt: den PREIS (in tausend CHF), das ALTER des Wagens, die Anzahl der PS, die gefahrene Kilometerleistung KM (in tausend Kilometern) und einen Index über das Markenprestige der Automarke (MARKE). Sie schätzen folgendes Modell mit EViews

$$PREIS_t = \beta_1 + \beta_2 ALTER_t + \beta_3 PS_t + \beta_4 KM_t + \beta_5 MARKE_t + \varepsilon_t$$

und erhalten folgende Regressionsergebnisse:

Dependent Variable: PREIS				
Method: Least Squares				
Date: 12/16/03 Time: 10:00				
Sample:	1	95		
Included observations:		95		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C		8.7539	0.5465	0.5861
ALTER	-0.1254	0.0675	-1.8575	0.0665
PS	0.1469		2.1764	0.0321
KM	-0.3460	0.0425		0.0000
MARKE	2.5798	1.5463	1.6683	0.0987
R-squared		Mean dependent var		12.7456
Adjusted R-squared		S.D. dependent var		2.9000
S.E. of regression				
Sum squared resid	412.0765			

a) **Berechnen Sie die fehlenden Werte im Regressions-Output und machen Sie ihren Rechen-weg deutlich.**

Ansatz zur Berechung des Koeffizienten/Standardfehlers/t-Werts unter H_0:

$$\beta_i = 0 \rightarrow t = \frac{b_i}{se(b_i)}$$

i) **b_1** **[1]**

$$b_1 = t \cdot se(b_1) = 0.5465 \cdot 8.7539 = 4.784$$

ii) **den Standardfehler von b_3** **[1]**

$$se(b_3) = \frac{b_3}{t} = \frac{0.1469}{2.1764} = 0.0675$$

iii) **den t-Wert von b_4** **[1]**

$$t(b_4) = \frac{b_4}{se(b_4)} = \frac{-0.3460}{0.0425} = -8.1412$$

iv) **Zeigen Sie, dass die Summe der quadrierten Abweichungen der abhängigen Variablen von ihrem Mittelwert (SST) 790.54 beträgt.** **[1,5]**

Es gilt:

$$\hat{var}(Y) = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T-1} = \frac{SST}{T-1} \qquad (analog\ zu\ 6.1.10b)$$

$$\Rightarrow \qquad sd(Y) = \sqrt{\frac{SST}{T-1}} \Rightarrow SST = sd(Y)^2 \cdot (T-1) = 2.9^2 \cdot 94 = 790.54$$

v) das Bestimmtheitsmaß **[1,5]**

$$R^2 = \left(\frac{SSR}{SST}\right) = 1 - \frac{SSE}{SST} = 1 - \frac{412.0765}{790.54} = 0.4788$$

vi) das angepasste Bestimmheitsmaß **[1,5]**

$$\bar{R}^2 = 1 - \frac{SSE/(T-K)}{SST/(T-1)} = 1 - \frac{412.0765/90}{790.54/94} = 1 - \frac{4.578528}{8.41} = 0.4556$$

vii) den Standardfehler der Regression **[1,5]**

Es gilt: $\hat{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T-K} = \frac{SSE}{T-K} \qquad (HGJ: 7.2.6)$

$$\Rightarrow se(reg) = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{SSE}{T-K}} = \sqrt{\frac{412.0765}{95-5}} = 2.1398$$

b) Interpretieren Sie kurz statistisch (Interpretation des p-Wertes genügt) und inhaltlich die Regressionskoeffizienten b_2 (Alter) und b_3 (PS). **[4]**

- b_2 (Alter des Wagens): Steigt das Alter des Wagens um eine Einheit, so nimmt der Preis eines Gebrauchswagens um 0.1245 Einheiten, also um 124.5 CHF ab; Der Koeffizient ist auf dem 10%-Signifikanzniveau statistisch von Null unterschieden.
- b_3 (PS): Steigt die Anzahl der PS um eine Einheit, so nimmt der Preis eines Gebrauchswagens um 0.1469 Einheiten, also im 146.9 CHF zu; Der Koeffizient ist auf dem 5%-Signifikanzniveau statistisch von Null unterschieden

c) Wie verändert sich der Preis eines Gebrauchtwagens wenn Sie vor dem Kauf noch eine 10 km lange Testfahrt machen? **[1]**

- b_2 (Alter des Wagens): Steigt das Alter des Wagens um eine Einheit, so nimmt der Preis eines Gebrauchtwagens um 0.1245 Einheiten, also um 124.5 CHF ab; Der Koeffizient ist auf dem 10%-Signifikanzniveau statistisch von Null unterschieden.

d) Bilden Sie einen 95%igen Intervallschätzer für β_4 (Kilometerleistung). Wie ist das Intervall zu interpretieren? **[5]**

- $t_c = t_{\alpha/2;df} = t_{0.025;90} = 1.987$; $b_4 \pm t_c \cdot se(b_4)$; $KI: -0.346 \pm 1.987 \cdot 0.0425$
- → $KI: [-0,4304;\ -0,2615]$
- Repeated Sampling: In wiederholten Stichproben enthalten 95% der so konstruierten Intervalle den wahren aber unbekannten Koeffizient.

e) Testen Sie die Aussage, "Preise für Oldtimer sind deswegen so hoch, weil Gebrauchtwagen mit jedem Jahr das sie älter werden um mindestens 25,- Franken teurer werden" auf einem Konfidenzniveau von 90%. **[4]**

- $H_0: \beta_2 \geq 0.025$; $H_1: \beta_2 < 0.025$

- $t_{b_2} = \dfrac{-0.1254 - 0.025}{0.0675} = -2.22815 < -1.291 \quad \left(t_{\alpha=0.1;df=90}\right)$
- → H_0 ablehnen.

f) Wieviel kostet ein Audi 80 des Baujahres 1992 mit 75 PS, 125.000 km auf dem Tacho? (Audi hat einen Markenprestigewert von 15). **[3]**

(Klausur geschrieben in 2003: Alter=11)

- 1*4.7843+11*-0.1254+75*0.1469+125*-0.346+15*2.5798= 9.8694
- Der Wagen kostet 9869.4 CHF.

Aufgabe 2: [14]

a) Erläutern Sie Inhalt und Bedeutung der Aussage(n) des Gauss-Markov Theorems. **[4]**

- Inhalt: Unter allen linearen, unverzerrten Schätzern und unter den Annahmen SR 1-5 haben die KQ Schätzer die kleinste Varianz.
- Bedeutung: Vergleich alternativer Verfahren, KQ gewinnt bei Varianz. Aussage gilt nur für Klasse der linearen, unverzerrten Schätzer.

b) Nennen Sie vier Faktoren, die die Präzision der Schätzung von Steigungsparametern im Rahmen multivariater Regressionen bestimmen und erläutern Sie die dahinterstehenden Zusammenhänge für zwei dieser Faktoren. **[6]**

- Im multiplen Regressionsmodell mit K=3:

$$var\left(b_2\right) = \frac{\sigma^2}{\sum_{t=1}^{T}(x_{t2} - \overline{x}_2)^2 \cdot (1 - r_{23}^2)} \quad \text{(HGJ, S. 154/5)}$$

- σ^2 → hohe Streuung, weniger Information zum genauen Verlauf
- T → mehr Beobachtungen, mehr Information und Präzision
- Variation um Mittelwert → breitere Informationsbasis günstig
- r_{23}^2 → je höher, desto schwieriger ist es, Effekte zu trennen, weniger Präzision

c) Nehmen Sie Stellung zu der Aussage „Ein Intervallschätzer ist informativer als ein Punktschätzer“. **[4]**

- Die Aussage ist korrekt, da Information über die Präzision der Schätzung enthalten ist; Information über den Standardfehler geht mit ein.

Aufgabe 3: [5]

Nehmen Sie Stellung zu der folgenden Aussage „Wenn es einen Zusammenhang zwischen einer Variable Y und einer Variable X gibt, lässt sich dieser korrekt mittels eines Kleinstquadrateschätzers bestimmen und die gewählte funktionale Form ist bedeutungslos“

- Die Aussage ist falsch; KQ ist unverzerrt, wenn lineare Form (in Parametern) gegeben.
- Falls funktionale Form falsch:
  - R^2 nicht optimal
  - systematische Effekte in e

− Zusammenhang nicht optimal repräsentiert

Aufgabe 4: (Wahr oder falsch / Wahr-falsch-weil) [14]

Wahr oder Falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr), für nicht zutreffende f (für falsche) ein. (Für jede richtige Antwort gibt es 1 Punkt, für jede falsche Antwort wird 1 Punkt abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.) Bei mit * gekennzeichneten Fragen gibt es einen Zusatzpunkt für eine kurze Begründung der Antwort.

F	Der geschätzte Steigungsparameter im einfachen linearen Modell ist unabhängig von der gewählten Skalierung der erklärenden Variable
W	Die Annahme, dass der Störterm im multivariaten Regressionsmodell normalverteilt ist, ist für Erwartungstreue und Varianz der Kleinstquadrateschätzer unerheblich.
F	Im einfachen linearen Regressionsmodell steigt die Kovarianz zwischen zwei Parameterschätzern β_1 und β_2 wenn die Streuung der abhängigen Variable fällt.
F	Wenn wir eine Nullhypothese nicht verwerfen, ist sie mit Wahrscheinlichkeit $(1-\alpha)\times 100$ Prozent wahr.
W	Wenn eine Variable Y in logarithmierter Form auf eine Variable X in ebenfalls logarithmierter Form regressiert wird, so ist der geschätzte Steigungsparameter als Elastizität interpretierbar.
F	Der Jarque-Bera Test nutzt Informationen über Standardfehler und Schiefe der Verteilung einer Zufallsvariable, um zu prüfen, ob sie normalverteilt ist.
F	Bei der Interpretation einzelner Steigungsparameter im Rahmen multivariater Regressionen wird unterstellt, dass sich alle erklärenden Variablen gleichzeitig um den gleichen relativen Betrag ändern.
F	Der Typ I Fehler beschreibt die Wahrscheinlichkeit mit der die Alternativhypothese verworfen wird obgleich sie zutreffend ist.
W	Bei steigender Stichprobengröße sinkt die Wahrscheinlichkeit eines Typ II Fehlers
F	Statistische Signifikanztests werden unter Festlegung der Typ II Fehlerwahrscheinlichkeit durchgeführt.
(*) **F**	Die Varianz der Summe zweier unkorrelierter Zufallsvariablen ist immer kleiner als die Varianz der Summe zweier korrelierter Zufallsvariablen. → $Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)$; Kovarianz kann sowohl positiv als auch negativ sein
(*) **F**	Die Korrelation ρ gibt ein Maß für die statistische Unabhängigkeit zweier Zufallsvariablen. → Korrelation ist Maß für linearen Zusammenhang. Bei Unabhängigkeit ist Korr=0, aber nicht zwingend umgekehrt: Korr=0 bedeutet nicht automatisch Unabhängigkeit