Aufgabe 1: [21]

Mit den Daten von 177 Mietwohnungen einer Schweizer Stadt wurde versucht, die Determinanten des Mietzinses empirisch zu ermitteln, indem eine Kleinstquadratschätzung des Regressionsmodells

$$MIETE = \beta_1 + \beta_2 NWF + \beta_3 ALTER + \beta_4 DZENT + \varepsilon$$

durchgeführt wurde. (Dabei bedeuten die Variablen MIETE: Mietpreis 1999 in CHF, NWF: Nettowohnfläche in qm, ALTER: Alter der Wohnung in Jahren, DZENT: Distanz zum Stadtzentrum in km)

EViews liefert folgende Regressionsergebnisse:

Dependent Variable: MIETE
Method: Least Squares
Date: 12/11/02 Time: 08:00
Sample: 1 177
Included observations: 177

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C		144.7024	6.335762	0.0000
NWF	11.40628		7.947048	0.0000
ALTER	-6.469049	0.937492		0.0000
DZENT		20.54634	-3.433036	
R-squared		Mean dependent var	1303.944	
Adjusted R-squared		S.D. dependent var	526.3415	
S.E. of regression				
Sum squared resid	26928912			

a) Berechnen Sie die fehlenden Werte im Regressions-Output. **[6]**

Ansatz zur Berechung des Koeffizienten/Standardfehlers/t-Werts unter H_0: $\beta_i = 0 \rightarrow t = \dfrac{b_i}{se(b_i)}$

i) b_1

$$b_1 = t \cdot se(b_1) = 6.335762 \cdot 144.7024 = 916.8$$

ii) den Standardfehler von b_2

$$se(b_2) = \frac{b_2}{t} = \frac{11.40628}{7.947048} = 1.435285$$

iii) den t-Wert von b_3

$$t(b_3) = \frac{b_3}{se(b_3)} = \frac{-6.469049}{0.937492} = -6.900378$$

iv) b_4

$$b_4 = t \cdot se(b_4) = -3.433036 \cdot 20.54634 = -70.53632$$

v) das Bestimmtheitsmaß und das angepasste Bestimmheitsmaß (Die Summe der quadrierten Abweichungen der abhängigen Variablen MIETE von ihrem Mittelwert beträgt: 48.758.231,3776).

$$R^2 = \left(\frac{SSR}{SST}\right) = 1 - \frac{SSE}{SST} = 1 - \frac{26928912}{48758231.3776} = 0.4477$$

$$\bar{R}^2 = 1 - \frac{SSE/(T-K)}{SST/(T-1)} = 1 - \frac{26928912/(177-4)}{48758231.3776/(177-1)} = 0.4381$$

vi) den Standardfehler der Regression

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

$$\Rightarrow se(reg) = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{SSE}{T-K}} = \sqrt{\frac{26928912}{177-4}} = 394.5357$$

b) Interpretieren Sie kurz statistisch und inhaltlich die Regressionskoeffizienten b_1 und b_2. [4]

- b_1 (statistisch): ist hoch signifikant von Null verschieden (t-Wert >2, p-Wert = 0)
- b_1 (inhaltlich): ist die Mindestmiete, die zu zahlen ist, wenn alle erklärenden Variablen 0 sind; ist inhaltlich schwer zu interpretieren;[1] Achsenabschnitt der Regressionsgeraden; braucht man zu besseren Anpassung der Regression.
- b_2 (statistisch): ist hoch signifikant von 0 verschieden, (t-Wert, p-Wert); hat geringen Standardfehler und ist somit präzise geschätzt
- b_2 (inhaltlich): Mietpreis je Quadratmeter Nettowohnfläche = 11.40 CHF in 1999; *alternativ*: Jeder zusätzliche Quadratmeter Nettowohnfläche erhöht die Miete um 11.40 CHF.

c) Bilden Sie einen 90%igen Intervallschätzer für das Alter der Wohnung. Was geben die Intervallgrenzen an? Wie ist das Intervall zu interpretieren? [6]

- 90%-Konfidenzintervall: $t_c = t_{\alpha/2;df} = t_{0.05;173} = 1.645$ $\quad KI: b_3 \pm t_c \cdot se(b_3)$

  hier: $b_3 \pm t_c \cdot se(b_3)$ = -6.469049 $\pm$ 1.645 $\cdot$ 0.937492 = [-8.011223; -4.926875]
- In wiederholten Stichproben liegt der wahre β_3-Koeffizient zwischen diesen beiden Intervallgrenzen.

d) Testen Sie, ob die Distanz zum Stadtzentrum bei einem 99%igen Konfidenzniveau einen signifikanten Einfluss auf den Mietzins hat. [2]

- Zweiseitiger t-Test
- t-Test für α = 1%, kritischer t-Wert (t_c)= $\pm$ 2.576
- H_0: $\beta_4 = 0$; H_1: $\beta_4 \neq 0$
- Da empirischer t-Wert = |-3.433| > 2.576 $\rightarrow$ H_0 verwerfen.
- Die Distanz zum Stadtzentrum hat bei einem 99%-igen Konfidenzniveau einen signifikanten Einfluss auf den Mietzins.

e) Testen Sie die Aussage des Mieterschutzbundes „ein Quadratmeter Wohnfläche kostet mindestens 10 Franken" auf einem 2,5%-Signifikanzniveau. [3]

- Einseitiger t-Test
- Die Aussage des Mieterschutzbundes kann auf 2,5% Signifikanzniveau verworfen werden.
- H_0: $\beta_2 = 0$; H_1: $\beta_2 > 10$ oder: H_0: $\beta_2 \leq 10$; H_1: $\beta_2 > 10$
- Empirischer t-Wert = (11.40628-10)/1.435285 = 0.9798

[1] Hier etwa: Hat eine Wohnung eine Nettowohnfläche von 0 qm (!), ist 0 Jahre alt und liegt im Stadtzentrum (Distanz = 0 km), so beträgt die Miete 916.80 CHF. → *Gutes Bsp., dass Interpretation von Konstanten meist wenig sinnvoll ist*.

– Kritischer t-Wert (t_c) = 1.960
– Da empirischer t-Wert < t_c $\rightarrow$ H_0 kann nicht verworfen werden.

Aufgabe 2: [6]

Sie sind daran interessiert, einen präzisen Schätzer für den linearen Zusammenhang zwischen Alter als erklärende Variable und einem Index für die Gesundheit einer Person als abhängiger Variable zu erhalten.

a) **Wenn Sie die Wahl haben, ihre Stichprobe im Seniorenwohnheim oder aus der Gesamtbevölkerung zu ziehen, haben Sie eine Präferenz? Warum? [4]**

Formel für die Varianz des Schätzers β_2: $\operatorname{var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}$

– Zusammenhang verbal erläutern:
  - Präziser Schätzer = geringer Standardfehler, geringe Varianz, hoher t-Wert;
  - Bei gegebener Varianz der abhängigen Variable ist Varianz von β_2 kleiner, wenn die Summe der quadrierten Abweichungen von X groß ist;
  - Hohe Varianz in den X ist vorteilhaft;
  - Gesamtbevölkerung wird bevorzugt, da dort alle Altersklassen vertreten sind, im Altersheim hingegen nur die obere Altersgruppe.

b) Wenn Sie mit einer Stichprobe von 100 oder von 800 Personen arbeiten können, was ziehen Sie vor und warum? [2]

– Größere Stichprobe ist für die Varianz des Schätzers vorteilhafter, da
  - $\operatorname{var}(b_1) = \sigma^2 \left[\frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} \right]$, d.h. mit steigendem T wird der Nenner größer und damit die Varianz kleiner ;
  - $\operatorname{var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}$, d.h. der Nenner steigt mit größerer Stichprobe (bei konstantem σ^2), somit wird die Varianz von β_2 kleiner.

Aufgabe 3: [5]

„Um Konfidenzintervalle für Schätzer des Steigungsparameters bestimmen zu können, sind Annahmen hinsichtlich der Verteilung des Fehlerterms nicht erforderlich." Nehmen Sie zu dieser Aussage Stellung und begründen Sie ihre Position.

– Die Aussage ist falsch;
– Die Annahme der Normalverteilung der Fehlerterme muss erfüllt sein: e ~ N(0, σ^2);
– ansonsten ist Konfidenzintervall verzerrt,
– da $b_2/se(b_2)$ nicht mehr t-verteilt ist, aber zur Berechnung des KI die t-Verteilung unterstellt wird (Einsetzen von t_c in KI-Formel: $b_2 \pm (t_c \cdot SE(b_2))$.

Aufgabe 4: (Wahr oder falsch) [7]

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 0,5 Punkte, für jede falsche Antwort werden 0,5 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.)

W	Zur Bestimmung des Kleinstquadrateschätzers ist die Annahme normalverteilter Fehlerterme nicht erforderlich.
W	Die t-Verteilung konvergiert bei steigenden Freiheitsgraden gegen die Normalverteilung.
W	Die Chi-Quadratverteilung konvergiert bei steigenden Freiheitsgraden gegen die Normalverteilung.
F	Die Summe von *m* standardnormal verteilten Variablen ist Chi-Quadrat verteilt mit *m* Freiheitsgraden.
F	Der t-Test wird unter der Annahme formuliert, dass die Nullhypothese falsch ist.
F	Die Wahrscheinlichkeit für einen Typ I Fehler ist umso höher, je kleiner das Signifikanzniveau ist.
W	Die Wahrscheinlichkeit für einen Typ II Fehler ist umso höher, je kleiner das Signifikanzniveau ist.
W	Je größer die Stichprobe, umso eher wird ein Typ II Fehler vermieden.
F	Je höher der p-Wert, umso höher die Wahrscheinlichkeit H_0 zu verwerfen.
W	Schätzungen mit großen Stichproben erlauben präzisere Vorhersagen, als Schätzungen mit kleinen Stichproben.
F	Stichproben mit großer Streuung der abhängigen Variable erlauben präzisere Vorhersagen, als Stichproben mit kleiner Streuung der abhängigen Variable.
W	Das R^2 steigt oder bleibt gleich, wenn zusätzliche erklärende Variablen im Modell berücksichtigt werden.
W	Für ein gegebenes R^2 ist das korrigierte R^2 umso höher, je größer die Differenz T-K (T=Anzahl der Beobachtungen, K=Anzahl der geschätzten Parameter).
F	Je größer die Stichprobe, umso erwartungstreuer sind KQ Schätzer.

Aufgabe 5: (Wahr oder falsch) [5]

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.)

W	Im multiplen Regressionsmodell steigt die Varianz eines geschätzten Steigungsparameters β_k, wenn die entsprechende erklärende Variable x_k stark mit anderen erklärenden Variablen im Modell korreliert ist.
F	Gegeben seien zwei Zufallsvariablen x, y und zwei Konstanten a, b. Die Varianz von Z = ax + by ist immer höher, wenn x und y nicht korreliert sind, als wenn x und y korreliert sind.
W	Aus einem (1-α) prozentigen Konfidenzintervall für den Parameter β können wir lernen, ob der geschätzte Parameter mit einer Typ I Fehlerwahrscheinlichkeit von α Prozent statistisch signifikant ist.

F	Eine hohe positive Kovarianz der im einfachen linearen Modell geschätzten Parameter β_1 und β_2 erhöht die Präzision der auf Basis dieser Schätzung gemachten Vorhersage.
W	Für eine gegebene Nullhypothese H_0: $\beta = c$ ist der p-Wert eines einseitigen Tests größer als der p-Wert eines zweiseitigen Tests.

Aufgabe 6: [6]

Ihnen wird mitgeteilt, dass eine Zufallsvariable, von der Ihnen 65 Beobachtungen vorliegen, folgende Merkmale hat:

Mittelwert = 0, Varianz = 4, Schiefe (Skewness) = 1, Wölbung (Kurtosis) = 5. Testen Sie am 1% Signifikanzniveau ob die Variable normalverteilt ist und erläutern Sie ihre Vorgehensweise genau. (Hinweis: Der p-Wert ihrer Teststatistik beträgt 0,0000197.)

– Jarque-Bera Test auf Normalverteilung:

– JB = $\frac{T}{6}\left(S^2 + \frac{(k-3)^2}{4}\right)$

– Mit S=1, T=65 und k=5 folgt

– $JB = \frac{65}{6}\left(1^2 + \frac{(5-3)^2}{4}\right) = \frac{65}{6}(1+1) = 21.67$

– kritischer Wert: $\chi^2_{0.99;2} = 9.21$

– $JB > \chi^2 \rightarrow$ H_0 (Fehler normalverteilt) verwerfen

– p = 0 < 0.01 $\rightarrow$ H_0 (Fehler normalverteilt) verwerfen

Aufgabe 7: [10]

Aus Versehen haben Sie die Daten für die Schätzung eines einfachen linearen Regressionsmodells doppelt eingelesen. Zeigen Sie algebraisch, welche Konsequenzen dies für den Schätzer des Steigungsparameters und für den Achsenabschnitt hat.

– $$b_2 = \frac{T\sum x_t y_t - \sum x_t \sum y_t}{T\sum x_t^2 - \left(\sum x_t\right)^2} = \frac{2\cdot T\cdot 2\cdot\left(\sum x_t y_t\right) - 2\cdot\sum x_t \cdot 2\sum y_t}{2\cdot T\cdot 2\cdot\sum x_t^2 - \left(2\cdot\sum x_t\right)^2}$$
$$= b_2$$

– $$\left.\begin{aligned} b_1 &= \overline{y} - b_2\cdot\overline{x} \\ &= b_1 \end{aligned}\right\} \text{Änderung nur, wenn sich } b_2 \text{ ändert.}$$

– → Es ändert sich nichts.