Wirtschaftsstatistik und Ökonometrie

Übungsblatt 13

Abgabe: Freitag, 07.02.2025 12:00 Uhr in Moodle

R-Aufgabe 1 (3 + 3 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 14 Punkte): Betrachten Sie den Datensatz `homeprice` (`library("UsingR")`), welcher die Variablen

$list$	Listenpreis (in Tausend \$)
$sale$	Verkaufspreis (in Tausend \$)
$full$	Anzahl von komplett ausgestatteten Badezimmern
$half$	Anzahl von kleinen Badezimmern
$bedrooms$	Anzahl von Schlafzimmern
$rooms$	Anzahl von Zimmern insgesamt
$neighborhood$	Subjektive Beurteilung der Wohngegend (Skala von 1 bis 5)

für 29 im Jahr 2001 verkaufte Häuser in Maplewood, NJ, enthält.

(a) Testen Sie zu einem Signifikanzniveau von $\alpha = 5\%$ für das lineare Regressionsmodell

$$sale = \beta_0 + \beta_1 full + \beta_2 rooms + \beta_3 neighborhood + u$$

unter den Annahmen MLR.1-6 die Nullhypothese $\hat{\beta}_3 = 0$ gegen eine zweiseitige Alternative. Finden Sie das dazugehörige Konfidenzintervall für $\hat{\beta}_3$ anhand der passenden t-Statistik und zeichnen Sie deren Verteilung (unter H_0) in ein Diagramm. Kennzeichnen Sie dabei innerhalb der Abbildung den Ablehnbereich sowie den Wert der Teststatistik.

(b) Unter den Annahmen MLRA.1-4 und MLRA.5' gilt, dass die Teststatistik asymptotisch normalverteilt ist. Berechnen Sie das entsprechende asymptotische Konfidenzintervall für $\hat{\beta}_3$ und fügen Sie die Dichte der entsprechenden Normalverteilung in obiges Diagramm ein. Kennzeichnen Sie erneut den entsprechenden Ablehnbereich auf Grundlage dieser asymptotischen Normalverteilung.

(c) Betrachten Sie die Ergebnisse aus Teil (a) und (b) und vergleichen Sie diese.

(d) Testen Sie zu einem Signifikanzniveau $\alpha = 5\%$ für das (unrestringierte) Modell

$$sale = \beta_0 + \beta_1 full + \beta_2 half + \beta_3 bedrooms + \beta_4 rooms + \beta_5 neighborhood + u,$$

ob die Variablen $half$ und $bedrooms$ gemeinsam einen Einfluss auf den Verkaufspreis haben. Gehen Sie dabei davon aus, dass die Annahmen MLR.1-MLR.6 erfüllt sind und verwenden Sie eine geeignete F-Statistik.

(e) Testen Sie die Hypothese aus Teil (d) mit dem Lagrange Multiplier Test.

(f) Unterscheiden sich die p-Werte für die Tests aus (d) und (e)?

Betrachten Sie nun die folgenden beiden Modelle:

$$sale = \beta_0 + \beta_1 full + \beta_2 rooms + \beta_3 neighborhood + u,$$
$$sale = \beta_0 + \beta_1 half + \beta_2 bedrooms + \beta_3 neighborhood + u.$$

(g) Welche Korrelation würde man für $full$ und $half$ bzw. $rooms$ und $bedrooms$ erwarten? Wie hoch sind die tatsächlichen Korrelationen?

(h) Sind die Regressoren der Modelle signifikant bzw. gemeinsam signifikant zu einem Niveau von 5%?

(i) Welches Modell würde bevorzugt werden?

(j) Würde $\bar{R}^2$ fallen, falls die Variable *neighborhood* aus den Regressionen entfernt wird? Begründen Sie Ihre Antwort.

R-Aufgabe 2 (2 + 3 + 1 = 6 Punkte): Betrachten Sie erneut den Datensatz `ergebnisse.txt` von Blatt 12 zusammen mit der linearen Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log x_2 + \beta_4 x_3 + \beta_5 x_1 x_3 + \beta_6 x_4 + u.$$

Das Signifikanzniveau sei durchweg auf $\alpha = 5\%$ festgesetzt.

(a) Welche Regressoren sind unter der Annahme der Homoskedastizität signifikant?

(b) Welche Regressoren sind unter der Annahme der Heteroskedastizität signifikant? Berechnen Sie dazu die heteroskedastizitätsrobusten Standardabweichungen, t-Statistiken sowie die zugehörigen p-Werte. Unterscheiden sich die Werte stark von denen aus Teil (a)?

(c) Führen Sie einen White-Test durch, um zu überprüfen, ob in diesem Modell Heteroskedastizität vorliegt. Dabei soll eine Regression von $\hat{u}_i^2$ auf $\hat{y}_i$ und $\hat{y}_i^2$ durchgeführt werden.

R-Aufgabe 3 (1 + 1 + 3 = 5 Punkte): Die Datei `schule.txt` enthält folgende Daten über 1.692 Schulen in Michigan:

- y Anteil der Viertklässler (in %), die eine ausreichende Note in Mathe erreicht haben (*Bestehen*)
- x_1 Anteil der Schüler (in %), die ein kostenloses Mittagessen erhalten haben (*Lunch*)
- x_2 Anzahl der Schüler (*Schueler*)
- x_3 Ausgaben (in \$) der Schule pro Schüler (*Ausgaben*)

Es wird folgender linearer Zusammenhang vermutet:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 \log x_3 + u$$

(a) Schätzen Sie die Regressionsparameter in **R** mit Hilfe von `lm` und interpretieren Sie β_1, β_2 und β_4 in Worten. Wie sollte man das lineare Modell ändern, um β_1 und β_2 besser interpretieren zu können.

(b) Testen Sie mithilfe des Breusch-Pagan-Tests zum Niveau $\alpha = 0.01$, ob in diesem linearen Modell Heteroskedastizität vorliegen könnte. (Es soll keine **R**-Funktion verwendet werden, die den Test automatisch durchführt, `lm` darf aber verwendet werden.)

(c) Schätzen Sie die normale Kovarianzmatrix von $\hat{\beta}$ und die White-Kovarianzmatrix von $\hat{\beta}$. Vergleichen Sie die Standardfehler von $\hat{\beta}_0, \dots, \hat{\beta}_4$.

Hinweis: Die Lösungen zu den R-Aufgaben müssen ausreichend erklärt werden. Dazu sollte in der .R-Datei, die auf Moodle als Lösung hochgeladen wird, der Code ausreichend kommentiert werden!