Universität Duisburg-Essen
Lehrstuhl für Ökonometrie
Dr. Yannick Hoga
MSc. Martin Arnold

Wintersemester 2022/23

Übungsblatt 6 — Inferenz in Regressionsmodellen I

Zu diesem Übungsblatt empfehlen wir neben der Lektüre von Kapitel 5 des Lehrbuches *Introduction to Econometrics* von *Stock & Watson* eine Aufarbeitung mithilfe der Kapitel 5 in unserem Online-Companion *Introduction to Econometrics with R*.

Aufgabe 1 — Anwendungsbeispiel: Boston Housing Data

Das Paket `MASS` enthält den Datensatz `Boston`, welcher wiederum die Variable `medv` (median house value) für 506 Gemeinden rund um Boston enthält. Im folgenden wollen wir ein lineares Modell aufstellen mit `medv` als abhängiger Variable und bis zu 13 erklärenden Variablen. Diese sind z. B. `rm` (average number of rooms per house), `age` (average age of houses), und `lstat` (percent of households with low socioeconomic status).

(a) Verschaffen Sie sich einen Überblick über den Datensatz.

(b) Passen Sie ein einfaches lineares Modell an mit `medv` als abhängiger und `lstat` als erklärender Variable. Geben Sie die p-Werte und die Standardfehler der geschätzten Koeffizienten an. Berechnen Sie auch das R^2 und die F-Statistik für das Modell.

(c) Berechnen Sie ein 90%-Konfidenzintervall für die geschätzten Koeffizienten.

(d) Schätzen Sie die Werte von `medv` für den Fall, dass die Variable `lstat` die Werte 5, 10 und 15 annimmt.

(e) Plotten Sie `medv` und `lstat` zusammen mit der geschätzten Regressionsgerade.

(f) Wiederholen Sie die Aufgaben (b) und (c) unter Verwendung heteroskedastie-robuster Standardfehler. Ausführliche Erläuterungen zur Berechnung heteroskedastie-robuster Standardfehler in Regressionsmodellen mit Anwendungsbeispielen in `R` finden Sie in Kapitel 5.4 im Online-Companion.

(g) Passen Sie ein Regressionsmodell an, welches `lstat` und `age` als erklärende Variablen für `medv` enthält.

(h) Passen Sie ein Regressionsmodell an, welches `medv` auf alle 13 Variablen regressiert.

(i) Passen Sie ein Regressionsmodell an, welches `medv` auf alle Variablen im Datensatz außer `age` regressiert.

(j) Wiederholen Sie die Aufgabenteile (g) `-` (i) unter Verwendung bei Heteroskedastie gültiger Standardfehler.

Aufgabe 2 — Bew. zur Stichprobenverteilung von $\widehat{\beta}_1$

Diese Aufgaben dient dazu die behauptete erste Gleichung auf Folie 3-53 nachzuweisen. Betrachten Sie das Regressionsmodell

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n.$$

Zeigen Sie, dass für den KQ-Schätzer $\widehat{\beta}_1$ gilt:

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right) u_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.$$

Aufgabe 3 — Anwendungsbeispiel: Lohnregression II

Aus Daten einer Stundenlohnerhebung (in Euro) von 250 zufällig ausgewählten Männern und 280 nicht-männlichen Personen, errechnete ein Wissenschaftler folgendes Regressionsergebnis:

$$\widehat{Lohn} = \underset{(0.23)}{12.52} + \underset{(0.36)}{2.12} \times Mann, \quad R^2 = 0.06, \quad SER = 4.2,$$

wobei $Mann$ eine binäre Variable ist mit $Mann = 1$, falls die Person männlich ist, und $Mann = 0$, falls nicht. Definiere die Lohngeschlechterdifferenz als die Differenz des durchschnittlichen Stundenlohns zwischen Männern und anderen Geschlechtern.

(a) Wie hoch ist die geschätzte Lohngeschlechterdifferenz?

(b) Ist die Lohngeschlechterdifferenz signifikant von 0 verschieden auf einem Niveau von 1%? Benutzen Sie für Ihre Antwort den p-Wert.

(c) Konstruieren Sie ein 95%-Konfidenzintervall für die Lohngeschlechterdifferenz.

(d) Was ist der Durchschnittslohn der Geschlechtergruppen im Datensatz?

(e) Ein anderer Wissenschaftler benutzt dieselben Daten, aber er regressiert $Lohn$ auf $NichtMann$, eine binäre Variable die 1 ist, falls die Person kein Mann ist und 0 für einen Mann. Bestimmen Sie die Regressionsergebnisse dieser Regression:

$$\widehat{Lohn} = \underline{\hspace{2em}} + \underline{\hspace{2em}} \times NichtMann, \quad R^2 = \underline{\hspace{2em}}, \quad SER = \underline{\hspace{2em}}.$$