

Übung zur Vorlesung „Empirische Ökonomie 1“

Übungsblatt 2: Lineares Regressionsmodell II

Aufgabe 1:

Sie haben mit einer großen Stichprobe eine OLS-Regression durchgeführt. Der geschätzte Parameter $\hat{\beta}_1$ beträgt 0,751. Die t -Statistik für einen Test der Nullhypothese $H_0 : \beta_1 = 0,5$ beträgt 0,275. Gehen Sie davon aus, dass die Annahmen A1-A3 des linearen Regressionsmodells für das geschätzte Modell erfüllt sind.

- Wird die Nullhypothese auf dem 5%-Signifikanzniveau abgelehnt?
- Wie groß ist die Varianz des geschätzten Parameters $\hat{\beta}_1$?
- Betrachten Sie die Nullhypothese $H_0 : \beta_1 = 3$. In welchem Bereich liegt der p -Wert, d.h. das Signifikanzniveau, zu dem H_0 gerade noch abgelehnt wird?

Aufgabe 2:

Öffnen Sie den Beispiel-Datensatz von Ramanathan data2-1 „SAT scores“ in GRETL. Die Beobachtungseinheiten sind eine Zufallsstichprobe von 427 Schülern. Die Variablen sind SAT scores, also standardisierte Testergebnisse, aufgeteilt nach Themengebieten:

`vsat` Verbaler Test
`msat` Mathematischer Test

Betrachten Sie ein Regressionsmodell der Form

$$\text{vsat}_i = \beta_0 + \beta_1 \cdot \text{msat}_i + u_i,$$

mit dem Index $i = 1, \dots, 427$. Gehen Sie davon aus, dass die Annahmen A1-A3 des linearen Regressionsmodells erfüllt sind.

- Testen Sie die Nullhypothese $H_0 : \beta_1 = 0$ bei einem Signifikanzniveau von $\alpha = 10\%$. Was ist das Ergebnis?
- Testen Sie die Nullhypothese $H_0 : \beta_0 = 300$ bei einem Signifikanzniveau von $\alpha = 1\%$. Was ist das Ergebnis?
- Betrachten Sie in Ihrem GRETL-Output die Gütemaße R^2 und SER . Wie berechnen sich diese und wie lassen Sie sich interpretieren?

Aufgabe 3:

In dieser Aufgabe beschäftigen wir mit dem Zusammenhang zwischen dem Alter einer Person und deren Lohn. Dazu verwenden wir Umfragedaten aus dem *Current Population Survey*. Der Name des GRETTL-Datensatzes ist `cpsage.gdt`. Sie können ihn auf Ihrer Festplatte speichern und dann in GRETTL unter Datei → Öffne Daten → Benutzerdatei. Er enthält Angaben zu einer Zufallsstichprobe von 7711 jungen US-amerikanischen Personen im Alter zwischen 25 und 34 Jahren, die im Jahr 2008 vollzeit beschäftigt waren. Den Datensatz finden Sie auf unserer Veranstaltungsseite.

Die vorhandenen Variablen sind:

`age` Alter [Jahre]
`ahe` durchschnittlicher Stundenlohn [\$/Stunde]

- Erzeugen Sie ein Histogramm der Variable „ahe“. Beschreiben Sie die Verteilung der Variable und beurteilen Sie ob die Annahme A3 linearen Regressionsmodells für diese Variable erfüllt ist.
- Erzeugen Sie nun ein Streudiagramm mit der Variable „age“ auf der X-Achse und der Variable „ahe“ auf der Y-Achse. Beschreiben Sie welchen Zusammenhang Sie zwischen diesen beiden Variablen erkennen und begründen Sie warum dieser beobachtet wird.
- Betrachten Sie ein Regressionsmodell der Form

$$\text{ahe}_i = \beta_0 + \beta_1 \text{age}_i + u_i, \quad i = 1, \dots, n$$

und gehen Sie davon aus, dass die Annahmen A1-A3 des linearen Regressionsmodells erfüllt sind. Um wie viele Dollar pro Stunde steigt der vorhergesagte Lohn einer Person, wenn sie ein Jahr älter wird?

- Wie lautet die Interpretation von $\hat{\beta}_0$ in dem in Teilaufgabe (c) geschätzten Modell?
- Testen Sie mit Hilfe Ihres Schätzergebnisses aus (c) die Nullhypothese: „Wenn eine Person ein Jahr älter wird, so erhöht sich der Stundenlohn um einen Dollar“ zum 1% Signifikanzniveau. Was ist das Ergebnis?
- Lassen Sie sich in GRETTL das 99% Konfidenzintervall von β_1 anzeigen und interpretieren Sie dieses.