

Übung zur Vorlesung „Empirische Ökonomie 1“

Übungsblatt 3: Lineares Regressionsmodell III

Aufgabe 1:

Betrachten Sie ein Regressionsmodell der Form

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

wobei

$Y_i =$ Stundenlohn von Person i [EUR]

$D_i =$ „Dummyvariable“ für das Geschlecht von Person i :

$D_i = 0$ wenn i weiblich ist,

$D_i = 1$ wenn i männlich ist

Sie haben das obige Modell für eine große Stichprobe geschätzt und die OLS-Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$ erhalten.

- Wie lautet die allgemeine Interpretation der geschätzten Koeffizienten in diesem Modell?
- Stellen Sie die Nullhypothese auf, die benötigt wird um zu testen, ob der Lohn von Frauen und Männern unterschiedlich ist.
- Beschreiben Sie was man unter Homoskedastie und Heteroskedastie versteht. Welche Bedingung muss in diesem Beispiel für Homoskedastie gelten?

Nehmen Sie für die folgenden Teilaufgaben an, dass Sie über eine Zufallsstichprobe von 2000 Individuen verfügen, von denen 600 Frauen und 1400 Männer sind. Sie haben folgende Informationen über diese Individuen: Der Durchschnittsstundenlohn der Frauen in der Stichprobe beträgt $\bar{Y}_F = 12,50$ EUR, der der Männer $\bar{Y}_M = 15,00$ EUR. Die Stichprobenstandardabweichung des Durchschnittsstundenlohnes ist $s_{Y_F} = 6$ bei den Frauen und $s_{Y_M} = 9,6$ bei den Männern.

- Wie hoch ist der Mittelwert von D in der Stichprobe?
- Wie hoch ist der mit OLS geschätzte Parameter $\hat{\beta}_0$?
- Wie hoch ist der mit OLS geschätzte Parameter $\hat{\beta}_1$?

Aufgabe 2:

In dieser Aufgabe beschäftigen wir uns mit der Schätzung von sogenannten Bildungsrenditen: Um wieviel steigt das spätere Einkommen z.B. durch ein Studium? Dazu verwenden wir Umfragedaten aus dem *U.S. National Longitudinal Survey of Youth*. Der Name des GRETLL-Datensatzes ist `NLSY.gdt`. Er enthält Angaben zu einer Zufallsstichprobe von 2061 jungen US-amerikanischen Männern im Alter zwischen 24 und 34 Jahren und steht Ihnen auf unserer Veranstaltungsseite zur Verfügung.

Die vorhandenen Variablen sind:

<code>age</code>	Alter [Jahre]
<code>wage</code>	Gehalt [\$/Monat]
<code>college</code>	Dummyvariable für College-Abschluss [0=nein, 1=ja]
<code>iq</code>	Intelligenzquotient (IQ)

Betrachten Sie ein Regressionsmodell der Form

$$\text{wage}_i = \beta_0 + \beta_1 \text{college}_i + u_i, \quad i = 1, \dots, n \quad (1)$$

- Schätzen Sie das Modell mit der Methode der kleinsten Quadrate. Verwenden Sie „robuste Standardfehler“. Wie hoch ist in der Stichprobe das durchschnittliche Einkommen (in \$/Monat) von Männern mit College-Abschluss?
- Wie lautet der geschätzte Standardfehler von $\hat{\beta}_1$?
- Wiederholen Sie die Regression ohne „robuste Standardfehler“ (d.h. mit unterstellter Homoskedastie der Fehlerterme) und vergleichen Sie diese Ergebnisse mit Ihrer Schätzung aus Teilaufgabe (a).

Aufgabe 3: Wahr-Falsch-Aussagen

Geben Sie bei jeder Aussage an, ob Sie sie für wahr oder falsch halten, und begründen Sie, warum die betreffende Aussage wahr bzw. falsch ist.

- Wenn der Fehlerterm u heteroskedastisch ist, dann ist der OLS-Schätzer verzerrt.
- Wenn Heteroskedastie vorliegt, dann ist der OLS-Schätzer immer noch effizient.
- Wenn die Annahmen 1–4 des Gauss-Markov Theorems zutreffen, kann man keinen effizienteren linearen Schätzer als den OLS-Schätzer finden, aber es könnte einen konsistenten nichtlinearen Schätzer mit geringerer Varianz geben.
- Auch in kleinen Stichproben liefert OLS unter den Annahmen A1–A3 konsistente Schätzergebnisse.