

Übung zur Vorlesung „Empirische Ökonomie 1“

## Übungsblatt 5: Das multiple lineare Regressionsmodell

### Aufgabe 1:

In dieser Aufgabe beschäftigen wir uns erneut mit der Schätzung von sogenannten Bildungsrenditen. Dazu verwenden wir wieder den Datensatz `NLSY.gdt`, der Ihnen auf unserer Veranstaltungsseite zur Verfügung steht.

Im Übungsblatt 4 haben wir „Fähigkeiten“ als ausgelassene Variable diskutiert. Gehen Sie nun davon aus, dass der IQ diese Fähigkeiten perfekt misst. Betrachten Sie das Regressionsmodell

$$\text{wage}_i = \beta_0 + \beta_1 \text{college}_i + \beta_2 \text{iq}_i + u_i, \quad i = 1, \dots, n. \quad (1)$$

Gehen Sie davon aus, dass dieses Modell korrekt spezifiziert ist und die Annahmen M1 bis M4 erfüllt sind.

- (a) Um wieviel \$/Monat unterscheidet sich das Durchschnittseinkommen von Männern mit College-Abschluss von dem von Männern ohne Abschluss aber mit gleichem IQ?
- (b) Um wieviel \$/Monat erhöht sich das erwartete Einkommen eines Mannes in der relevanten Grundgesamtheit durch einen um 1 Punkt höheren IQ bei gleicher Ausbildung?
- (c) Wieviel \$/Monat beträgt das erwartete Einkommen eines Mannes in der relevanten Grundgesamtheit, der einen Collegeabschluss gemacht hat und einen IQ von 100 Punkten hat?

## Aufgabe 2:

In dieser Aufgabe interessieren wir uns dafür, ob auf dem deutschen Arbeitsmarkt die Löhne nach dem Geschlecht diskriminiert werden.

Dazu verwenden wir den Datensatz `loehne.gdt`, der Informationen über eine Zufallsstichprobe von 657 Arbeitnehmerinnen und Arbeitnehmern in Westdeutschland aus dem Jahr 1996 enthält. Er steht Ihnen auf unserer Veranstaltungsseite zur Verfügung.

Die für uns relevanten Variablen sind:

<code>bildung</code>	Bildung in Jahren
<code>lohn</code>	Bruttolohn [EUR/Stunde]
<code>mann</code>	Dummy, =1 wenn männlich, =0 sonst

- (a) Betrachten Sie zunächst ein einfaches Modell mit `lohn` als abhängiger und `mann` als erklärender Variable. Verdienen Frauen oder Männer mehr? Begründen Sie ihre Antwort.
- (b) Testen Sie die Nullhypothese „Die durchschnittliche Bildung (gemessen in Jahren) unterscheidet sich in der Bevölkerung nicht zwischen Männern und Frauen“ zu einem Signifikanzniveau von 10%. Was ist das Ergebnis?
- (c) Führen Sie nun eine OLS-Regression durch, in der Sie `lohn` als abhängige und `mann` und `bildung` als erklärende Variablen verwenden. Nehmen Sie an, dass die Annahmen M1 bis M4 für dieses Modell erfüllt sind. Testen Sie für dieses Modell die Nullhypothese „Der Durchschnittslohn unterscheidet sich in der Bevölkerung nicht zwischen Männern und Frauen“ zu einem Signifikanzniveau von 1%. Was ist das Ergebnis?
- (d) Interpretieren Sie die geschätzten Koeffizienten des Regressionsmodells aus Teilaufgabe (c).

### Aufgabe 3:

Wir verwenden weiterhin den Datensatz `loehne.gdt` aus der vorhergehenden Aufgabe. Nehmen Sie nun an, Sie seien an dem Lohnunterschied zwischen den Sektoren interessiert. Die für uns relevanten Variablen sind:

`lohn`      Bruttolohn [EUR/Stunde]  
`sektor1`   Dummy, =1 wenn Primärer Sektor (Landwirtschaft)  
`sektor2`   Dummy, =1 wenn Sekundärer Sektor (Verarbeitung)  
`sektor3`   Dummy, =1 wenn Tertiärer Sektor (Dienstleistungen)

Alle beobachteten Individuen arbeiten in einem der drei Sektoren.

- (a) Können Sie ein Regressionsmodell mit `lohn` als abhängiger Variablen schätzen, in dem neben der Konstanten die drei Variablen `sektor1`, `sektor2` und `sektor3` als erklärende Variablen verwendet werden?
  
- (b) Führen Sie die in (a) beschriebene Regression in Gretl durch.