

## Methoden der Ökonometrie - Übung 3

### Aufgabe 1

Betrachten Sie den Datensatz `mtcars` (der Befehl `data(mtcars)` lädt den Datensatz direkt in ihre Arbeitsumgebung). Der Datensatz enthält unter anderem die Variablen `mpg` (miles per gallon), `hp` (horse power) und `wt` (weight).

Das Modell

$$\text{mpg}_t = \beta_0 + \beta_1 \text{hp}_t + \beta_2 \text{wt}_t + u_t, \quad u_t \sim (0, \sigma^2)$$

soll mit OLS geschätzt werden.

- Implementieren Sie die OLS-Formel für das Modell selbst (d.h. nutzen Sie nicht `lm()`) und schätzen Sie die Parameter  $\beta_0$ ,  $\beta_1$  und  $\beta_2$ .  
Hinweis: Wenn Sie alles richtig gemacht haben, sollten Sie  $\hat{\beta}_0 = 37.22727$ ,  $\hat{\beta}_1 = -0.03177$ ,  $\hat{\beta}_2 = -3.87783$  erhalten.
- Führen Sie nun dieselbe Regression mit Hilfe der Funktion `lm()` durch.  
Hinweis: ein Beispiel für ein mit `lm()` geschätztes multiples Regressionsmodell finden Sie am Ende von [Kapitel 6.1](#) in “[Introduction to Econometrics with R](#)”.

### Aufgabe 2

- Führen Sie die folgenden Schritte aus:
  - Erzeugen Sie 150 auf dem Intervall  $[0, 15]$  [stetig uniformverteilte](#) ( $= \mathcal{U}(0, 15)$ -verteilte) Zufallszahlen und weisen Sie diese der Variable `x` zu. Sie können dafür die Funktion `runif()` benutzen.
  - Erzeugen Sie 150  $N(0, 16)$ -verteilte Zufallszahlen (`rnorm()`) und weisen Sie diese der Variable `u` zu.
  - Erzeugen Sie den Vektor `y` wie folgt: `y <- 7 + 3*x + 4*x^3 + u`.
- Plotten Sie `y` gegen `x`. Welcher Zusammenhang zwischen `x` und `y` ist erkennbar?
- Schätzen Sie ein passendes Modell mit Hilfe von `lm()`. Fügen Sie die geschätzte Regressionslinie zum Plot hinzu (nicht ganz einfach, aber trotzdem versuchen).  
Hinweis: Als Hilfestellung können Sie den Abschnitt über nicht lineare Effekte in “[Introduction to Econometrics with R](#)” nutzen.

### Aufgabe 3

Anhand einer Simulationsstudie soll gezeigt werden, dass das arithmetische Mittel und die korrigierte Stichprobenvarianz einer u.i.v. Stichprobe unverzernte Schätzer für den Erwartungswert und die Varianz einer Grundgesamtheit sind. Zeigen Sie außerdem, dass die nicht korrigierte Stichprobenvarianz verzerrt ist.

Gehen Sie dazu wie folgt vor:

- Generieren Sie 20.000 Stichproben der Größe 15 aus einer normalverteilten Grundgesamtheit mit Erwartungswert  $\mu = 3$  und Varianz  $\sigma^2 = 25$ .

- b) Berechnen Sie für jede Stichprobe  $i$  das arithmetische Mittel ( $\bar{x}_i$ ) und die korrigierte Stichprobenvarianz ( $\bar{s}_i^2$ ).
- c) Stellen Sie die Verteilung für das arithmetische Mittel und die korrigierte Stichprobenvarianz graphisch dar.
- d) Approximieren Sie den Erwartungswert für beide Schätzer wie folgt:

$$E[\bar{X}_i] \approx \frac{1}{10000} \sum_i \bar{x}_i \quad \text{und} \quad E[\bar{S}_i^2] \approx \frac{1}{10000} \sum_i \bar{s}_i^2.$$

Spricht ihr Ergebnis für die Unverzerrtheit der beiden Schätzer?

- e) Führen Sie die Simulationsstudie unter gleichen Voraussetzungen für die nicht korrigierte Stichprobenvarianz durch. Was stellen Sie fest?

Hinweis: Eine Hilfestellung und Hintergründe finden Sie u.a. im Kapitel 2.2. von [“Introduction to Econometrics with R”](#).

## Aufgabe 4

Laden Sie den Datensatz „CPS1988“ aus dem AER-Paket.

```
# Falls das Paket noch nicht installiert ist
# install.packages("AER")
library(AER)
data(CPS1988)
```

Die Variablen beschreiben Folgendes: Lohn, Schulzeit in Jahren, Berufserfahrung in Jahren, Hautfarbe, wohnhaft in Metropolregion, Region, Teilzeitstelle der befragten Personen (?CPS1988).

1. Erstellen Sie eine Übersicht deskriptiver Statistiken aller Variablen. Nicht alle Variablen sind numerisch. Welche nicht?
2. Regressieren Sie den Lohn (**wage**) auf Schulzeit (**education**), Berufserfahrung (**experience**), und die Hautfarbe (**ethnicity**) ohne Konstante. Wie können die geschätzten Koeffizienten interpretiert werden?
3. Die Regression aus Aufgabenteil 2 soll nun nochmal mit Konstante durchgeführt werden. Welches Problem ergibt sich dadurch für den OLS Schätzer? Führen Sie die Regression jetzt mit Konstante in R durch. Wie umgeht R das eben angesprochene Problem? Wie sind nun die Koeffizienten zu interpretieren?
4. Überprüfen Sie, ob die durchgeführten Regressionen in Aufgabenteil 2 und 3 zu identischen Ergebnissen führen.
5. Abschließend soll die Regression aus 3 um eine Interaktion zwischen **education** und **ethnicity** ergänzt werden. Schreiben Sie das resultierende Modell auf. Führen Sie dann die entsprechende Regression in R durch. Wie werden die Koeffizienten interpretiert?