

## Aufgabenblatt 1

---

### Überblick

*Querschnittsdaten* beinhalten Beobachtung der gleichen Variablen von verschiedenen Individuen zu einem Zeitpunkt. Manchmal wird die Datenerhebung parallel an verschiedenen Orten durchgeführt oder zu einem späteren Zeitpunkt wiederholt. Dann ergibt sich eine Sammlung an strukturell gleichen Datensätzen, für die jeweils das gleiche Modell angenommen werden kann. Eine naheliegende Idee ist, anstatt separate Modelle zu schätzen, die Daten in einer *Pool* zusammenzuführen.

### Aufgaben

1. Wofür kann das folgende Modell eingesetzt werden?

$$wage = \beta_0 + \beta_1 educ + u \quad (1)$$

2. Unter welcher Annahme ist Modell (1) korrekt spezifiziert?
3. Mit dem Datensatz `wooldridge::cps78_85` (die `::` Notation bedeutet, dass auf das Objekt `cps78_85` im R Paket `wooldridge` zugegriffen wird) kann Modell (1) geschätzt werden. Lesen Sie die Daten in R ein. Unter `?wooldridge::cps78_85` finden Sie eine Beschreibung.
4. Die Variable `educ` ist im Datensatz vorhanden, `wage` fehlt. Aber wir können die `wage` Variable aus dem Datensatz konstruieren – wie?
5. Bei dem Datensatz handelt es sich um einen *gepoolten* Datensatz – warum?
6. Verschaffen Sie sich einen Überblick von der (konstruierten) `wage` und der `educ` Variable und unterscheiden Sie dabei auch zwischen den beiden Erhebungen.
7. Schätzen Sie das *gepoolte Regressionsmodell* in (1) und interpretieren Sie die Koeffizienten. Sind die Koeffizienten statistisch signifikant?
8. Überprüfen Sie, ob die Annahme der Homoskedastie verletzt ist.
9. Welche Auswirkungen hat Heteroskedastie auf die KQ-Schätzung?
10. Wie können wir (hier) mit dem Problem der Heteroskedastie umzugehen?
11. Schätzen Sie nun ein *log-level Modell*, das für 1985 eine eigene Konstante zulässt:

$$\log(wage) = \beta_0 + \beta_1 educ + \gamma_0 y85 + u \quad (2)$$

Welche Interpretation hat  $\gamma_0$ ?

12. Inkludieren Sie zusätzlich den Interaktionsterm ( $y85 \cdot educ$ ) in das Modell (2). Welche Interpretation hat der zugehörige Koeffizient?

13. Wenn Sie das Modell

$$\log(wage) = \beta_0 + \beta_1 educ + u \quad (3)$$

einmal für den gepoolten Datensatz und dann für die Daten aus 1978 und 1985 getrennt schätzen, welche Unterschiede in den Koeffizienten und Varianzen erwarten Sie dann?

14. Die Daten sollten nur gepoolt werden, wenn sich die Koeffizienten der einzelnen Stichproben nicht signifikant voneinander unterscheiden. Bitte führen Sie einen Chow-Test durch, um die Gleichheit der Koeffizienten zu testen.
15. Der Chow-Test basiert auf der Annahme, dass in beiden Datensätzen die Fehlervarianz identisch sind. Wenn das nicht der Fall ist, gibt es trotzdem eine Möglichkeit, diesen Test anzuwenden – welche?