

## Übung Multivariate Analyseverfahren – SoSe 2016

Dr. Volker Ludwig

### Übungsblatt 2 – Abgabe bis 18.05.2016

Wir stellen einen Paneldatensatz für die Jahre 1992 bis 2010 aus den Originaldaten des SOEP zusammen, mit dem wir die Zeitverwendung für Hausarbeit analysieren können. Erweitern Sie das Do-file zu Übungsblatt 1 (03\_Übungsblatt1.do). Erstellen Sie ein neues Do-file (04\_Übungsblatt2.do). Verwenden Sie Schleifen, um alle Jahre 1992,..., 2010 zu matchen.

1. Legen Sie Ihre Arbeitsverzeichnisse fest, in dem Sie globale Macros `pfad1` und `pfad2` definieren. (`pfad1` ist der Ordner, in dem die SOEP-Originaldaten liegen; `pfad2` der Ordner, in dem Sie Ihre rekodierten Daten speichern wollen.)
2. Matchen Sie die Information zur Hausarbeit (in Std., typischer Werktag) aus allen Jahren 1992,..., 2010 und speichern Sie Personennummer (`id`), Befragungsjahr (`year`) und die Variable zur Hausarbeit (`hw`) im neuen Datensatz `hw1.dta`. (Hinweis: Variablen zum Zeitaufwand für Hausarbeit in den Personendaten `$p.dta`:

```
ip0103 jp0203 kp0803 lp0203 mp0203 np0203 op0503 ///
pp02a3 qp05a3 rp0207 sp1103 tp1007 up0203 vp0207 ///
wp6203 xp0207 yp1203 zp0203 bap0303
```

3. Erstellen Sie einen Datensatz `hw2.dta` aus (`$pgen`), der für die Jahre 1992,..., 2010 nur die folgenden Variablen enthält:
  - Personennummer (neuer Variablenname: `id`)
  - Befragungsjahr (`year`)
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Personennummer PartnerIn (`pid`)
  - Familienstand (`famstd`)
  - Bruttomonatsverdienst in Euro (`labgro`)
  - Erwerbsstatus (`lfs`) (**neu**)
4. Erstellen Sie einen Datensatz `hw3.dta` (aus `ppfad.dta`), der für jede Person die folgenden zeitkonstanten Variablen enthält:
  - Personennummer (`id`)
  - Geschlecht (`sex`)
  - Geburtsjahr (`gebjahr`)
  - Teilstichprobe (`psample`) (**neu**)

Der Datensatz soll zusätzlich die zeitveränderliche Variable zum Wohnort in Ost-/Westdeutschland enthalten (`east`). Erstellen Sie diese Variable aus den Variablen `$sampreg`.

(Hinweis: Benennen Sie die Variablen `isampreg, ..., basampreg` in einer Schleife um in `east1992, ..., east2010`. Verwenden Sie dann den Befehl `reshape`, um den Datensatz ins Long Format zu bringen.)

5. Erstellen Sie einen Paneldatensatz `hw4.dta` aus (`$hgen`), der für die Jahre 1992,..., 2010 und für jeden Haushalt nur die folgenden Variablen enthält:
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Befragungsjahr (`year`)
  - (Korrigiertes) Netto-Haushaltseinkommen (`hhinc`)
  
6. Erstellen Sie einen Paneldatensatz `hw5.dta` aus (`$pequiv`), der für die Jahre 1992,..., 2010 nur die folgenden Variablen enthält:
  - Personennummer (`id`)
  - Anzahl HH-Mitglieder im Alter von 0-14 J. (`kids14`)
  - Befragungsjahr (`year`)
  - Region Wohnort Ost-/Westdeutschland (`east2`)
  
7. Spielen Sie jetzt die Daten aus `hw1.dta, ..., hw5.dta` der Reihe nach zusammen. Speichern Sie die Daten als `housework.dta`.
  
8. (\*) Erstellen Sie einen (dyadischen) Paneldatensatz `housework_dyad.dta` für heterosexuelle Paare. Gehen Sie aus von `housework.dta`.
  - Erstellen Sie einen Datensatz `women.dta`, der nur Frauen enthält, die mit einem/r Partner/in zusammen leben (`pid>0`).
  - Erstellen Sie einen Datensatz `men.dta`, der nur Männer enthält, die mit einem/r Partner/in zusammen leben.  
Der Datensatz soll die Variablen `pid hw gebjahr sex labgro famstd lfs` enthalten.
  - Matchen Sie die Personenjahre aus `women.dta` und `men.dta` mit `merge`. (Hinweis: Damit sie die Daten matchen können, müssen Sie in `men.dta` zuerst die Personennummer `pid` des Partners in `id` umbenennen. Außerdem müssen Sie die Variablen `hw, ..., lfs` umbenennen; z.B. könnten Sie in einer Schleife die Endung `_m` an die Variablennamen anhängen.)
  
9. Schicken Sie Ihr Do-file an [vludwig@soziologie.uni-muenchen.de](mailto:vludwig@soziologie.uni-muenchen.de). Geben Sie im Betreff der email das Kürzel „MVSO16/IhrNachname“ an.

**Aufgaben mit (\*) sind Zusatzaufgaben.**