

Abteilung  
Arbeitsmarktpolitik  
und Beschäftigung

Hauptseminar:  
Analyse von Längsschnittdaten  
mit GSOEP und STATA

Dozenten:  
Christian Brzinsky  
Christoph Hilbert

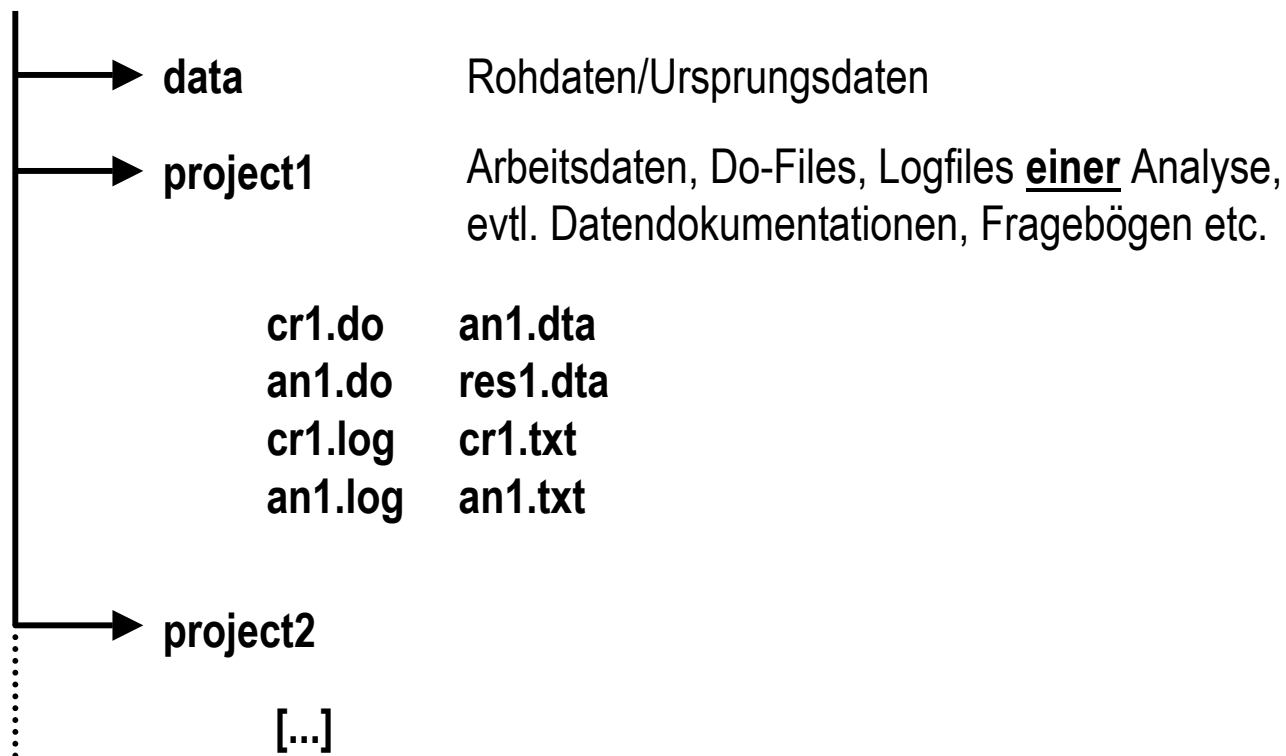
Wintersemester 03/04

# **STATA III: Daten- und Analysevorbereitung**

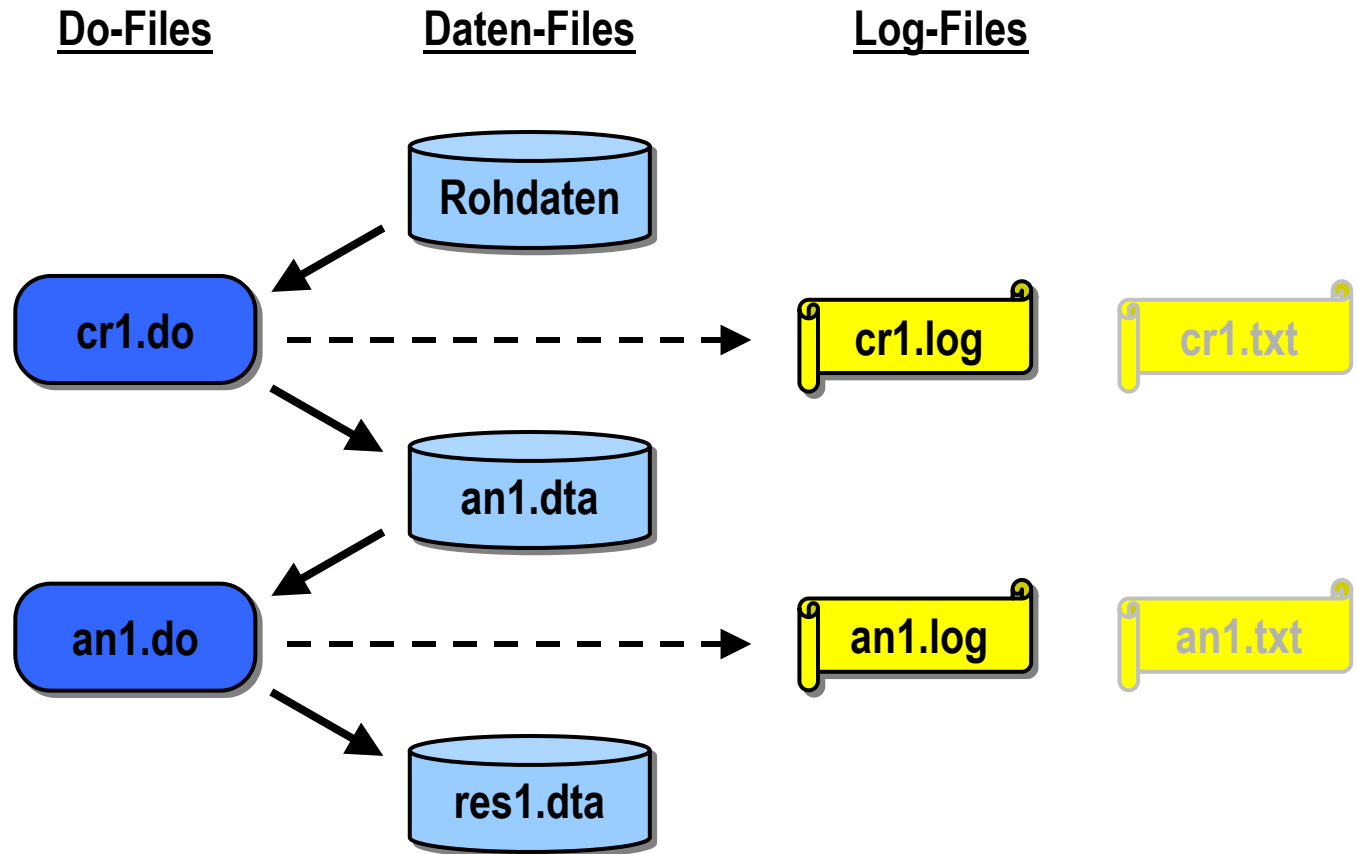
17. November 2003

# Analyse-Organisation und File-Benennung

K:\m014-16



# Beziehungen der Arbeits-Files einer Analyse



# Befehle zum Erstellen neuer Variablen

## Erstellen von neuen Variablen:

### > *generate*

by: [var] generate [var] = [value] if [exp] in [range]

### > *replace*

by: [var] replace [var] = [value] if [exp] in [range]

## Zuweisen von Labels:

### > *label variable*

label variable [var] „[name]“

### > *label define, label values*

label define [labname] [value1] „[name1]“ [value2] „[name2]“ etc.

label values [var] [labname]

## Erstellen neuer Variablen (Beispiel 1)

1. Erstellen Sie eine neue Variable mit dem Namen **lire**, die den Wert ‚0‘ hat.
2. Weisen Sie dieser Variable den Wert ‚1‘ zu, wenn die Parteipräferenz (**rp112**) der Personen Ihrer Meinung nach ‚rechts/konservativ‘ ist.
3. Geben Sie der neuen Variable **lire** ein Variablen-Label und zwei (sinnvolle) Werte-Labels.
4. Erzeugen Sie eine Kreuztabelle aus **lire** und den verschiedenen „Sorgen“-Variablen (**rp11401** bis **rp11410**). Kann man Ihrer Meinung nach von „eher linken“ bzw. „eher rechten“ Sorgen sprechen?

## Erstellen neuer Variablen (Beispiel 2a)

1. Schauen Sie sich die Variable „Geburtsjahr“ (**rp13002**) an. Wie hoch ist der Anteil der unter 25jährigen im aktuellen Datensatz? Wie hoch ist der Anteil der Rentner/innen?
  - > Zur Vermeidung umständlicher Kopfarbeit (Fehleranfälligkeit!) wäre das Alter zum Befragungszeitpunkt interessant.
2. Berechnen Sie eine neue Variable **alter**, die aus dem Geburtsjahr berechnet wird (Hinweis: Welle R des SOEP enthält Daten aus dem Jahr 2001)
3. Schauen Sie sich die neue Variable an. Was genau bedeutet ein Alter von beispielsweise 24 Jahren?

## Erstellen neuer Variablen (Beispiel 2b)

1. Für Aussagen über bestimmte Altersgruppen kann eine Aggregation (Klassenbildung) nützlich sein.
2. Bilden Sie aus der Variable **alter** neue (Dummy-)Variablen folgender Kategorien:
  - unter 25 Jahre **dumalt1**
  - 25 bis unter 35 Jahre **dumalt2**
  - 35 bis unter 45 Jahre **dumalt3**
  - 45 bis unter 55 Jahre **dumalt4**
  - 55 bis unter 65 Jahre **dumalt5**
  - 65 Jahre und älter **dumalt6**
3. Bilden Sie eine Variable **altkat**, die alle sechs Dummy-Variablen zu einer zusammenfasst.

## Erstellen neuer Variablen (Beispiel 2b)

4. Weisen Sie allen neuen Variablen sinnvolle Variablen- und Werte-Label zu.
5. Erzeugen Sie eine Kreuztabelle aus Alterskategorien und Sorgen oder Parteipräferenz (**rp112** und **lire**).
6. Wozu sind die unterschiedlichen Variablentypen der Alterskategorien zu verwenden? Welches Skalenniveau haben die erzeugten Variablen?

### Frage zum Weiterdenken:

- > Was lässt sich bisher über den Zusammenhang zwischen Sorgen und Parteipräferenz bzw. Alter aussagen? Was nicht?

## Erstellen neuer Variablen (Beispiel 2c)

### Methode zur Zusammenfassung einer Variable

```
recode [var] ([old value] = [new value])
```

Erzeugen Sie aus **alter** eine neue zusammengefasste Variable mit neuem Namen.

## Erstellen neuer Variablen (Beispiel 3a)

1. Löschen Sie den im Speicher verbliebenen Datensatz und schalten Sie kurzfristig die Aufzeichnung des *cmdlog* aus.
2. Öffnen Sie die Datei **K:\mo14-16\kkstata\data1.dta**.
3. Sortieren Sie den Datensatz nach der Variable für die Interviewernummer.
4. Lassen Sie sich die ersten 20 Ausprägungen dieser Variable anzeigen.
5. Generieren Sie eine Variable, die für jedes Interview (=Fall) angibt, wie viele Interviews der/die Interviewer/in insgesamt geführt hat.
6. Lassen Sie sich eine Zusammenfassung der Variable **intcount** ausgeben. Was fällt hierbei auf?

## Erstellen neuer Variablen (Beispiel 3b)

1. Geben Sie ein weiteres Beispiel für die Verwendung von `_N` in STATA ein:

```
use hierarch.dta, clear
```

```
sort hhnr
```

```
by hhnr: generate hheink = sum(eink)
```

```
by hhnr: replace hheink = hheink[_N]
```

2. Was wird mit dieser Befehlsfolge erzeugt?

## Abschluss einer STATA-Sitzung

1. Schließen Sie Ihren *log*- und Ihren *cmdlog*-File.
2. Schauen Sie sich die eben erstellten *log*-Files im STATA-Text-Viewer an + versuchen Sie, alle Schritte nachzuvollziehen.
3. Schließen Sie den STATA-Text-Viewer und öffnen Sie Ihren *cmdlog*-File im *do*-File-Editor.
4. Speichern Sie diese Datei als *do*-File (!!!) im Verzeichnis **K:\mo14-16\project2**.
5. Editieren Sie den *do*-File (vgl. **K:\project1\uebung1.do**) und bringen Sie ihn ohne Fehlermeldung zum Laufen.
6. Suchen Sie in den Dateien **rp.dta** und **rpgen.dta** nach für Sie interessanten Variablen und wenden die bisher gelernten Befehle an.