

# **Stata I: Analyseorganisation und erste Schritte**

03. November 2003

## Ziel dieser Veranstaltung ist zu lernen:

1. Benutzung der Hilfe-Funktion (*help, search*)
2. Ein Analyse-File zu erstellen
3. die Arbeit eines Tages zu dokumentieren (*log*)
4. Eine Analyse so aufzubauen, dass sie jederzeit wieder nachvollziehbar und reproduzierbar ist (*do, doedit, cmdlog*)
5. Weitere Kommandos zur Datenorganisation kennen lernen:
  - use
  - save
  - erase
  - sort
  - drop
  - drop if
  - keep
6. Literatur: Kap. Kohler / (Kreuter 2001), Kap. 2 und 3

## Die Hilfe-Funktion in Stata

- > Mit der Stata-Hilfefunktion lassen sich nahezu alle Fragen einfach beantworten, die sich während der laufenden Datenarbeit ergeben
- > Ergänzend gibt es noch umfangreiche Handbücher, die sich vor allem zum Einarbeiten in spezielle Themengebiete eignen
- > Die Stata-Onlinehilfe besteht aus 2 Kommandos: **help** und **search**
  - **help** verwendet man, wenn das Kommando bekannt ist, jedoch detailliertere Informationen erwünscht sind
  - **search** sucht ein Schlüsselwort in der Datenbank und im Internet (unspezifischer)
- > Alternativ kann auch das interaktive Hilfefenster verwendet werden. Der Aufruf erfolgt über den Befehl *whelp* oder über das Pull-Down-Menü "Help". Es bietet jedoch gegenüber der Kommandosteuerung keine Vorteile. Auch im Results-Fenster können mit der Maus direkt Befehle oder Weblinks angesteuert werden.
- > **Beispiel:** *search help* liefert folgenden Output



Stata Results

. search help

Keyword search

Keywords: help

Search: (1) Official help files, FAQs, Examples, SJs, and STBs

Search of official help files, FAQs, Examples, SJs, and STBs

```

GS      . . . . . Getting Started manual
[U]     Chapter 1 . . . . . Read this -- it will help
        (help whatsnew7to8)
[U]     Chapter 2 . . . . . Resources for learning and using Stata
        (help news, sj, stata, viewer)
[U]     Chapter 8 . . . . . Stata's on-line help and search facilities
        (help help, search, viewer)
[U]     Chapter 21 . . . . . Programming Stata
        (help program, macro, quotes, syntax, examplehelpfile)
[U]     Chapter 31 . . . . . Commands everyone should know
[RI]    help . . . . . Obtain on-line help
        (help help)
[RI]    search . . . . . Search Stata documentation
        (help search)
[RI]    view . . . . . View files and logs
        (help view)

FAQ     . . . . . How to fix Stata's Help under Windows
        . . . . . A. Riley
4/99    Stata's Help stopped working: how do I fix it?
        http://www.stata.com/support/faqs/win/whelp.html

FAQ     . . . . . Contacting Stata Technical Support
2/03    How do I contact Technical Support?
        http://www.stata.com/support/faqs/techsup/techsup.html

FAQ     . . . . . Other sources of support
4/99    What are other sources of support?
        http://www.stata.com/support/faqs/techsup/slists.html

```

[R] - Reference Manual.

[U] - User's Guide.

[G] - Graphics Manual.

FAQ - Stata's web site.

help - einzige Quelle ist das  
Help-File

cmdlog off

## Dokumentation von Arbeitstagen: log-Files

- > Es ist sinnvoll und einfach möglich, von Arbeitstagen Protokolle anzulegen, die alle Inputs und Outputs dokumentieren
  - Nachvollziehen von Argumentationslinien (trial & error)
  - Vermeidung von Doppelarbeit
  - Arbeitsnachweis
- > das Kommando **log using** weist Stata an, eine Datei anzulegen in welcher die Ausgaben aller folgenden Befehle geschrieben werden
- > Zum Befehl **log using** gehört die Angabe des Dateinamens, z.B. **an1** für **"Analyse 1"**
- > Die Aufzeichnung erfolgt entweder in ASCII (text) oder in SMCL (Stata Markup and Control Language; enthält Formatierungen)
- > **print log** – druckt log-Files
- > **view log** – zeigt das log-File auf dem Monitor an



**Review**

```
log using an1
use rpgen, replace
sum rbilzeit
mvdecode rbilzeit, mv(-1,
sum rbilzeit
sum rbilzeit, detail
log close
doedit an1
doedit an1.smcl
help log
view an1
view an1.smcl
view an1.smcl
```

**Variables**

Target: Command Window

```
nation01
rpsbil
rpbbil01
rpbbil02
rpbbil03
rpsbila
rpbbila
rpsbilo
rfamstd
rbilzeit
rerwzeit
rtatzeit
rvebzeit
ruebstd
lfs01
is8801
isei01
mps01
nace01
siops01
egp01
klas01
```

Stata Results

Stata Viewer [view an1.smcl]

Back Refresh Search Help Contents What's New News

Command: view an1.smcl

```
log: C:\Eigene Dateien\stata\an1.smcl
log type: smcl
opened on: 27 Oct 2003, 21:28:38

. use rpgen, replace
(RPGEN (12/23/02 18:40:19-431 DATABASE) DB01 12/27/02 14:22:10)
```

```
. sum rbilzeit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rbilzeit	11087	11.05493	3.744953	-2	18

```
. mvdecode rbilzeit, mv(-1, -2)
rbilzeit: 547 missing values generated
```

```
. sum rbilzeit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rbilzeit	10540	11.70579	2.480451	7	18

```
. sum rbilzeit, detail
```

Dauer der Ausbildung, in Jahren

Percentiles	Smallest	Obs	Sum of Hgt.	Mean	Std. Dev.
1%	7	7			
5%	9	7			
10%	9	7			
25%	10.5	7	10540	10540	
50%	11			11.70579	2.480451
	Largest				
75%	12	18			
90%	15	18			6.152638
95%	18	18			1.126428
99%	18	18			3.944075

```
have not reset set
will be fixed to
log is being created.

etermines how the log
extension is added

ype setting will be
Stata.
```

cmdlog off

Stata Command



- Review
- sum rbilzeit
- mvdecode rbilzeit, mv(-1,
- sum rbilzeit
- sum rbilzeit, detail
- log close
- doedit an1
- doedit an1.smcl
- help log
- view an1
- view an1.smcl
- view an1.smcl
- doedit an1.smcl
- doedit an1.smcl

- Variables
- Target: Command Window
- nation01
- rpsbil
- rpbbil01
- rpbbil02
- rpbbil03
- rpsbila
- rpbbila
- rpsbilo
- rfamstd
- rbilzeit
- rerwzeit
- rtatzeit
- rvebzeit
- ruebstd
- lfs01
- is8801
- isei01
- mps01
- nace01
- siops01
- egp01
- klas01

```

is written. If filename was specified without an extension, the appropriate extension is added
ng will be

an1.smcl - Stata Do-file Editor
File Edit Search Tools

(smcl)
(com)(sf)(ul off)(txt)(.-)
      log: (res)C:\Eigene Dateien\stata\an1.smcl
      (txt)log type: (res)smcl
      (txt)opened on: (res)27 Oct 2003, 21:28:38
      (txt)
      (com). use rpgen, replace
      (txt)(RPGEN (12/23/02 18:40:19-431 DATABASE) DB01      12/27/02 14:22:10)

      (com). sum rbilzeit

      (txt)      Variable (c |)      Obs      Mean      Std. Dev.      Min      Max
      (hline 13)(c +)(hline 56)
      rbilzeit (c |)(res)      11087      11.05493      3.744953      -2      18
      (txt)
      (com). mvdecode rbilzeit, mv(-1, -2)
      (txt)rbilzeit:(res)(col 15)547(txt) missing values generated

      (com). sum rbilzeit

      (txt)      Variable (c |)      Obs      Mean      Std. Dev.      Min      Max
      (hline 13)(c +)(hline 56)
Line number: 1

doedit an1.smcl

```

Stata Command

## Reproduzierbarkeit von Analysen: do-Files

Wissenschaft erhebt den Anspruch der Objektivität. Zentrales Kriterium für Objektivität ist nach Popper die **intersubjektive Nachvollziehbarkeit**: anderen soll es möglich sein, die Forschungsergebnisse mit den gleichen Methoden zu bestätigen oder zu kritisieren. Voraussetzung ist eine sorgfältige Dokumentation

### Was sind Do-Files?

- > Einfache Textdateien mit Stata-Befehlen, die hintereinander ausgeführt werden
- > Zentrales Werkzeug , um Reproduzierbarkeit zu erreichen
- > Auch für die eigene Arbeit ist eine Dokumentation hilfreich: sollen Analysen aktualisiert oder beispielsweise die Zielgruppe verändert werden, kann auf ‚frühere‘ Arbeiten zurückgegriffen werden.

### Vom interaktiven Arbeiten zum do-File:

- > Voraussetzung: Dokumentation mit log-Files
- > Öffnen eines log-Files im Editor (*view log \*.log*)
- > Speichern in einem \*.do-File
- > Output bis auf relevante Befehle löschen

## Organisation von do-Files

### Sinnvoll ist eine Unterscheidung von:

#### 1. Create-do-Files,

mit denen Datensätze erzeugt werden (z.B. *crdata1.do*)

#### 2. Analyzing-do-Files,

mit denen Datensätze analysiert werden (z.B. *an1.do*)

#### 3. Master-File:

Für jedes Projekt wird ein Master angelegt. Es handelt sich dabei um ein do-File, in der die einzelnen do-Files eines Projektes (*sowohl cr-Files, als auch an-Files*) nacheinander ausführen.

## Befehle, die in keinem do-File fehlen sollten

```
-----an1.do  
version 8.0  
set more off  
capture log close  
log using an1.log  
....  
-----
```

**version** - gibt an, in welcher Version das do-File geschrieben ist. Damit ist sicher gestellt, dass das do-File auch fehlerfrei läuft

**set more off** - stellt sicher, dass das do-File auch ohne Unterbrechung abläuft, das heisst, die 'automatische Bremse' des Ausgabebildschirm ist ausgeschaltet

**capture log close** - Schließen eines evtl. geöffneten log-Files. *capture* bewirkt, dass keine Fehlermeldung angezeigt wird, wenn gar kein log-File offen ist.

**log using an1.log, replace** - Aufzeichnung eines log-Files. Dateiname richtet sich nach dem des do-Files. Option '*replace*', da log-File möglicherweise bereits existiert

## Effizientes Erstellen von do-Files: *cmd-log*

- > ***cmdlog*** - erlaubt es, nur die Befehle, die während einer interaktiven Sitzung gemacht werden aufzuzeichnen. Damit handelt es sich um einen Ausschnitt aus dem vollständigen log-File.
- > ***log using*** und ***cmdlog using*** öffnen ein log-File. ***log close*** and ***cmdlog close*** schließen es. Dazwischen können "Aufzeichnungspausen" mit *log off* and *cmdlog off*, ind *log on* and *cmdlog on*
- > ***log*** und ***cmdlog*** können parallel, nacheinander oder einzeln ausgeführt werden
- > Command Logs sind immer einfache ASCII-Files (\*.txt) und können deshalb schnell in do-Files umgewandelt werden.
- > Alternativ: gleich als do-Files gespeichert werden, indem die Erweiterung beim Speichern angegeben wird (*cmdlog using an1.do*)
- > ***cmdlog*** (wie auch ***log***) ohne Argument zeigt den Status des *Loggings* an
- > **Optionen**
  - ***append***: spezifiziert, dass Ergebnisse an das Ende des bisherigen Files angehängt werden. Wenn das File noch nicht existiert, wird ein neues erstellt.
  - ***replace***: wenn das File bereits existiert, so wird es überschrieben
  - Erfolgt die Spezifizierung nicht, dann gibt Stata eine Fehlermeldung aus, so das File bereits existiert

## Die meisten Kommandos folgen dem gleichen Aufbau:

### **. *command varlist if exp in range, option***

#### **> *command***

Beschreibt das, was wir tun wollen, wie summarize, list, save, etc.

#### **> *varlist***

Die Variable(n), auf die wir dieses Kommando anwenden möchten. Varlist ist meist optional. Wird keine Variable angegeben, so erfolgt der Prozess auf allen Variablen.

#### **> *if exp(ression)***

Spezifiziert die Fälle, die wir bearbeiten möchten. *list rbilzeit if rbilzeit > 10* zeigt die Variable *rbilzeit*, aber nur die Fälle, deren Bildungszeit mehr als 10 Jahre beträgt. Achtung: Gleichheitsbedingungen immer mit doppeltem Gleichheitszeichen (==) angeben!

#### **> *in range***

prinzipiell die gleiche Funktion wie *if exp*, allerdings bezieht es sich auf Beobachtungen (Fälle) *list rbilzeit in 1/10* listet die Beobachtungen *1 bis 10* auf.



```
*-----  
version 8.0  
set more off  
capture log close  
log using crgsoep1.log, replace  
*-----grogsoep1.do-----  
*Generieren eines Testdatensatzes aus dem GSOEP-----  
*zu Einkommen und Bildungszeit, Welle R (2001)-----  
*-----  
  
*-----[ automatically pull PPFAD ]-----  
  
use "c:\data\soep19\ppfad.dta"  
keep hnr persnr sex gebjahr psample rhhnr rnetto  
  
*-----[ balanced / unbalanced design ]-----  
  
keep if (rnetto == 1)  
  
sort hnr persnr  
save "c:\data\ppfad.dta"  
*clear  
  
*-----[ automatically pull PHRF ]-----  
  
use "c:\data\soep19\phrf.dta"  
keep hnr persnr prgroup rphrf  
  
sort hnr persnr  
save "c:\data\phrf.dta"  
clear  
  
*-----[ automatically create pmaster ]-----  
  
use "c:\data\soep19\ppfad.dta"  
sort hnr persnr  
merge hnr persnr using "c:\data\phrf.dta"  
  
drop if _merge == 2  
drop _merge
```

## Welche "Fenster" gibt es in Stata?

### Interaktiver Output einer Stata-Sitzung

- > Stata Results
- > Review
- > Variables

### Projektorganisation und Dokumentationen

- > Log-Files
- > Command-Logs
- > Do-Files