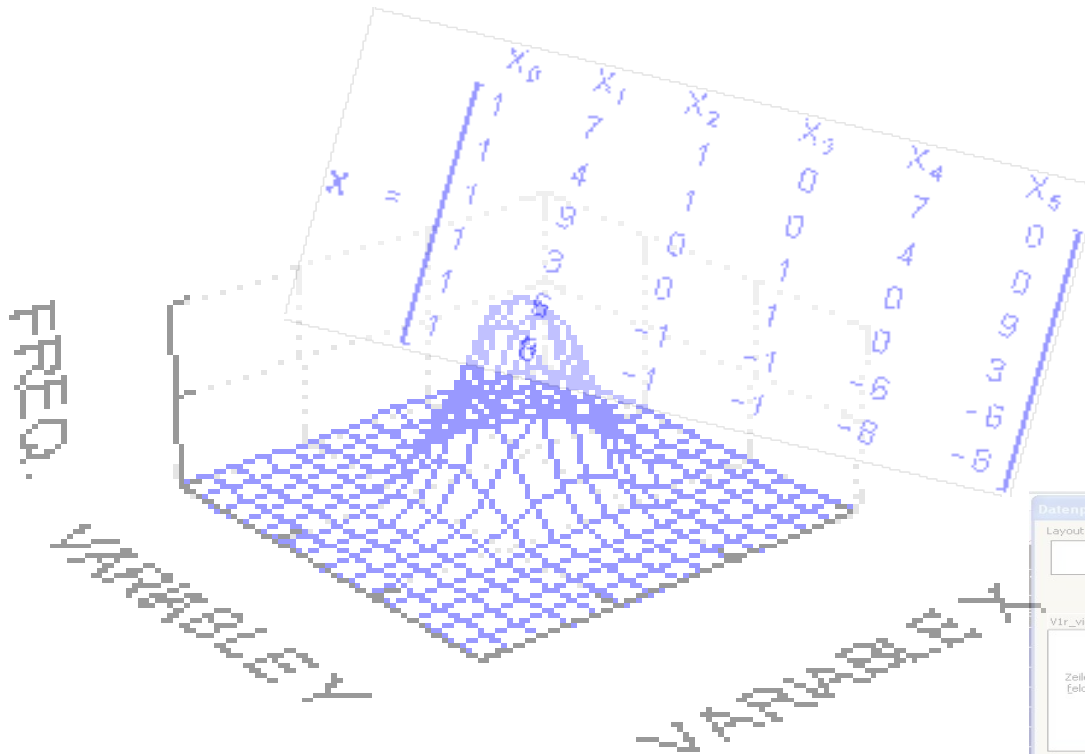
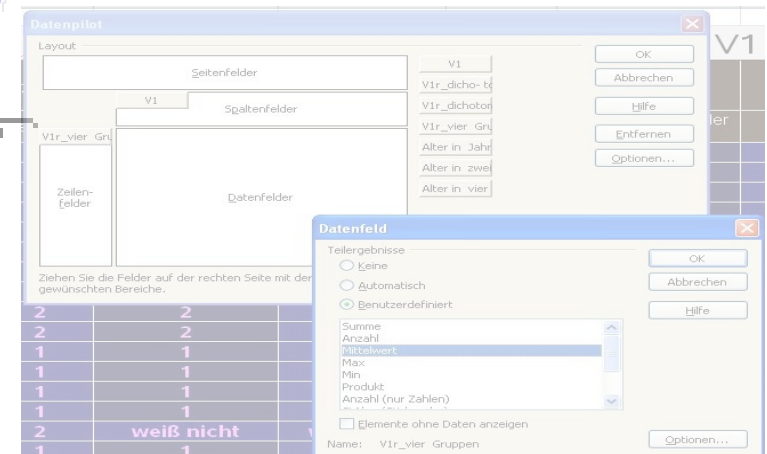


Statistik IV

1. Grundlagen und Datenaufbereitung

Ziehen Sie die Felder auf der rechten Seite mit der gewünschten Bereiche.

2	2	
2	2	
1	1	
1	1	
1	1	
1	1	
2	weiß nicht	
1	1	

Terminplanung

<i>Nr.</i>	<i>Termin</i>	<i>Inhalt</i>
1	14.04.09	Einführung <ul style="list-style-type: none"> • Organisatorisches und Scheinvoraussetzungen • Statistik mit Softwareunterstützung?
2	28.04.09	Grundlagen Stata <ul style="list-style-type: none"> • Aufbau des Programms • Umgang mit Daten, Dokumentation Datenaufbereitung für statistische Auswertungen I
3	12.05.09	Datenaufbereitung für statistische Auswertungen II
4	26.05.09	Deskriptive Statistik und graphische Darstellung
5	09.06.09	Zusammenhangsanalyse I
6	23.06.09	Zusammenhangsanalyse II
7	07.07.09	Zusammenhangsanalyse III Probeklausur Besprechung der Probeklausur
8	14.07.09	Klausur (90 Minuten)

Übersicht

- **Grundlagen**
 - Arbeitsumgebung
 - Daten
 - Programm
- **Datenaufbereitung I**

Arbeitsumgebung

- Virtueller Semiarordner:

<http://www.uni-goettingen.de/de/111208.html>

- Ausgangsverzeichnis:

V:\lehre\stata_leibold

- Literatur zu STATA im permanenten Handapparat

ALLBUS

ALLBUS 2008

- Allgemeine Bevölkerungsumfrage der Sozialwissenschaften 2008
- Seit 1980 sich regelmäßig wiederholender (alle 2 Jahre) repräsentativer Bevölkerungsquerschnitt (n = 2.800 - 3.500)
- Seit 1986 incl. der deutschen ISSP-Stichprobe
- ALLBUS 2008 + ISSP
 - Politische Einstellung
 - Mediennutzung
 - Freizeitverhalten und Religion (ISSP)
 - Standarddemographie (sehr umfangreich)

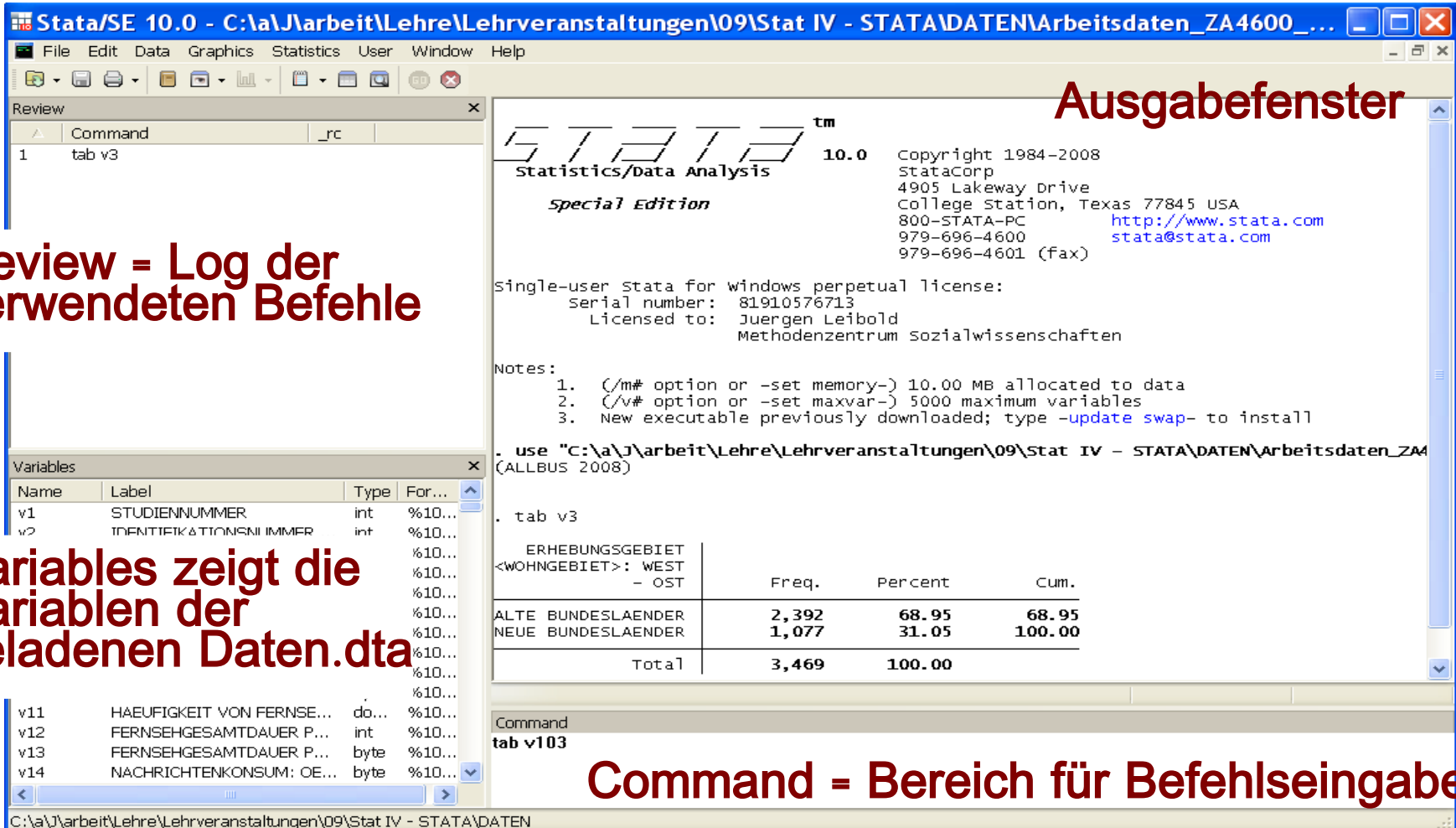
ALLBUS Dokumentation

- GESIS – Datenaufbereitung und Dokumentation
- Online verfügbar oder per DVD
 - Daten im SPSS (SAV-Format) und STATA (DTA-Format)
 - Codebooks bzw. Datenhandbücher
 - Fragebögen
- Letztere befinden sich im Ordner Dokumentation

- Bevor wir die Daten mit STATA behandeln sollten wir uns mit den Fragenprogramm des ALLBUS vertraut machen.

STATA 10 – Aufbau des GUI

STATA 10



Ausgabefenster

Review = Log der verwendeten Befehle

Variables zeigt die Variablen der geladenen Daten.dta

Command = Bereich für Befehlseingaben

Copyright 1984-2008
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user stata for windows perpetual license:
Serial number: 81910576713
Licensed to: Juergen Leibold
Methodenzentrum Sozialwissenschaften

Notes:
1. (/m# option or -set memory-) 10.00 MB allocated to data
2. (/v# option or -set maxvar-) 5000 maximum variables
3. New executable previously downloaded; type -update swap- to install

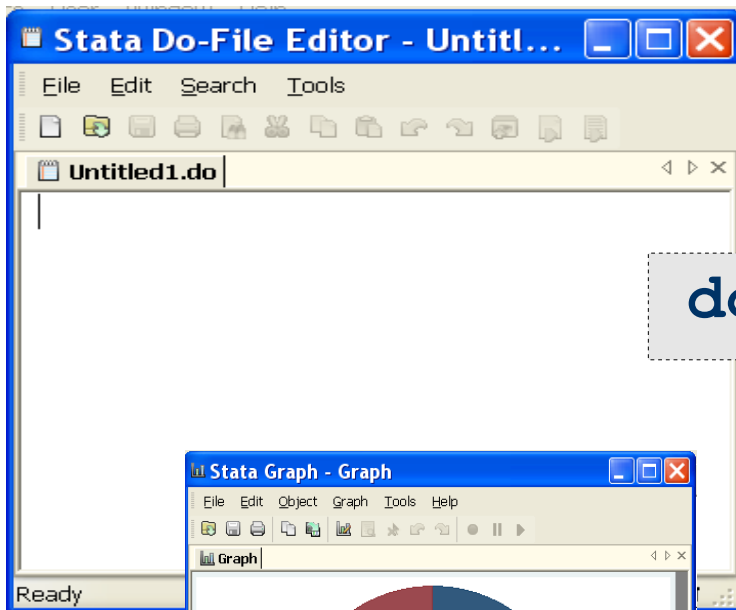
```
. use "C:\a\j\arbeit\Lehre\Lehrveranstaltungen\09\Stat IV - STATA\DATEN\Arbeitsdaten_ZA4600_..."
(ALLBUS 2008)
```

```
. tab v3
```

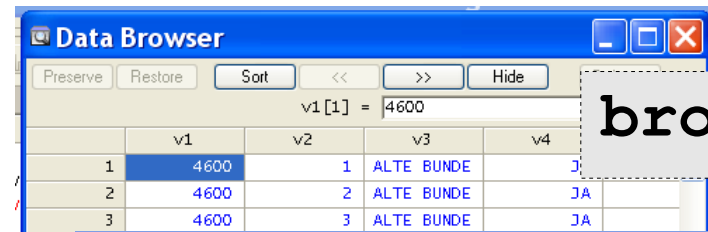
	Freq.	Percent	Cum.
ERHEBUNGSGEBIET <WOHNGBIET>: WEST			
- OST			
ALTE BUNDESSTAENDEN	2,392	68.95	68.95
NEUE BUNDESSTAENDEN	1,077	31.05	100.00
Total	3,469	100.00	

```
Command
tab v103
```

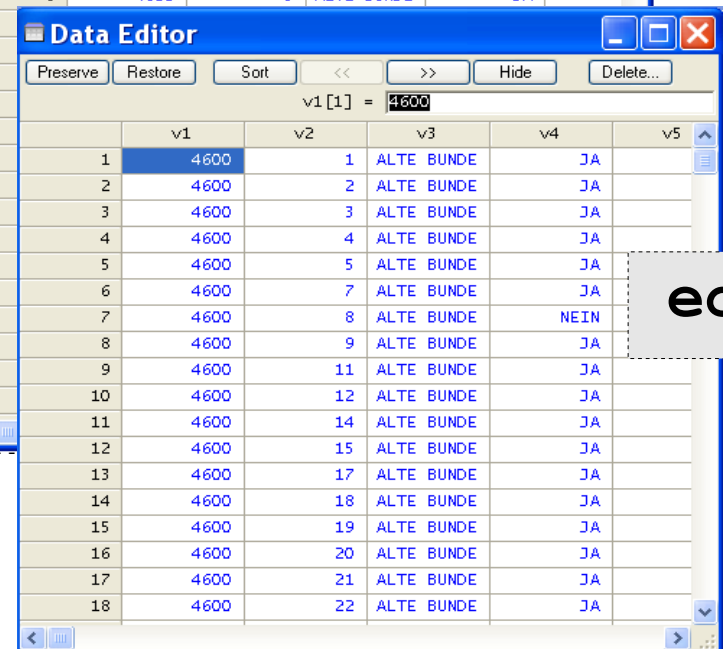
Weitere Fenster



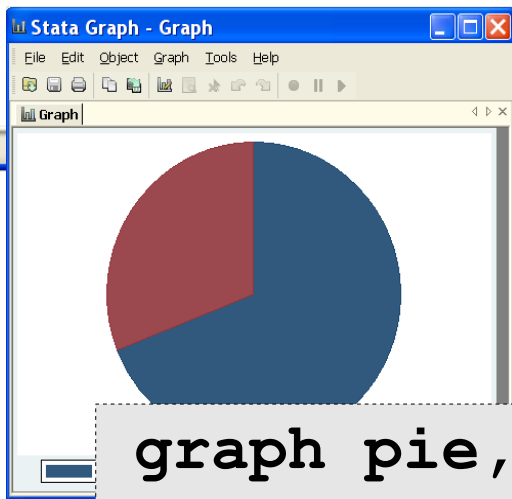
doedit



browse

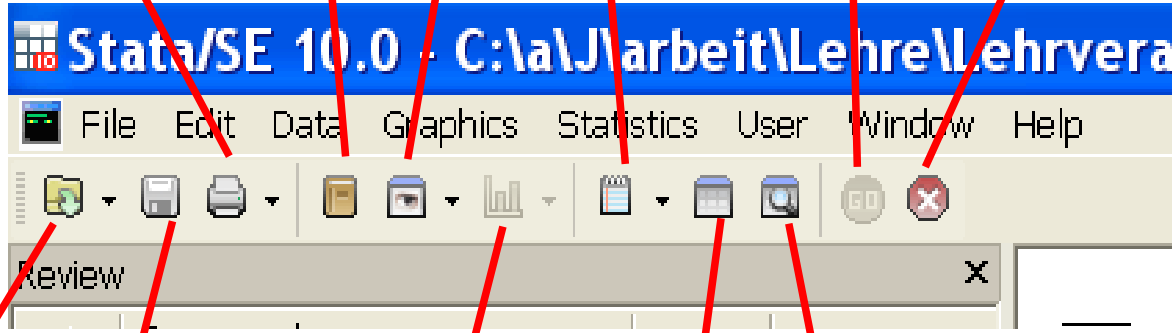


edit



graph pie, over(v3)

Icon-Bar



The screenshot shows the Stata 10.0 icon bar with the following labels and arrows:

- Drucken** (Print) points to the printer icon.
- Dateien öffnen** (Open files) points to the folder icon.
- Dateien speichern** (Save files) points to the floppy disk icon.
- Logfile** points to the document icon.
- Viewer** points to the document with magnifying glass icon.
- Do-file** points to the document with list icon.
- Graph-Fenster** (Graph window) points to the bar chart icon.
- Daten-editor** (Data editor) points to the document with list icon.
- Daten-browser** (Data browser) points to the document with magnifying glass icon.
- Ausgabe beenden** (End output) points to the 'GD' icon.
- Ausgabe fortsetzen** (Resume output) points to the 'X' icon.

Arbeiten mit STATA

Log-File der Ausgabe

- Klicken Sie auf das Log-File-Icon.
- Geben Sie dem neuen Log-File einen Namen.
- Ab jetzt werden die Ausgaben der Sitzung in der angegebenen Datei geloggt.

Datensätze aufrufen

1. Windows-Explorer (Doppelklick auf Datendatei)
2. Befehl: `use "v:\.....\ALLBUS08.dta"`
3. Icon → Datei öffnen
4. Menü → File; Open
5. [Strg]+[o]

Datensatz speichern

1. Befehl: `save "v:\.....\ALLBUS08_Arbeitsdaten.dta"`
2. Icon → Datei speichern
3. Menü → File; Save AS
4. [Shift]+[Strg]+[s]

1. Aufgabe

- Daten mit STATA aufrufen und im eigenen Arbeitsbereich abspeichern.

Mittel zur Datenübersicht

- Data Browse
- **describe** / Auflistung der Variablen
- **summarize** / Fälle/Mean/Std.Dev/Min/Max
- **list** / Auflistung aller Ausprägungen
nach Fällen sortiert
- **[des; sum; list] v3**
- / Auflistungen nur für v3
- **list in 1** oder **list in 2/4**
/Auflistung nur für ausgewählte Fälle
- **codebook v3** /Auflistung der Häufigkeiten mit Wert und Label
- **lookfor Bundestag**
/sucht nach einer Zeichenfolge (hier Bundestag)

Aufgabe 2

■ Orientierungseinheit

- Welche Variablen befinden sich am Anfang und Ende des Datensatzes?
- Wie viele Variablen sind im Datensatz enthalten?
- Welche Variablen beinhalten das Alter und das Geschlecht der befragten Person?
- Welchen Mittelwert hat die Variable Altersvariable?
- Mit welchen Werten ist die Item-Nonresponse der Variablen v22 – v25 und v633 codiert?

Do-File als Dokumentation

- Review-Fenster
- rechte Maustaste
 - Select All
- rechte Maustaste
 - Send to Do-File-Editor
 - Datei speichern

Fehlende Werte / Missings

- `tabulate v633` / Häufigkeiten
- `mvdecode v633, mv (99.96=.a)`
/ Missing setzen
- `replace v633 = 99 if v633 == .a`
 - oder
- `mvencode v633, mv (.a=99)`
/ Missing in Wert zurückgesetzt
- Missingwerte beginnen immer mit einem Punkt und danach kommt ein Buchstabe [`.a` oder z.B. `.z`]
- Es sind damit nur 27 verschiedene Missingwerte möglich

Missingbereiche

- `replace v633= .a if v633 >=80`
- `tabulate v633`
- `tabulate v633 , mis`

foreach - mehrere Variablen gleichzeitig (hier mit Missings versehen)

- `codebook v22 v23 v24 v25`
- ```
foreach var01 of varlist v22 v23 v24 v25 {
replace `var01'=.a if `var01' >7
}
```

# Aufgabe 3

- Für die Variablen zum *Vertrauen in Institutionen* im ALLBUS 2008 Missings setzen

# Neue Variablen erzeugen

## Index - Vertrauen in Institutionen der Bundespolitik

- `generate i_bund = v73 + v79` /erzeugt den Summenindex
- `tab i_bund`
- `drop i_bund` /wirft die Variable weg
- `foreach var of varlist v73 v79 {`  
`replace `var'=.a if `var'>90`  
`}`
- `generate i_bund = v73 + v79`
- `tab i_bund`



# Werte zusammenfassen

- `recode v633 (min/19 = 1) (19.0001/25=2)  
(25.0001/80=3) , gen(BMI)`

# Variablen und Werte benennen

- `label variable BMI "BMI-3Kat"`
- `label define BMI_lb 1"mager" 3"dick"`
- `label value BMI BMI_lb`

# Aufgabe 4

- Berechnen Sie einen Summenindex für das Vertrauen in Medien.
- Berechnen Sie einen multiplikativen Index für das Vertrauen in Medien und Bundespolitik aus den beiden neuen Indizes.
- Erzeugen Sie eine Variable, die orientiert am Median nur zwei Werte enthält 0=geringes und 1=großes Vertrauen des Medien\*Bund-Index enthält
- Vergeben Sie Labels für die Variablen und Wertelabels für die letzte Variable

# Umbenennen und Clonen von Variablen

- `tab i_bund`
- `rename i_bund Index01`
- `clone Index02=Index01`
- `tab Index01 Index02`

# Extension to generate [egen]

- Erstellen Sie einen Index auf Basis des Mittelwerts einer Variablenliste unter Berücksichtigung der Missing-Values
  - `egen i_bund=rowmean(v73 v79)`
- Erstellen Sie eine Variable, die die Anzahl der Missings einer Liste beinhaltet.
  - `egen i_bundmis=rowmiss(v73 v79)`
- Erstellen Sie eine Variable, die die Anzahl der gültigen Antworten einer Variablenliste beinhaltet
  - `egen i_bundnonmis=rownonmiss(v73 v79)`

# Aufgabe 5

- Erzeugen Sie aus allen Vertrauensvariablen einen Index der auf dem Mittelwert beruht.
- Vergeben Sie Labels für die Variablen und Wertelabels für die letzte Variable
- Speichern Sie Ihren Datensatz

# Hausaufgabe

- Versehen Sie alle Variablen des Datensatzes mit STATA-Missings
- Hinweis: Die Arbeit kann aufgeteilt werden