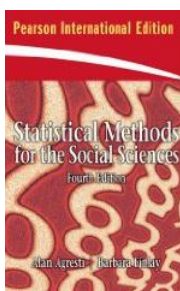


Worum geht es hier?

- ▶ Eine von zwei zentralen Vorlesungen im Aufbaumodul I
- ▶ Für BA-Kernfach
- ▶ Vertiefung und Erweiterung der Grundkenntnisse
- ▶ **Plus praktische Anwendung**
- ▶ Berufsqualifizierende Übung (am Computer) zur Vorlesung (verpflichtend)
- ▶ Weitere Bestandteile des Moduls: VL Methoden II + Seminar Berufsfeldqualifikation (z. B. Theorie der Umfrageforschung) + Übung „Anwendung von Forschungsmethoden“ (z. B. Auswertung von Mikro-Daten)
- ▶ **Gemeinsame Klausur für Statistik II/Methoden II**

Literatur etc.



- ▶ Mehr Lehrbücher:
 - ▶ Fox, John: Applied regression analysis, linear models, and related methods
 - ▶ Gelman, Andrew & Hill, Jennifer: Data Analysis Using Regression and Multilevel/Hierarchical Models
 - ▶ Kennedy, Peter: A Guide to Econometrics (5e)
 - ▶ Wooldridge, Jefferey: Introductory Econometrics
- ▶ Homepage: <http://kai-arzheimer.com/Statistik-II/>



- ▶ Fox, John: A mathematical primer for social statistics.
- ▶ Gill, J.: Essential Mathematics for Political and Social Research.



Nach-/Vorbereitung

- ▶ Weitere Literatur im ReaderPlus Nr. 6007
- ▶ Registrierungscode ba-pol-statistik-2-2009
- ▶ Für heute: Kohler/Kreuter Seite 7-28, <https://www.zdv.uni-mainz.de/readerplus/mav/1/miv/0/nph-ld/14545/31197.07803!1!1256220117!/document.pdf?Log=1>
- ▶ Außerdem: Kohler/Kreuter Seite 145-157, <https://www.zdv.uni-mainz.de/readerplus/mav/1/miv/0/nph-ld/14545/31197.07803!1!1256220610!/document.pdf?Log=1>
- ▶ Für nächste Woche: in Kohler/Kreuter stöbern

Was steht auf dem Programm?

Vorlesung	Inhalt
1	Wiederholung, Stata, Deskriptive Statistik
2	Mittelwerte, Zusammenhangsmaße, Hypothesentests in Stata
3	Was ist Regression?
4	Partielle Korrelation, Drittvariablenkontrolle und Gewichtung
5	Multikollinearität und Kohortenanalyse
6	ANOVA, Transformationen
7	Schätzverfahren, Annahmen, Verletzungen

Was steht auf dem Programm? II

Vorlesung	Inhalt
8	Einführung Logit und Probit
9	Multinomiale Logitmodelle
10	Einführung Count/Event Data
11	Einführung Zeitreihen
12	Einführung Panel/TSCS
13	Einführung Mehrebenenanalyse
14	Einführung Faktorenanalyse

Was sind Daten?

- ▶ Empirische Beobachtungen/Messungen an einem politischen Objekt (z.B. Person, Staat, Parteiprogramm, Parlament)
- ▶ Quantitativ
- ▶ Verschiedene „Skalenniveaus“
 1. Ratio-Skalen
 2. Intervall-Skalen
 3. Ordinal-Skalen
 4. Nominal-Skalen
- ▶ Speicherung auf Tontafeln/Papyrus/Papier (früher)
- ▶ Heute mit Computer (Lochkarten, Magnetbänder, Festplatten, CD/DVD ...)

Die Welt früher: Lochkarten

KUNDEN NR.	KA	A-I		TEXT	A-I		BETRAG		BERLINER PRESSE-VERTRIEB
		J-R	J-R		J-R	DM	PF		
0 0 0 0 0 0	0	S-Z	0 0	S-Z	0 0	0 0 0 0 0 0	0 0		
1 1 1 1 1 1	1	AJS	1 1	AJS	1 1	1 1 1 1 1 1	1 1		
2 2 2 2 2 2	2	BKT	2 2	BKT	2 2	2 2 2 2 2 2	2 2		
3 3 3 3 3 3	3	CLU	3 3	CLU	3 3	3 3 3 3 3 3	3 3		
4 4 4 4 4 4	4	DMV	4 4	DMV	4 4	4 4 4 4 4 4	4 4		
5 5 5 5 5 5	5	ENW	5 5	ENW	5 5	5 5 5 5 5 5	5 5	SUMMENKARTE	
6 6 6 6 6 6	6	FOX	6 6	FOX	6 6	6 6 6 6 6 6	6 6		
7 7 7 7 7 7	7	GPY	7 7	GPY	7 7	7 7 7 7 7 7	7 7		
8 8 8 8 8 8	8	HQZ	8 8	HQZ	8 8	8 8 8 8 8 8	8 8		
9 9 9 9 9 9	9	IR	9 9	IR	9 9	9 9 9 9 9 9	9 9		
1 2 3 4 5 6	7		8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31		32 33 34 35 36 37 38 39 40				

Was ist eine „(Roh)datenmatrix“?

- ▶ „Rohe“ (nicht-bearbeitete) Meßwerte → Tabelle
 - ▶ Speicherung intern und auf Festplatte/Speichermedium (Datei/Format)
1. Zeilen = „Fälle“ (Untersuchungsobjekt = Personen, Länder, Départements etc.)
 2. Spalten = „Variablen“ (Eigenschaft = Nationalität, Links-Rechts-Wert etc.)

	cntry	idno	lrscale	trstplt	trstplc	vote
1.	IT	3183200	6	2	7	yes
2.	SE	202429	6	3	6	yes
3.	AT	602	5	no trust	8	yes
4.	AT	1934	5	6	9	not elig
5.	IT	3583300	left	6	8	yes
6.	IT	3457000	4	7	9	yes
7.	AT	1216	4	4	7	yes
8.	IT	3120400	6	5	7	yes
9.	IT	3029500	left	5	complete	yes

Statistik II Einführung (11/1)

Wie kommen die Daten in den Computer?

- ▶ Früher: pencil & paper →Eingabe von Hand (Fehler)
- ▶ Heute: (oft) CAPI/CATI, Web oder Datenbanken →„maschinenlesbare“ Daten
- ▶ **Bitte an Online-Befragung teilnehmen**
- ▶ **Aber:** Vielzahl von Formaten (Arbeit + Fehler)
- ▶ Unzahl von verschiedenen *logischen* Formaten z. B. für Zeit und Datum (1. November 2009, 01/11/09, 11/01/2009, Nov 1 2009 ...)
- ▶ Unterschiedliche *physische Speicherkonventionen* (Unix vs. DOS/Windows, 32/64 bit, big-endian/little-endian)
- ▶ Vielzahl von Programmen mit (proprietären) Formaten (logischer Aufbau der Dateien im Computer)

Statistik II Einführung (12/1)

Was sind die wichtigen maschinenlesbaren Formate?

1. Generische (Text) Formate
 - ▶ ASCII (frei, fest, kommasepariert)
 - ▶ Extension z. B. *.txt, *.dat, *.csv
 2. Proprietäre (binäre) Formate
 - ▶ Excel (*.xls)
 - ▶ SPSS (*.sav, *.por)
 - ▶ Stata (*.dta)
 - ▶ ...
- ▶ (Fast) alle Programme lesen Textformate
 - ▶ Probleme: Ineffizient (Dateigröße), Details
 - ▶ Proprietäre Formate erfordern Konversionsprogramme/-routinen
 - ▶ Probleme: Details, konzeptuelle Unterschiede (z. B. missing values)

Wozu braucht man ein Statistikprogramm?

1. Zugriff auf Daten
2. Verwaltung von Daten
 - ▶ Auffinden von Informationen
 - ▶ Kombinieren von Daten aus verschiedenen Quellen
 - ▶ *Kontrolliertes* Verändern/Kopieren von Rohdaten + Dokumentation → Skriptfähigkeit
3. *Reproduzierbare* Analyse von Daten → Skriptfähigkeit
4. Erstellen von Grafiken und Tabellen → Skriptfähigkeit wünschenswert

Welche wichtigen Programme gibt es?

1. Spezielle Programme für besondere Verfahren:
 - ▶ AMOS, Lisrel, MPlus, Mwin
 - ▶ Interoperabilität, beschränkte Möglichkeiten zur Datenverwaltung
 2. Generelle Programme für (fast) alle Aufgaben
 - ▶ Excel (nicht wirklich)
 - ▶ SAS (Dinosaurier, in manchen kommerziellen Bereichen)
 - ▶ SPSS (point & click, Marktforschung, noch an Universitäten)
 - ▶ R (open source, Programmiersprache, an Hochschulen in *fortgeschrittenen* Bereichen)
 - ▶ Stata (zunehmend in Wirtschafts- und Sozialwissenschaften)
- ▶ Vorteile von Stata
- ▶ Kann (fast) alles, was man bis zur Promotion braucht
 - ▶ Hervorragende Dokumentation, aktive Nutzergemeinschaft
 - ▶ Relativ leicht programmier- und erweiterbar
 - ▶ Wer Stata kann, lernt sehr schnell SPSS

Wie kann man Stata benutzen?

- ▶ Anmelden auf Windows-Terminalserver
ts.zdv.uni-mainz.de (aus PC-Pool oder von zu Hause,
Erläuterung: <http://www.zdv.uni-mainz.de/3222.php>)
- ▶ Stata aus Start-Menu auswählen
- ▶ Befehle per Menü/Eingabebox oder über Eingabezeile
(bevorzugt) eingeben
- ▶ Befehle und Variablennamen können abgekürzt werden
- ▶ Ausführen → Return-Taste; Befehle müssen nicht mit Punkt,
Komma oder Semikolon abgeschlossen werden
- ▶ Komplexere Befehle per Editor
bearbeiten/speichern/ausführen → *.do-Files
- ▶ Ein- und Ausgaben können protokolliert werden →
Reproduzierbarkeit und Automatisierung

Stata gestartet

The screenshot shows the Stata 10.1 interface with the following components:

- Menüs und Buttons:** The top menu bar includes File, Edit, Data, Graphics, Statistics, User, Window, and Help.
- bisherige Kommandos (History):** A list of commands entered, including `use`, `set mem`, `clear`, and `tab`.
- Variablen:** A list of variables with their names, labels, types, and formats.
- Ergebnisse:** The output window showing the command `tab v3` and its results, including a table for `erhebungsgebiet:`.
- Eingaben:** The command input line at the bottom.

erhebungsgebiet:	Freq.	Percent	Cum.
west - ost			
alte bundeslaender	1,982	67.28	67.28
neue bundeslaender	964	32.72	100.00
Total	2,946	100.00	

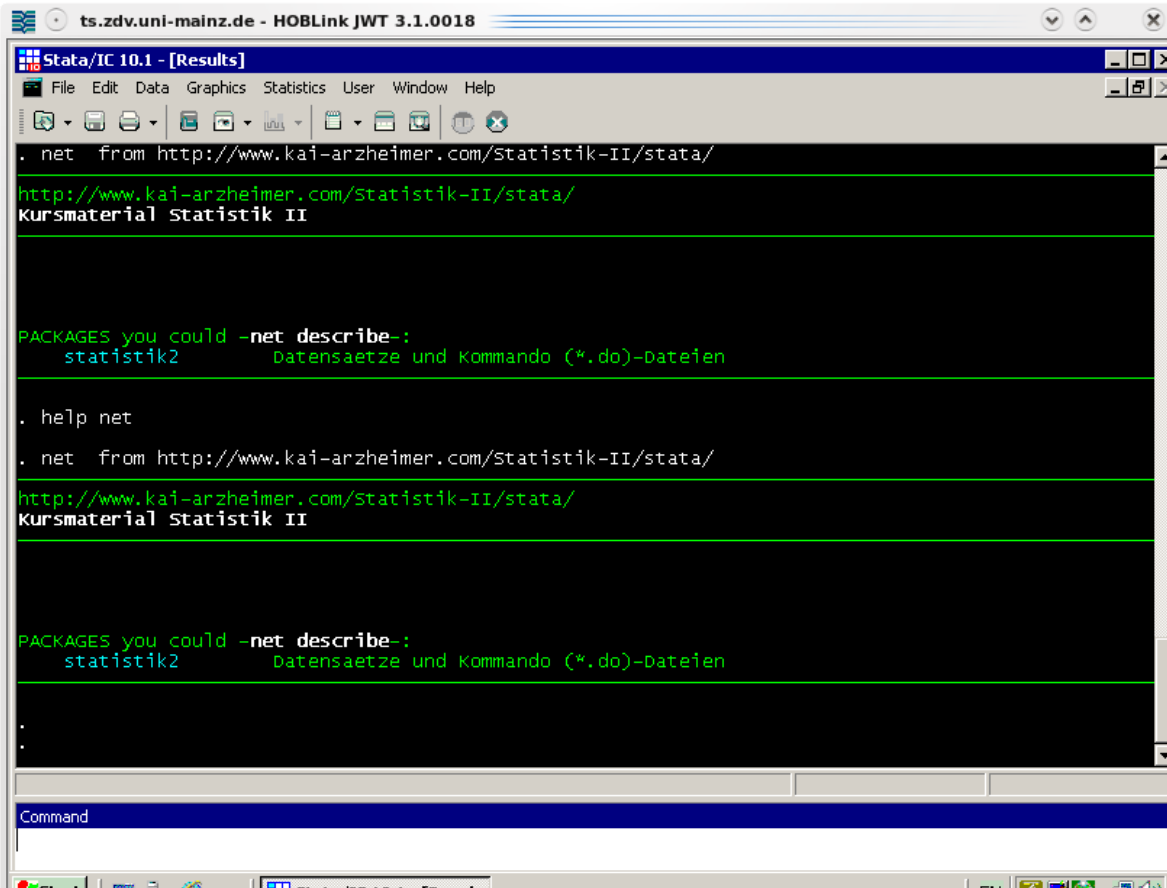
Statistik II Einführung (17/1)

Interaktivität

- ▶ Datensätze und Kommandodateien stehen über die Homepage zur Verfügung
- ▶ Zugriff ist aus Stata heraus möglich
- ▶ net from
`http://www.kai-arzheimer.com/Statistik-II/stata/`
- ▶ Per Klicken oder per net get allbus2008 kann das Paket allbus2008 geladen und lokal installiert werden
- ▶ Das Paket enthält eine Version des ALLBUS 2008 sowie eine *.do Datei mit Kommandos

Statistik II Einführung (18/1)

Paket zur Vorlesung



The screenshot shows a Stata/IC 10.1 window with the following content:

```
ts.zdv.uni-mainz.de - HOBLink JWT 3.1.0018
Stata/IC 10.1 - [Results]
File Edit Data Graphics Statistics User Window Help
. net from http://www.kai-arzheimer.com/Statistik-II/stata/
http://www.kai-arzheimer.com/Statistik-II/stata/
kursmaterial Statistik II

PACKAGES you could -net describe-:
  statistik2      Datensaeetze und Kommando (*.do)-Dateien

. help net
. net from http://www.kai-arzheimer.com/Statistik-II/stata/
http://www.kai-arzheimer.com/Statistik-II/stata/
kursmaterial Statistik II

PACKAGES you could -net describe-:
  statistik2      Datensaeetze und Kommando (*.do)-Dateien
.
```

Statistik II Einführung (19/1)

Interaktivität

- ▶ Datensätze und Kommandodateien stehen über die Homepage zur Verfügung
- ▶ Zugriff ist aus Stata heraus möglich
- ▶ `net from`
`http://www.kai-arzheimer.com/Statistik-II/stata/`
- ▶ Per Klicken oder per `net get allbus2008` kann das Paket `allbus2008` geladen und lokal installiert werden
- ▶ Das Paket enthält eine Version des ALLBUS 2008 sowie eine `*.do` Datei mit Kommandos

Statistik II Einführung (20/1)

Aufbau eines Stata-Befehls

- ▶ Viele optionale Komponenten, aber einfache Struktur

```
[by varlist:] command [varlist] [in range] [if exp]  
[filename], [options]
```

- ▶ `by...`: – Befehl für Subgruppen wiederholen
- ▶ `varlist` – Befehl auf diese Variablen anwenden
- ▶ `range` – Befehl auf Gruppe von Fällen anwenden (Reihenfolge)
- ▶ `if` – Befehl auf Gruppe von Fällen anwenden (logische Bedingung)
- ▶ `filename` – Dateinamen (oft mit `using` eingeleitet)
- ▶ `,` `options` – Optionen für Befehl festlegen

```
use allbus2008, clear
```

Was ist der ALLBUS 2008?

- ▶ „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften“ (+ISSP)
- ▶ „Goldstandard“
- ▶ Mehr-Themen-Befragung, seit 1980 alle zwei Jahre mit neuen Befragten wiederholt
- ▶ Regelmäßige Replikation einzelner Fragenblöcke
- ▶ Design?
- ▶ Themen 2008 u. a.
 - ▶ Politische Einstellungen, politische Partizipation, politische Entfremdung und Protest
 - ▶ Staatsbürgerschaft, Herkunft, Nationalstolz
 - ▶ Medien
 - ▶ Soziales Kapital und Freizeit
 - ▶ Wirtschaft

Häufigkeitstabellen

- ▶ Wie häufig kommen einzelne Kategorien einer Variablen vor?
- ▶ Z. B. Straftäter härter bestrafen?
- ▶ Wie findet man die passende Variable?
- ▶ describe
- ▶ Besser: lookfor straf
- ▶ Label = Etikett für interne Werte
 - ▶ Variablennamen
 - ▶ Numerische Werte von Variablen
- ▶ Gesuchte Variable: v26
- ▶ Häufigkeitstabelle: tabulate v26 oder tab v26

Statistik II Einführung (23/1)

Allbus Fragebogen

[F015 a]

⇒ Antwortschema zu Frage 15 : Aussagen bitte vorlesen! Liste 15 liegt vor!

	Stimme voll und ganz zu	Stimme eher zu	Weder noch	Stimme eher nicht zu	Stimme überhaupt nicht zu	Weiß nicht	KA
Einwanderer sollten verpflichtet werden, sich den deutschen Sitten und Gebräuchen anzupassen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Politik sollte sich aus der Wirtschaft heraushalten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zum Schutz der Umwelt sollten härtere Maßnahmen getroffen werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gleichgeschlechtliche Ehen sollten gesetzlich anerkannt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Frauen und Männer sollten bei Bewerbungen und Beförderungen gleich behandelt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Straftäter sollten härter bestraft werden als bisher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

⇒ Fortsetzung nächste Seite!

Stimme voll und ganz zu	Stimme eher zu	Weder noch	Stimme eher nicht zu	Stimme überhaupt nicht zu
-------------------------	----------------	------------	----------------------	---------------------------

Statistik II Einführung (24/1)

Häufigkeitstabellen

- ▶ Wie häufig kommen einzelne Kategorien einer Variablen vor?
- ▶ Z. B. Straftäter härter bestrafen?
- ▶ Wie findet man die passende Variable?
- ▶ describe
- ▶ Besser: lookfor straf
- ▶ Label = Etikett für interne Werte
 - ▶ Variablennamen
 - ▶ Numerische Werte von Variablen
- ▶ Gesuchte Variable: v26
- ▶ Häufigkeitstabelle: tabulate v26 oder tab v26

Statistik II Einführung (25/1)

describe

```
Stata/TC 10.1 - http://www.kai-arzheimer.com/Statistik-II/stata/allbus2008.dta - [Results]
Command
1 set mem 5m
2 use http://www.kai-arzheimer.com/... 198
3 use http://www.kai-arzheimer.com/... 199
4 list
5 labstat 100
6 tabstat v3
7 man labstat
8 tab v31
9 d v31
10 tab v31
11 tab v23
12 help lookfor
13 d
14 describe
15 help describe
16 describe

describe
-----
Contains data from http://www.kai-arzheimer.com/Statistik-II/stata/allbus2008.dta
obs: 3,469 ALLBUScompact 2008, Stata-Testversion
vars: 490 25 Aug 2009 16:25
size: 1,973,661 (62.4% of memory free)

storage display value
variable name type format label variable label
-----
v1 int %10.0g v1 STUDIENNUMMER
v2 int %10.0g v2 IDENTIFIKATIONSNUMMER DES BEFRAGTEN
v3 byte %10.0g v3 ERHEBUNGSGEBIET <WOHNGEBIET>: WEST - OST
v4 byte %10.0g v4 DEUTSCHE STAATSBÄNDER: BBT
v5 byte %10.0g v5 ATTRAKTIVITÄT DES BEFR., INTERVIEWSTART
v6 byte %10.0g v6 WIRTSCHAFTSLAGE IN DER BRD HEUTE
v7 byte %10.0g v7 WIRTSCHAFTSLAGE, BEFR. HEUTE
v8 byte %10.0g v8 WIRTSCHAFTSLAGE DER BRD IN 1 JAHR
v9 byte %10.0g v9 WIRTSCHAFTSLAGE, BEFR. IN 1 JAHR
v10 byte %10.0g v10 REGIERUNGSVERANTWORTUNG WIRTSCHAFTSLAGE
v11 double %10.0g v11 HÄUFIGKEIT VON FERNSEHEN PRO WOCHE
v12 int %10.0g v12 FERNSEHGESAMTDAUER PRO TAG IN MINUTEN
v13 byte %10.0g v13 FERNSEHGESAMTDAUER PRO TAG, KAT.
v14 byte %10.0g v14 NACHRICHTENKONSUM: OFFENTLICHES TV
v15 double %10.0g v15 KONSUMHÄUFIGK.: NACHRICHTEN OFF. TV
v16 byte %10.0g v16 NACHRICHTENKONSUM: PRIVATES TV
v17 double %10.0g v17 KONSUMHÄUFIGK.: NACHRICHTEN PRIVATES TV
v18 double %10.0g v18 HÄUFIGKEIT TAGESZEITUNG LESEN PRO WOCHE
v19 byte %10.0g v19 INTERNET FÜR INFORMATION ÜBER POLITIK?
v20 byte %10.0g v20 HÄUFIGKEIT INTERNET FÜR POLITIKFÖ
v21 byte %10.0g v21 EINWANDERER ZU ANPASSUNG VERPFLICHTEN?
v22 byte %10.0g v22 POLITIK AUS WIRTSCHAFT HERAUSHALTEN?
v23 byte %10.0g v23 HÄRTERE MASSNAHMEN FÜR UMWELTSCHUTZ?
v24 byte %10.0g v24 ANERKENNUNG GLEICHGESCHLECHTLICHER EHE?
v25 byte %10.0g v25 BERUF: GLEICHBEHANDLUNG V. MANN UND FRAU
v26 byte %10.0g v26 HÄRTERE STRAFEN FÜR STRAFTÄTER?
v27 byte %10.0g v27 REGIERUNG: SOZ. SICHERUNG IM WICHTIGSTEN?
v28 byte %10.0g v28 UNTERTEILUNG ZU GUNSTEN EINFACHER LEUTE?
v29 byte %10.0g v29 EINWANDERER GUT F. DEUTSCHE WIRTSCHAFT?
v30 byte %10.0g v30 DEUTSCHES MILITÄR F. KRIEG GEGEN TERROR
v31 unter allen Umständen Freiheit achten?
v1 unter allen Umständen Freiheit achten?

Variable List
-----
Name Label
v1 STUDIENNUMMER
v2 IDENTIFIKATIONSNUMMER DES BEFRAGTEN
v3 ERHEBUNGSGEBIET <WOHNGEBIET>: WEST - OST
v4 DEUTSCHE STAATSBÄNDER: BBT
v5 ATTRAKTIVITÄT DES BEFR., INTERVIEWSTART
v6 WIRTSCHAFTSLAGE IN DER BRD HEUTE
v7 WIRTSCHAFTSLAGE, BEFR. HEUTE
v8 WIRTSCHAFTSLAGE DER BRD IN 1 JAHR
v9 WIRTSCHAFTSLAGE, BEFR. IN 1 JAHR
v10 REGIERUNGSVERANTWORTUNG WIRTSCHAFTSLAGE
v11 HÄUFIGKEIT VON FERNSEHEN PRO WOCHE
v12 FERNSEHGESAMTDAUER PRO TAG IN MINUTEN
v13 FERNSEHGESAMTDAUER PRO TAG, KAT.
v14 NACHRICHTENKONSUM: OFFENTLICHES TV
v15 KONSUMHÄUFIGK.: NACHRICHTEN OFF. TV
v16 NACHRICHTENKONSUM: PRIVATES TV
v17 KONSUMHÄUFIGK.: NACHRICHTEN PRIVATES TV
v18 HÄUFIGKEIT TAGESZEITUNG LESEN PRO WOCHE
v19 INTERNET FÜR INFORMATION ÜBER POLITIK?
v20 HÄUFIGKEIT INTERNET FÜR POLITIKFÖ
v21 EINWANDERER ZU ANPASSUNG VERPFLICHTEN?
v22 POLITIK AUS WIRTSCHAFT HERAUSHALTEN?
v23 HÄRTERE MASSNAHMEN FÜR UMWELTSCHUTZ?
v24 ANERKENNUNG GLEICHGESCHLECHTLICHER EHE?
v25 BERUF: GLEICHBEHANDLUNG V. MANN UND FRAU
v26 HÄRTERE STRAFEN FÜR STRAFTÄTER?
v27 REGIERUNG: SOZ. SICHERUNG IM WICHTIGSTEN?
v28 UNTERTEILUNG ZU GUNSTEN EINFACHER LEUTE?
v29 EINWANDERER GUT F. DEUTSCHE WIRTSCHAFT?
v30 DEUTSCHES MILITÄR F. KRIEG GEGEN TERROR
v31 UNTER ALLEN UMSÄNDEN FREIHEIT ACHTEN?
v1 UNTER ALLEN UMSÄNDEN FREIHEIT ACHTEN?
```

Statistik II Einführung (26/1)

Häufigkeitstabellen

- ▶ Wie häufig kommen einzelne Kategorien einer Variablen vor?
- ▶ Z. B. Straftäter härter bestrafen?
- ▶ Wie findet man die passende Variable?
- ▶ describe
- ▶ Besser: lookfor straf
- ▶ Label = Etikett für interne Werte
 - ▶ Variablennamen
 - ▶ Numerische Werte von Variablen
- ▶ Gesuchte Variable: v26
- ▶ Häufigkeitstabelle: tabulate v26 oder tab v26

Statistik II Einführung (27/1)

lookfor straf

```
Stata/TC 10.1 - http://www.kai-arzheimer.com/Statistik-II/stata/allbus2008.dta - [Results]
Command
1 set mem 5m
2 use http://www.kai-arzheimer.com/... 198
3 use http://www.kai-arzheimer.com/... 199
4 list
5 labstat 100
6 tabstat
7 man tabstat
8 tab v31
9 d v31
10 tab v31
11 tab v26
12 help lookfor
13 d
14 describe
15 help describe
16 desc v26
17 lookfor straf

Results
v47 byte %10.0g v47 HABE AN OFFENTL.DISKUSSSIONEN TEILGEN.
v48 byte %10.0g v48 HABE IN BÜRGERINITIATIVE MITGEARBEITET
v49 byte %10.0g v49 HABE SCHON IN PARTEI MITGEARBEITET
v50 byte %10.0g v50 HABE AN UNGENEHMIGTER DEMO TEILGENOMMEN
v51 byte %10.0g v51 HABE AN GENEHMIGTER DEMO TEILGENOMMEN
v52 byte %10.0g v52 HABE MICH NICHT AN WAHLEN BETEILIGT
v53 byte %10.0g v53 HABE ANDERE PARTEI GEWAHLT
v54 byte %10.0g v54 HABE AN UNTERSCHRIFTENSAMMLUNG TEILGEN.
v55 byte %10.0g v55 HABE KRITISCH KONSUMIERT
v56 byte %10.0g v56 HABE AN ONLINE-PROTESTAKTION TEILGEN.
v57 byte %10.0g v57 2. JAHR: POLITISCHE MEINUNG GESAGT
v58 byte %10.0g v58 2. JAHR: AN WAHLEN BETEILIGT
v59 byte %10.0g v59 2. JAHR: AN OFFENTL.DISKUSSSIONEN BET.
v60 byte %10.0g v60 2. JAHR: IN BÜRGERINITIATIVE MITGEARB.
v61 byte %10.0g v61 2. JAHR: SCHON IN PARTEI MITGEARBEITET
v62 byte %10.0g v62 2. JAHR: AN UNGENEHMIGTER DEMO TEILG.
v63 byte %10.0g v63 2. JAHR: AN GENEHMIGTER DEMO TEILGEN.
v64 byte %10.0g v64 2. JAHR: MICH NICHT AN WAHL BETEILIGT
v65 byte %10.0g v65 2. JAHR: ANDERE PARTEI GEWAHLT
v66 byte %10.0g v66 2. JAHR: UNTERSCHRIFTENSAMMLUNG
v67 byte %10.0g v67 2. JAHR: KRITISCH KONSUMIERT
v68 byte %10.0g v68 2. JAHR: ONLINE-PROTESTAKTION
v69 byte %10.0g v69 PARTEIPREFERENZ: VORHANDEN
v70 byte %10.0g v70 PARTEIPREFERENZ: BEFR.
v71 byte %10.0g v71 VERTRAUEN: GESUNDHEITSWESEN
v72 byte %10.0g v72 VERTRAUEN: BUNDESVERFASSUNGSGERICHT
v73 byte %10.0g v73 VERTRAUEN: BUNDESTAG
v74 byte %10.0g v74 VERTRAUEN: STADT-,GEMEINDEVERWALTUNG
v75 byte %10.0g v75 VERTRAUEN: JUSTIZ
v76 byte %10.0g v76 VERTRAUEN: FERNSEHEN
v77 byte %10.0g v77 VERTRAUEN: ZEITUNGSWESEN
v78 byte %10.0g v78 VERTRAUEN: HOCHSCHULEN,UNIVERSITAETEN
v79 byte %10.0g v79 VERTRAUEN: BUNDESREGIERUNG
v80 byte %10.0g v80 VERTRAUEN: POLIZEI
v81 byte %10.0g v81 VERTRAUEN: POLITISCHE PARTEIEN
v82 byte %10.0g v82 VERTRAUEN: KOMMISSION DER EU
v83 byte %10.0g v83 VERTRAUEN: EUROPÄISCHES PARLAMENT

lookfor straf
storage display value
variable name type format label variable label
v26 byte %10.0g v26 HÄRTERE STRAFEN FÜR STRAFTAETER?
```

Statistik II Einführung (28/1)

Häufigkeitstabellen

- ▶ Wie häufig kommen einzelne Kategorien einer Variablen vor?
- ▶ Z. B. Straftäter härter bestrafen?
- ▶ Wie findet man die passende Variable?
- ▶ describe
- ▶ Besser: lookfor straf
- ▶ Label = Etikett für interne Werte
 - ▶ Variablennamen
 - ▶ Numerische Werte von Variablen
- ▶ Gesuchte Variable: v26
- ▶ Häufigkeitstabelle: tabulate v26 oder tab v26

Statistik II Einführung (29/1)

Häufigkeiten: Härtere Strafen

```
. tab v26
```

HAERTERE STRAFEN FUER STRAFTAETER?	Freq.	Percent	Cum.
STIMME VOLL ZU	1,882	54.25	54.25
STIMME EHER ZU	781	22.51	76.77
WEDER NOCH	429	12.37	89.13
STIMME EHER NICHT ZU	210	6.05	95.19
STIMME GAR NICHT ZU	91	2.62	97.81
WEISS NICHT	62	1.79	99.60
KEINE ANGABE	14	0.40	100.00
Total	3,469	100.00	

```
. lab list v26
```

```
v26:
```

1. STIMME VOLL ZU
2. STIMME EHER ZU
3. WEDER NOCH
4. STIMME EHER NICHT ZU
5. STIMME GAR NICHT ZU
8. WEISS NICHT
9. KEINE ANGABE

```
. tab v26 if v26<8
```

HAERTERE STRAFEN FUER STRAFTAETER?	Freq.	Percent	Cum.
1. STIMME VOLL ZU	1,882	55.47	55.47
2. STIMME EHER ZU	781	23.02	78.49
3. WEDER NOCH	429	12.64	91.13
4. STIMME EHER NICHT ZU	210	6.19	97.32

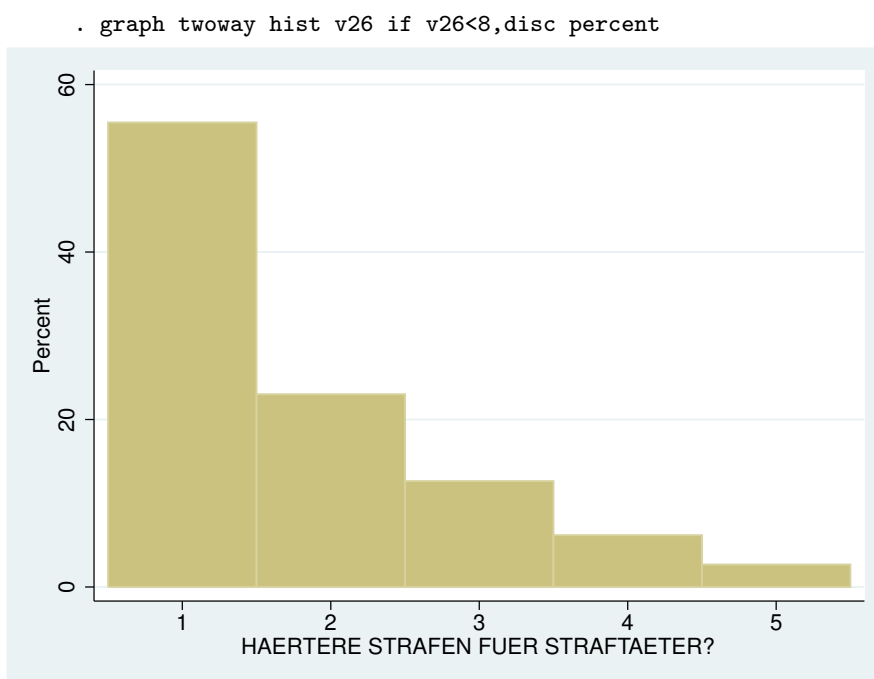
Statistik II Einführung (30/1)

Total 3,393 100.00

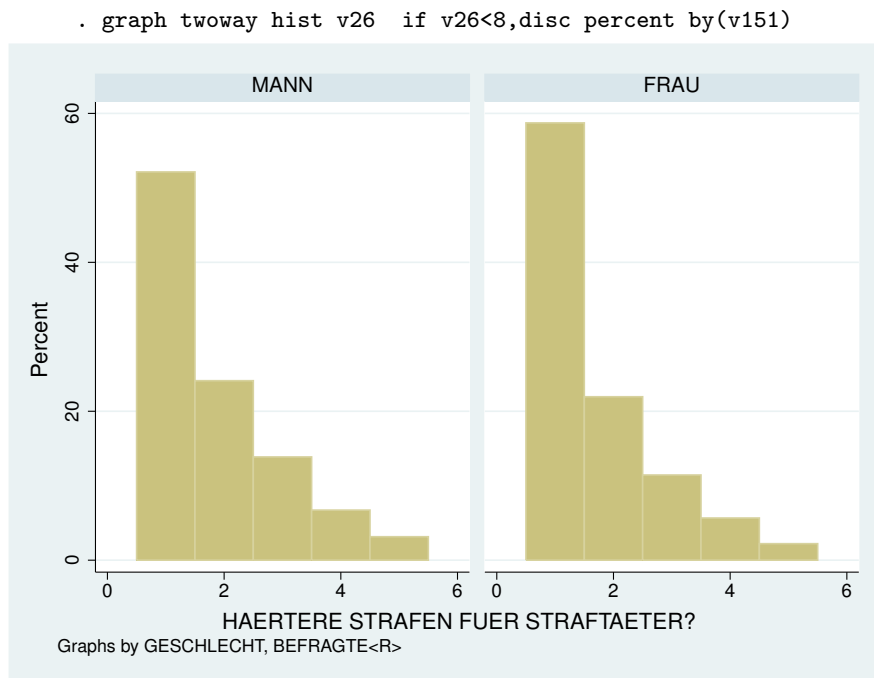
Grafikbefehle

- ▶ Aus historischen Gründen verschiedene Interfaces
- ▶ Teils sehr komplexe Optionen
- ▶ Handbücher, Kohler/Kreuter, Visual Guide
- ▶ Kategoriale Variable, eine Dimension – warum Histogramm eigentlich nicht geeignet?

Barplot/Histogramm



Barplot/Histogramm nach Geschlecht



Statistik II Einführung (33/1)

„Echte“ Balkendiagramme

- ▶ Erfordern einen Trick:
- ▶ Kategoriale Variablen müssen in eine Serie binärer Variablen (0/1 kodiert = „Dummies“) zerlegt werden
- ▶ Jede Variable mit k Kategorien kann durch $k - 1$ Dummies ersetzt werden (andere Kodierungen möglich)
- ▶ Beispiel Konfession („katholisch“, „protestantisch“, „andere“)
 - ▶ Dummies für katholisch/protestantisch (andere= weder katholisch noch protestantisch)
 - ▶ Dummies für andere/protestantisch (katholisch= weder andere noch protestantisch)
 - ▶ Dummies für andere/katholisch (protestantisch= weder andere noch katholisch)
- ▶ Dritter Dummy redundant (wichtig für Regression)

Statistik II Einführung (34/1)

Wie generiert man Dummies?

- Befehle um Variablen zu kopieren, erzeugen, verändern

```
. generate dummy1=0
. replace dummy1=1 if v26==1
(1882 real changes made)
```

```
. generate dummy2=v26==2
. tab dummy1 dummy2
```

dummy1	dummy2		Total
	0	1	
0	806	781	1,587
1	1,882	0	1,882
Total	2,688	781	3,469

```
. tab v26 dummy1
```

HAERTERE STRAFEN FUER STRAFTAETER?	dummy1		Total
	0	1	
STIMME VOLL ZU	0	1,882	1,882
STIMME EHER ZU	781	0	781
WEDER NOCH	429	0	429
STIMME EHER NICHT ZU	210	0	210
STIMME GAR NICHT ZU	91	0	91
WEISS NICHT	62	0	62
KEINE ANGABE	14	0	14
Total	1,587	1,882	3,469

v26 umkodieren/fehlende Werte löschen?

- Werte größer 5 sollten automatisch ignoriert werden (missing)
- Hohe Werte = Ablehnung: kontraintuitiv

```
. gen strafe = v26 if v26<8
(76 missing values generated)
```

```
. tab strafe
```

strafe	Freq.	Percent	Cum.
1	1,882	55.47	55.47
2	781	23.02	78.49
3	429	12.64	91.13
4	210	6.19	97.32
5	91	2.68	100.00
Total	3,393	100.00	

```
. replace strafe = 6 - strafe
(2964 real changes made)
```

```
. tab strafe
```

strafe	Freq.	Percent	Cum.
1	91	2.68	2.68
2	210	6.19	8.87
3	429	12.64	21.51
4	781	23.02	44.53
5	1,882	55.47	100.00

Aus der neuen Variable fünf Dummies erzeugen

- ▶ Mehrere Hilfsmittel
- ▶ Sehr schnell:

```
. tab strafe,gen(mehrstraf)
```

strafe	Freq.	Percent	Cum.
1	1,882	55.47	55.47
2	781	23.02	78.49
3	429	12.64	91.13
4	210	6.19	97.32
5	91	2.68	100.00
Total	3,393	100.00	

```
. d mehrst*
```

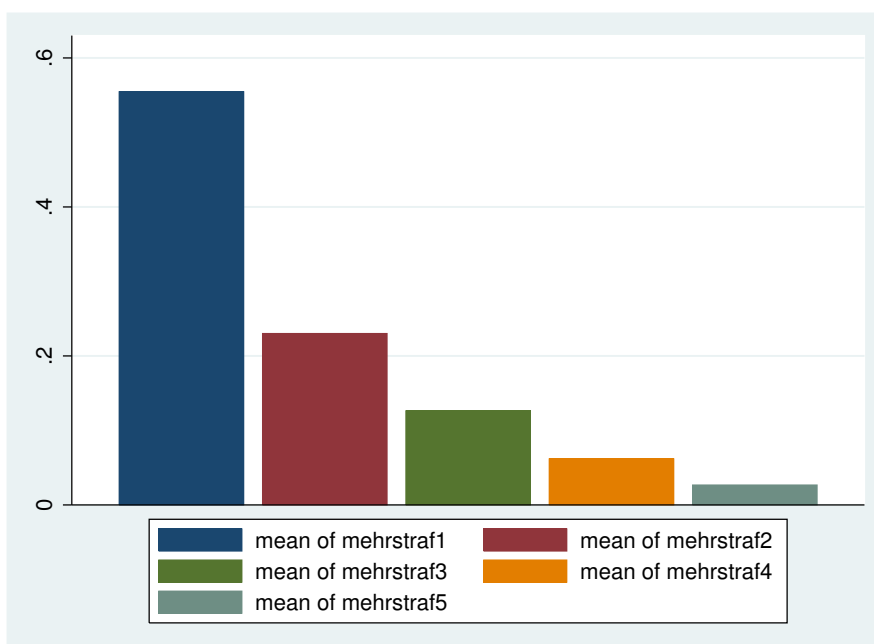
variable name	storage type	display format	value label	variable label
mehrstraf1	byte	%8.0g		strafe== 1.0000
mehrstraf2	byte	%8.0g		strafe== 2.0000
mehrstraf3	byte	%8.0g		strafe== 3.0000
mehrstraf4	byte	%8.0g		strafe== 4.0000
mehrstraf5	byte	%8.0g		strafe== 5.0000

```
. summ mehrstraf*
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mehrstraf1	3393	.5546714	.4970753	0	1
mehrstraf2	3393	.2301798	.4210098	0	1
mehrstraf3	3393	.1264368	.33239	0	1
mehrstraf4	3393	.0618921	.240995	0	1
mehrstraf5	3393	.0268199	.1615807	0	1

Statistik II Einführung (37/1)

Echtes Balkendiagramm

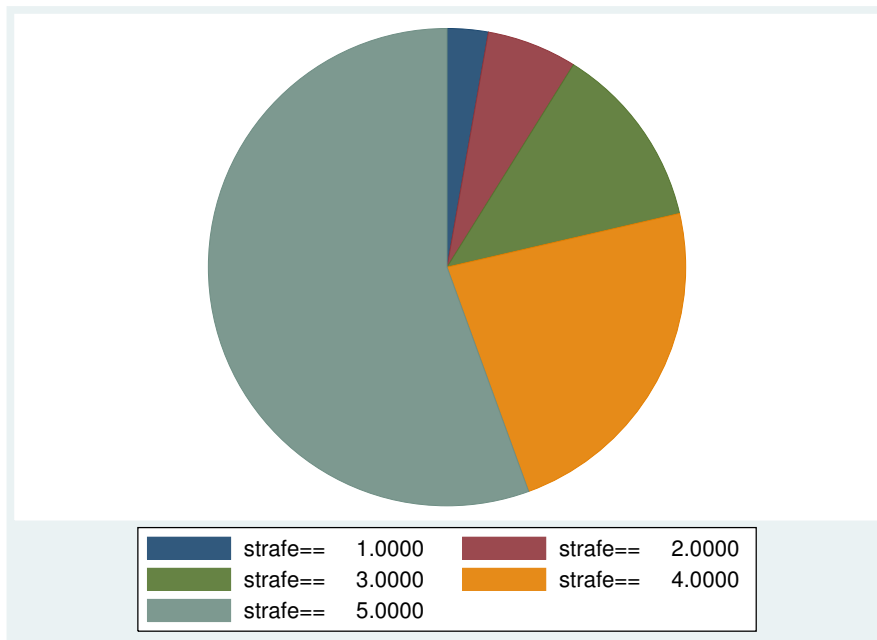


Statistik II Einführung (38/1)

Tortendiagramme...

- ▶ ...sind unerwünscht
- ▶ ...können mit dem gleichen Trick (Dummies) erzeugt werden

```
. graph pie mehrstraf*
```



Statistik II Einführung (39/1)

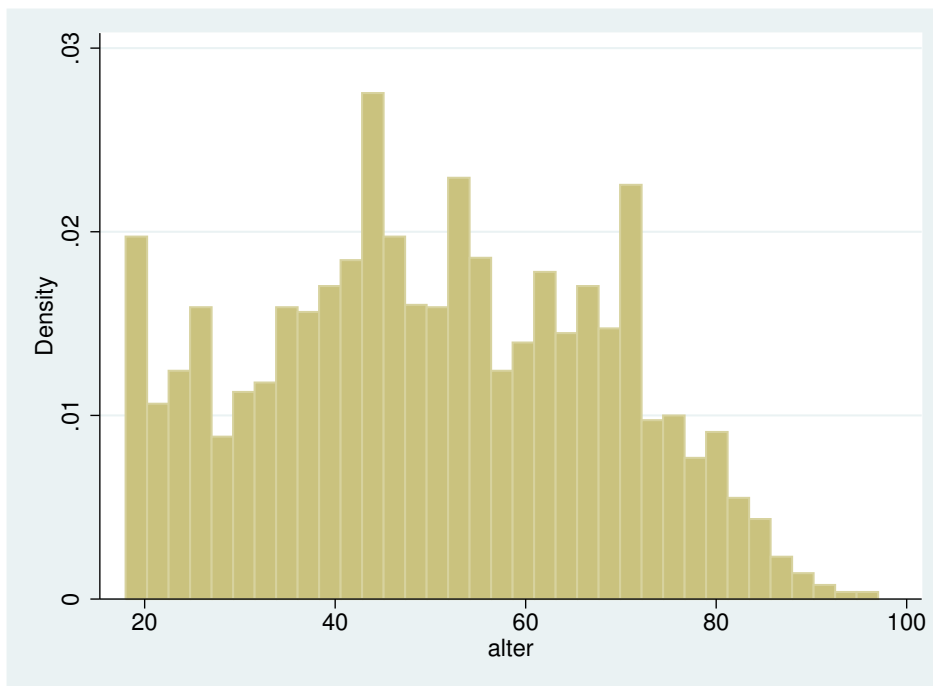
Histogramm und Dichteschätzung

- ▶ (Konzeptuell) kontinuierliche Variablen
- ▶ Z. B. Alter (18-999)
- ▶ Alter > 120 auf missing setzen

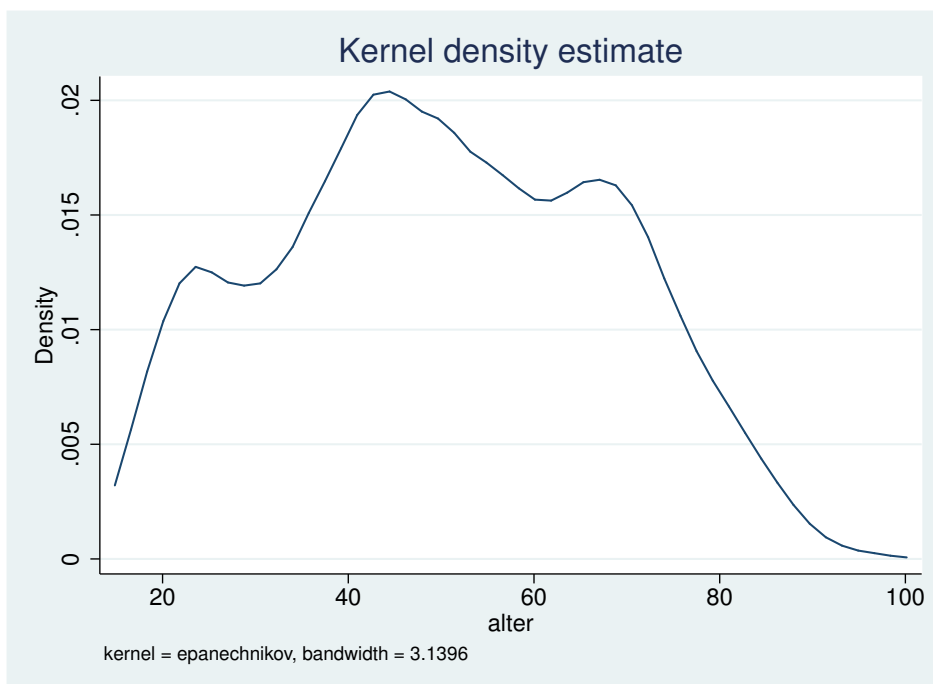
```
. gen alter = v154 if v154<121  
(12 missing values generated)  
. hist alter  
(bin=35, start=18, width=2.2571429)  
. kdensity alter  
. kdensity alter,by(v151)  
option by() not allowed  
r(191);  
. graph twoway kdensity alter,by(v151)
```

Statistik II Einführung (40/1)

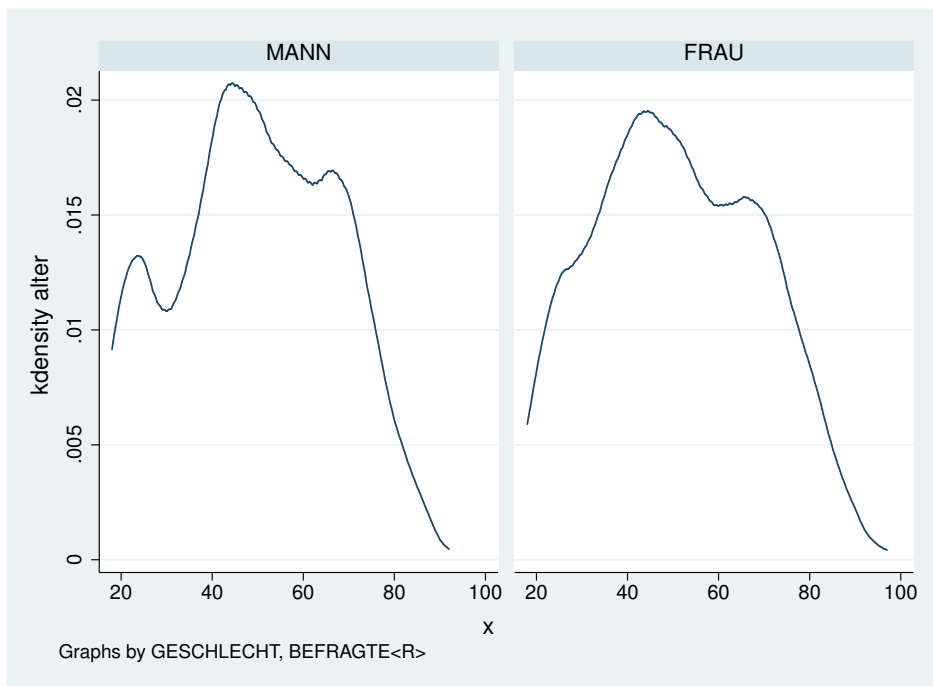
Alter: Histogramm



Alter: Dichteschätzung



Alter: Dichteschätzung nach Geschlecht



Statistik II Einführung (43/1)

Vergleich mit theoretischer Verteilung

- ▶ Ist das Alter normalverteilt?
- ▶ Normalverteilung (Dichte)
 - ▶ Symmetrisch
 - ▶ Glockenförmig
 - ▶ Ca. 90% der Fläche ± 1.64 Standardabweichungen vom Mittelwert; 95% der Fläche ± 1.96 Standardabweichungen
- ▶ Kumulierte Normalverteilung: Integral der Dichteverteilung
- ▶ Wichtiges Modell für viele reale und theoretische Verteilungen
- ▶ Dichte und kumulierte Verteilungen sind *Funktionen*
- ▶ x - Werten zwischen $\pm\infty$ werden y -Werte zugeordnet
- ▶ Stata kann diese und viele andere Funktionswerte bestimmen, plotten, ausgeben

Statistik II Einführung (44/1)

Normalverteilung

- ▶ Wieviel Prozent der Verteilung liegen links der Werte 1, 2, 3?
- ▶ Welche Werte trennen von links 2.5%, 10% und 50% der (Standard)Normalverteilung ab?

```
. display invnormal(.025)
-1.959964
. display invnormal(.10)
-1.2815516
. display invnormal(.50)
0
```

```
. display normal(1)
.84134475
. display normal(2)
.97724987
. display normal(3)
.9986501
```

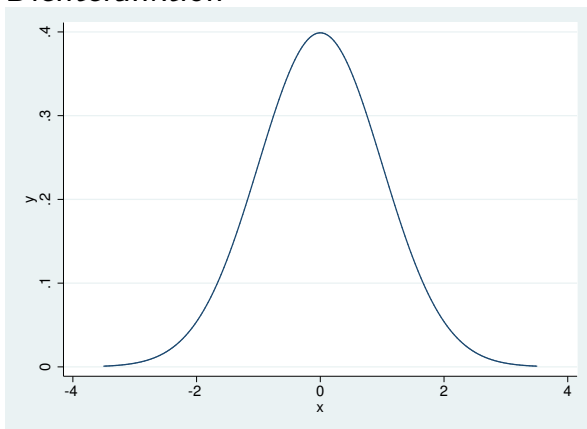
```
. display 2+2
4
. display "Hallo!"
Hallo!
```

- ▶ `normal`: *kumulierte* Normalverteilung
- ▶ `invnormal`: *Umkehrfunktion* zur kumulierten Normalverteilung
- ▶ `normpdf`: *Dichtefunktion* der Normalverteilung

Statistik II Einführung (45/1)

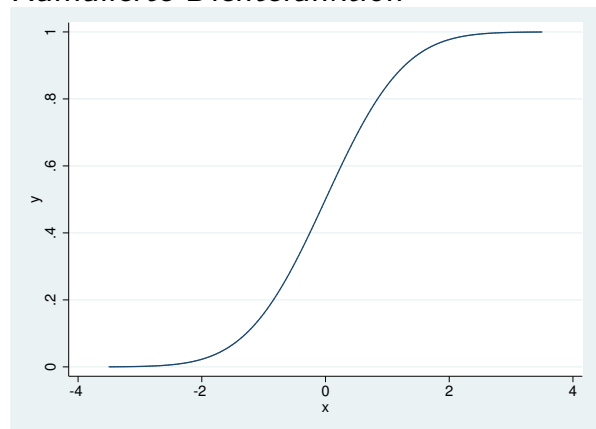
Normalverteilung graphisch

Dichtefunktion



```
. graph twoway function y=normden(x),range(-3.5 3.5)
```

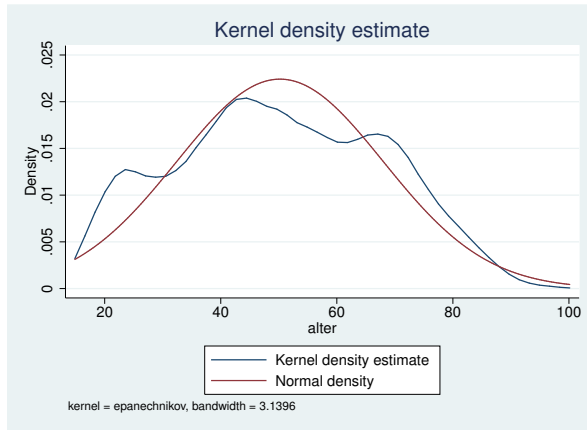
Kumulierte Dichtefunktion



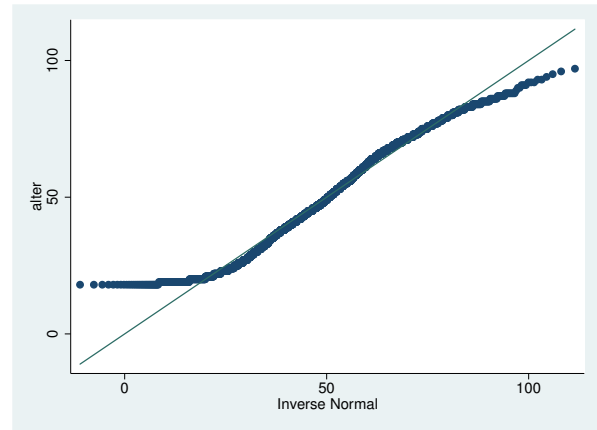
```
. graph twoway function y=normal(x),range(-3.5 3.5)
```

Alter und Normalverteilung

- ▶ Normalverteilung mit passendem Mittelwert/Varianz über Altersverteilung plotten
- ▶ Bequem: Option nutzen; genauer: Quantil-Plot



```
. kdensity alter,normal
```



```
. qnorm alter
```

Zwei kontinuierliche Variablen: Scatterplot

- ▶ Viele Variablen nur konzeptuell kontinuierlich
- ▶ Wenige Kategorien → Punkte im Plot klumpen
- ▶ Beispiel Alter (ok) vs. Attraktivität auf den Interviewer (1-11)
- ▶ „jitter“, „random noise“

```
. graph twoway scatter v5 alter,jitter(8) by(v151)  
. graph export scatter-2.eps  
(file scatter-2.eps written in EPS format)
```

- ▶ Software essentiell für die Verwaltung, Verarbeitung und Auswertung von Daten
- ▶ Stata Programm der Wahl für die Ausbildung in der Politikwissenschaft
- ▶ Alles, was in Statistik I und hier gelernt wurde/wird, mit Stata machbar
- ▶ Wichtig: **Üben**