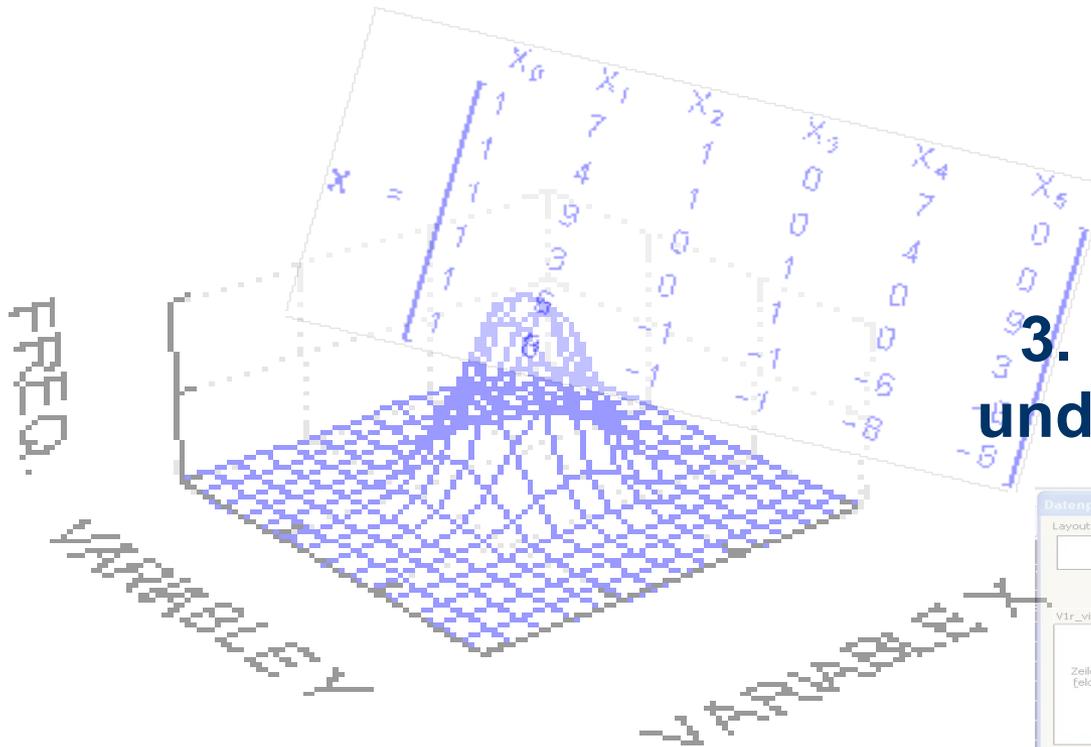
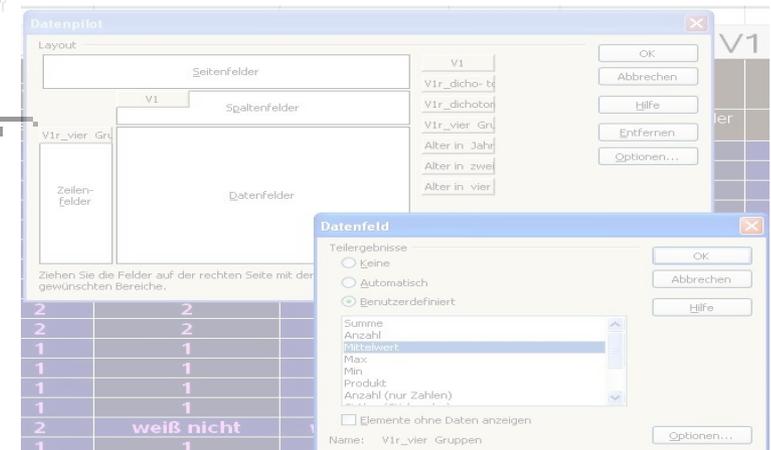


Statistik IV

3. Datenaufbereitung II und erste Auswertungen

Datenpilot

Layout:

- Seitenfelder: V1
- Spaltenfelder: V1
- Datenfelder: V1_vier Gr

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

Datenfeld

Teilergebnisse:

- Keine
- Automatisch
- Benutzerdefiniert

Summe
Anzahl
Mittelwert
Max
Min
Produkt
Anzahl (nur Zahlen)

Elemente ohne Daten anzeigen

Name: V1_vier Gruppen

Terminplanung

<i>Nr.</i>	<i>Termin</i>	<i>Inhalt</i>
1	14.04.09	Einführung • Organisatorisches und Scheinvoraussetzungen • Statistik mit Softwareunterstützung?
2	28.04.09	Grundlagen Stata • 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

- **Wiederholung**
- **Hausaufgabe 1**
- **Datenaufbereitung II**
- **erste Analysen**

STATA

memorize

`summarize v356`

`lookfor Alter`

MZS

Hausaufgabe 1

Ermittlung der Codes in Stata

- `numlabel v155, add` /* Codes bei tab anzeigen*/
- `numlabel v155, remove` /* Anzeige aufheben */
- `numlabel all, add` /* alle Codes bei tab anzeigen */

Alternative

- `label list v155`
/* Codes und Wertelabels direkt anzeigen */

Datentransformation II

Clonen und Umbenennen von Variablen

- Clonen einer Variable
 - `clonevar AlterK=v155`
- Umbenennen einer Variable
 - `rename AlterK Alter2`
- `tab Alter2`

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 der Variablen beruht.
- Vergeben Sie Labels für die Variablen und Wertelabels für die letzte Variable

Do-Files

Kommentare im Do-File

- ********* Zeilen mit Sternen am Beginn sind Kommentare
- **/*** Text zwischen wird als Kommentar interpretiert ***/**
- **//** Text hinter einem Doppelschrägstrich wird als Kommentar wahrgenommen
- **///** beginnt einen Kommentar, der den nächsten Umbruch mit einbezieht
 - Letzteres kann dazu genutzt werden, lange Befehle über mehrere Zeilen umzubrechen

Arbeiten mit Do-Files

- Grundstruktur

```
*** Kommentar zum DO-File
```

```
version 10.1
```

```
set more off /* Do läuft ohne Unterbrechung */
```

```
(.....) /* Bereich für Arbeitsbefehle */
```

```
exit
```

Neue Daten anlegen und Datensätze kombinieren

Daten eingeben

- **clear** /* leert den Datenbereich */
- **edit** /* startet das Eingabefenster */
- **input ID vx2 vx3** /* ermöglicht die Dateneingabe per Kommandozeile */
- Werte eingeben
 - 1 33 9
 - 2 19 1
 - 3 28 1
 - 4 45 2
 - 5 23 1
 - end

Stringvariablen

- Stringvariablen anlegen
 - `input str20 name`
 - Karl-Heinz
 - Dr. Edeltraut
 - Sven
 - Annika
 - Heinz-Hugo
- `replace name = "Dr. Sven" in 3`

Aufgabe 6

- Stringvariable zur Partei anlegen und Parteinamen für die 5 Fälle eingeben.
 - SPD, FDP, Grüne, SPD, CDU

Stringfunktionen

- `encode Partei, gen(party_n)`
/*erzeugt numerische Variablen in alphabetischer Reihenfolge */
- `gen spd = upper(Partei) == "SPD"`
/* erzeugt eine dichotome Variable */
- `gen Heinz=strpos(name, "Heinz") > 0`
/* erzeugt aus einem Substring eine dichotome Variable */

Aufgabe 7

- Erzeugen Sie für eine dichotome Variable für CDU und eine für Personen mit Dr.-Titel.
- Speichern Sie den Datensatz.

Daten einlesen

- Stata besitzt keinen Importfilter für Daten aus anderen Programmen
- Stata kann ASCII-Dateien (American Standard Code for Information Interchange) lesen.
 - `insheet using "V: (...) \DATEN\ASCII.txt", delimiter(";")`

Datensätze zusammenfügen und Variablen auswählen

- `merge ID using "C: (...) \DATEN\EINGABE.dta", unique sort`
- `drop vx3`
- `keep ID v2 v3 v4 name Partei spd Heinz`

Datenauswertungen

Tabulate

■ Univariate Tabellen

- `tab v93`
- `tab1 v93-v95, mis`

■ Bivariat Tabellen

- `tab v93 v3, exp chi2 taub gamma`
- `tab v93 v3, col row`
- `tab2 v3 v93 v94`

Statistische Kennwerte

- `tabstat v93` /* liefert den Mittelwert */
- `tabstat v93, stats(mean sum max min range median q sd variance)`

Hausaufgabe 2

- Erstellen Sie ausgehend vom Codebook des ALLBUS' 2008 einen kommentierten do-File, der eine interessante Auswertung liefert, die Sie in der nächsten Sitzung vorstellen können.