Introduction to Stata

Svend Juul

ISHR refers to the book: Svend Juul: Introduction to Stata for Health Researchers. College Station, TX: Stata Press 2006.

A questionnaire:

Sex:
Male 1
Female 2

Born:
Day:_____ Month:_____ Year:_____

Education completed:
Physician 1
Dentist 2
Other 3

Education completed when?
Month:_____ Year:_____

How many children do you have?

2

Sex:
Male 1
Female 2

Born:
Day: ____ Month: ____ Year: 19____

Education completed:
Physician 1
Dentist 2
Other 3 *(write)*

Education completed when?
Month: ____ Year: 19____

How many children do you have?

```
01107121961101198 93
02108121963101199 01
...
```

3

In this case, the content of the questionnaires were entered in the computer as ASCII data lines in fixed formats. ISHR chapter 6 tells about this and other methods for entering data.

A codebook:

Variable name	Variable definition	Coding (range)	Digits	Positions
id	ID number	1-99	2	1-2
sex	Sex of respondent	1 male 2 female	1	3
bday **bmon** **byear**	Born day Born month Born year	1-31 1-12 1930-1982	2 2 4	4-5 6-7 8-11
educ	Education	1 physician 2 dentist 3 other	1	12
emon **eyear**	Education completed month year	 1-12 1970-2002	 2 4	 13-14 15-18
kids	Number of children	0-8 9* unknown	1	19

4

The codebook documents the relationship between the information in the data lines (previous slide) and the final Stata dataset.

ISHR section 18.3.

Rectangular data set.

	Variables							
	id	**sex**	**age**					
Obser-vations	**1**	**M**	**37**					
	2	**M**	**21**					
	3	**F**	**45**					

5

Derived variables:

Variable name	Variable definition	Coding etc.
eage	Age by completion of education	Calculate from **bmon byear emon eyear**
eagegr	5 year age groups	Calculate from **eage**: 1 20-24 2 25-29 etc.

6

Types of variables (scales):

Interval scale	Age in years Weight in kg etc.
Rank (ordinal) scale	lean / average / obese
Nominal scale	Danish / Norwegian / Swedish / Other
Dichotomous (binary) scale	male / female -44 / 45+

7

On variable types in Stata: ISHR chapter 5.

Use numerical codes:

Information	**Code**
Body weight is 65 kg	"65"
Sex is male	"1"
Body weight is unknown	"999"

"999" is a **missing value**.

8

You instruct Stata with commands

- Type a command in the Command Window
- Use menus and dialogs to create a command
- Type commands in a do-file – to be executed later.

9

ISHR section 1.5

Stata is case-sensitive

- Command names are lowercase. **`list`** is a valid command name; **`List`** is not.
- Variable names usually are lowercase, and **`SEX`**, **`Sex`**, and **`sex`** are different variable names.
- Variable names start with a letter. Avoid > 10 characters. No special characters except _ (underscore). No **`æ ø å`**.
- **`sex køn nation var47 47v a a2.7 a2_7`**

10

ISHR section 4.1

```
. use D:\StataCourse\smoke.dta
. summarize
    Variable |  Obs      Mean  Std. Dev.  Min   Max
-------------+--------------------------------------
          id |  230     115.5  66.53946     1   230
         sex |  230   1.73913  .4400666     1     2
         age |  230   5.80435  14.27799    21    84
      weight |  227  64.08811  11.85798    43   110
      height |  227  166.9736  8.370788   150   194
      smoker |  230  .9217391  .7435055     0     2
     cigaret |  230  4.730435  7.186015     0    40
     cheroot |  230  .2086957  1.105727     0    10
        pipe |  230  .0347826  .2262446     0     2
```

11

The `summarize` command: ISHR section 10.1

Anatomy of commands

[*prefix*:]	*command*	[*varlist*]	[*qualifiers*]	[, *options*]
	summarize			
	summarize	**_all**		
	sum	**sex age**		
	summarize	**sex-weight**		
	summarize	**pro***		
	sort	**sex**		
by sex:	**summarize**	**weight**		
	summarize	**weight**	**if sex==1**	
	summarize		**in 1/10**	
	summarize			**, detail**

12

ISHR section 4.2

Calculations

```
generate bmi = weight/(height^2)

replace bmi = (weight-1)/(height^2) if clothes==1

egen hrqol = rowsum(q1-q10)

recode opage (35/120=3)(15/35=2)(0/15=1) , ///
  generate(opagr)
```

13

Calculation commands: ISHR chapter 8

`generate` section 8.1

`replace` section 8.1

Operators and functions section 8.2

`egen` section 8.3

`recode` section 8.4

Job creating the Stata data file **educ.dta** from **educ.txt**:

```
// gen_educ.do

cd "C:\docs\educ"

infix id 1-2 sex 3 bday 4-5       ///
  bmon 6-7 byear 8-11 educ 12     ///
  emon 13-14 eyear 15-18 kids 19  ///
  using "educ.txt"

// VARIABLE LABELS
label variable id "ID number"
label variable bday "Day of birth"
...

// VALUE LABELS
label define sexlbl  1 male  2 female
label values sex sexlbl
...

// MISSING VALUES
recode kids  9=.

save "educ.dta"
```

14

Do-files ISHR section 1.5

`cd` section 6.1

`infix` section 6.3

`label` section 7.1

Missing values section 5.3

`save` section 6.1

Once you have the Stata data file, you can easily create tables:

```
use c:\docs\educ\educ.dta
oneway kids educ , tabulate
```

```
. oneway kids educ , tabulate
```

Education	Mean	SD	Valid N
physician	1.6	0.8	30
dentist	3.4	1.6	11
other	2.0	0.9	14
Total	2.1	1.0	55

15

`use` section 6.1

`oneway` section 10.4

```
. tab2 educ sex
```

	Sex		
Education	male	female	Total
physician	15	15	30
dentist	4	7	11
other	5	9	14
Total	24	31	55

16

`tab2` section 10.3

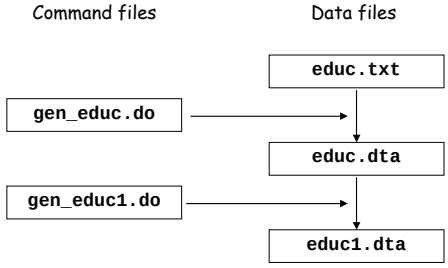
Next generation Stata dataset:

```
// gen_educ1.do

cd "c:\docs\educ"
use "educ.dta" [, clear]

generate bdate=mdy(bmon,bday,byear)
generate edate=mdy(emon,15,eyear)
generate eage=(edate-bdate)/365.25
recode eage (30/100=3)(25/30=2)(0/25=1) ,   ///
  generate(eagegr)

label variable eage "Age at completed education"
label variable eagegr   ///
  "Age group at completed education"
label define eagegr  1 "-24"  2 "25-29"  3 "30+"
label values eagegr eagegr

save "educ1.dta" [, replace]
```

17

`cd, use, save`	section 6.1
`generate`	section 8.1
Date variables	section 5.5
`label`	section 7.1

Which files did we use?



18

My suggested principle for giving name to a do-file that generates a new version of the dataset (the **gen_** prefix): ISHR section 18.4.

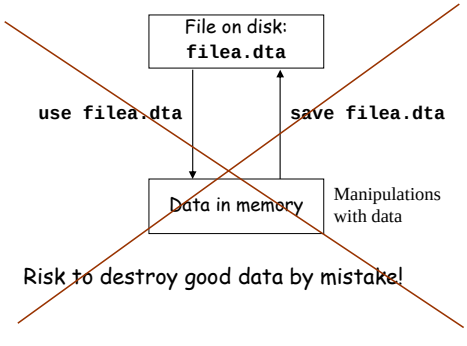
Types of files

educ.txt	Simple data file (ASCII file)
educ.dta	Stata dataset
gen_educ.do	Command file (do-file)

19

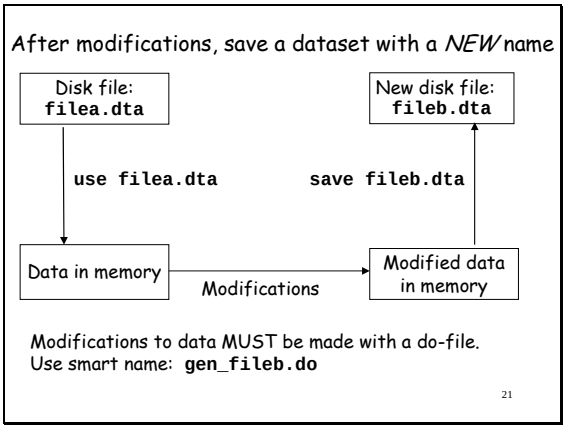
File types: ISHR chapter 3.

Let Windows display file name extensions: ISHR appendix B

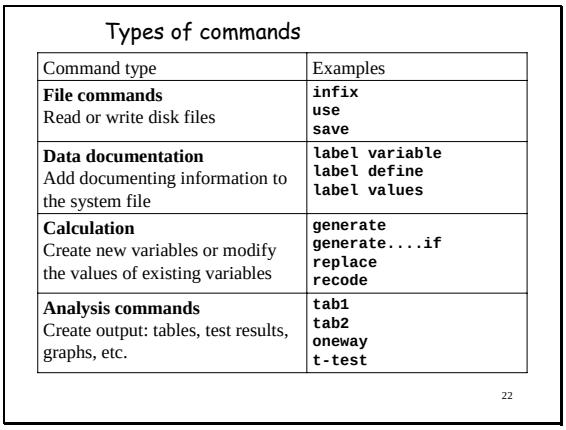


20

ISHR section 6.1



ISHR section 6.1



File commands: ISHR chapter 6

Documentation commands: ISHR chapter 7

Calculation commands: ISHR chapter 8

Simple analysis commands: ISHR chapter 10