

A very short introduction to Stata

Svend Juul, 15 June 2011

1

Rectangular data set

	Variables						
	id	sex	age				
Observations	1	1	37				
	2	2	21				
	3	2	.				
	4	1	37				

2

Use numerical codes:

Information	Code
Body weight is 65 kg	"65"
Sex is male	"1"
Body weight is unknown	"."
Sex is unknown	"9"

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

3

You instruct Stata by commands

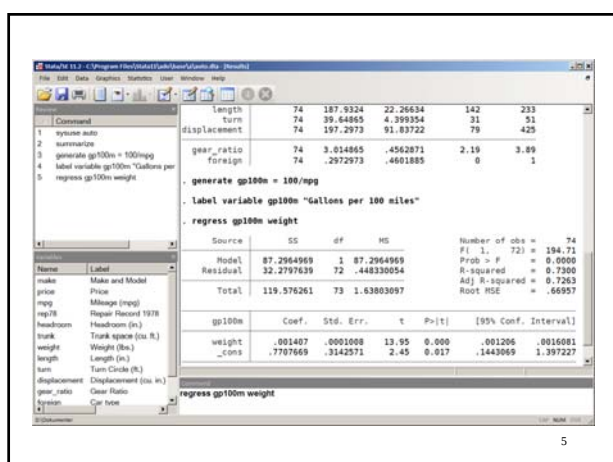
The only way (almost) to give instructions to Stata is by commands.

There are menus and dialogs, but only as a help to generate commands.

In Excel you can act more intuitively

- but reconstruction of what you did may be virtually impossible.

4



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 variables.
- 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**

6

```
. cd "D:\Statacourse\your name"
. use "smoke.dta"
. summarize
```

Variable	Obs	Mean	Std.Dev.	Min	Max
id	230	134.069	68.9916	1	250
sex	230	1.7391	.440066	1	2
age	230	55.8043	14.2779	21	84
weight	227	64.0881	11.8579	43	110
height	227	166.973	8.37078	150	194
smoker	230	.921739	.743505	0	2
cigaret	230	4.73043	7.18601	0	40
cheroot	230	.208695	1.10572	0	10
pipe	230	.034782	.226244	0	2

7

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	weight		, detail

8

Calculations

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

// If measured weight or height is missing,
// replace by self-reported weight and height.
replace bmi = (swt)/(sht^2) if missing(bmi)

egen hrqol = rowtotal(q1-q10)

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

9

Once you have a Stata dataset, analysis is easy (tables, estimation, statistical tests, graphs):

```
. use smoke.dta
. tab2 smoker sex
-> tabulation of smoker by sex
```

Smoker?	sex		Total
	male	female	
no	10	63	73
current	38	64	102
former	12	43	55
Total	60	170	230

10

Next generation Stata file

```
// gen_smoke1.do
cd "D:\Statacourse\your name"
use smoke.dta , clear
generate bmi = weight/((height/100)^2)
recode bmi (min/18.5=1)(18.5/25=2)(25/30=3) ///
(30/max=4) , generate(bmigr)

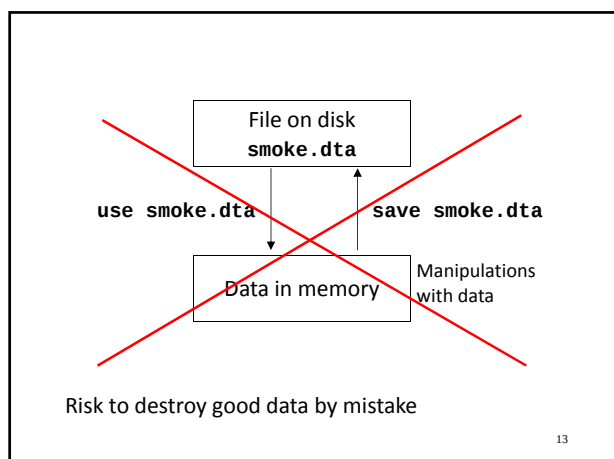
label variable bmi "Body Mass Index (kg/m^2)"
label variable bmigr "BMI, grouped"
label define bmilabel 1 "underwt" ///
2 "normalwt" 3 "overwt" 4 "obese"
label values bmigr bmilabel
save smoke1.dta , replace
```

11

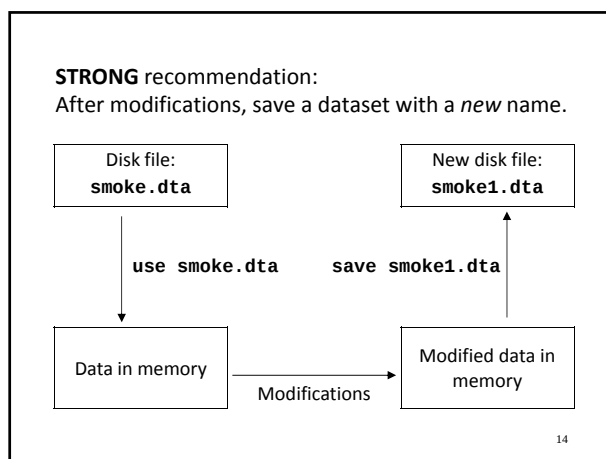
Types of files

smoke.dta	Stata dataset
smoke1.dta	
gen_smoke1.do	Command file (do-file)

12



13



14

Types of commands	
Command type	Examples
File commands Read and write disk files	use save
Calculation Create new variables from the information in existing variables	generate generate...if replace recode
Data documentation Add documenting information to the dataset	label variable label define label values
Analysis commands Create output: tables, test results, graphs, etc.	tab1 tab2 ttest regress

15

Managing output

You can *look* at the text in the Results window, but you cannot do much with it.

In the lab, an output log (a copy of what you saw in the Results window) is generated automatically.

The unofficial **nlog** command opens the output log in a text editor. Here you can remove junk and print selected parts.

Just give the command

nlog

16

Resources

Juul S, Frydenberg M. An introduction to Stata for health researchers, 3rd ed. College Station, TX: Stata Press 2010.

At <http://folkesundhed.au.dk/uddannelse/software> you find:

- Instructions on purchase of Stata and the book (Aarhus GradPlan).
- Short introduction to EpiData and other tidbits.

17