

Seminar 1

Introduction to Stata

There are some important windows in Stata:

- RESULTS: lists the results of Stata operations.
- REVIEW: lists executed commands.
- VARIABLES: lists the variables in memory.
- COMMAND: command window.

1 Importing Data

There are 3 possibilities to import data into Stata:

1. Use built-in databases in Stata. You can see them by going to File-> Example datasets-> Example datasets installed with Stata
`sysuse lifeexp.dta`

2. Use data in your computer that is already in Stata format.

```
use  
"C:\Users\Asus\Documents\AUCA\Teaching\Seminars\Seminar  
1\Data\hh1a.dta", clear
```

3. Use data in your computer that is in Excel format.

```
import excel  
"C:\Users\Asus\Documents\AUCA\Teaching\Seminars\Seminar  
1\Data\Excel bad.xlsx", clear sheet("Excel bad") firstrow
```

2 Examining Data

For this purpose we will use one of the datafiles from "Life in Kyrgyzstan" database:

```
use "C:\Users\Asus\Documents\AUCA\Teaching\Seminars\Seminar  
1\Data\hh1a.dta", clear
```

1. Describe the dataset:

```
describe
```

2. Explore the observations by using the command:

```
browse (br)
```

3. Display the first ten observations of each variable by using the command:

```
list in 1/10
```

List the first 15 observations for the variables `age` and `h108`:

```
list age h108 in 1/15
```

List observations from 1 until the last observation for individuals of 18 years of age and being a grandchild of the head of the household:

```
list in 1/1 if age==18 & h104==9
```

4. Inspect a variable `age` and its distribution. It is also useful for identifying outliers or unusual values, or for spotting non-integers in a variable that should only contain integers:

```
inspect age
```

3 Descriptive Statistics

1. Obtain descriptive statistics of each variable:

```
summarize (sum)
```

2. Obtain descriptive statistics of the variable `age` with additional information about the distribution:

```
sum age, detail
```

What is the age at the 50th percentile, and why is it different from the mean?

3. Tabulate the variable `h105`

```
tab h105
```

4 Modifying Data

1. Rename the variable `h105` to `ethnicity`

```
rename h105 ethnicity
```

2. Generate a new binary (dummy) variable `married`, which takes the value of 1 for a married person and zero otherwise:

```
gen married= 1 if h108== 1  
replace married= 0 if h108== 2 | h108== 3 | h108== 4 |  
h108== 5 | h108== 6
```

There is an alternative way to do it:

```
gen married_new= 1 if h108== 1  
replace married_new= 0 if h108!= 1
```

Compare the two newly created variables for `married`. Why are they different? Which method is correct, and why?

3. Generate a logarithmic transformation of a variable `age`:

```
gen log_age=log(age)
```

In empirical applications, which of the two variables `age` or `log_age` is likely to be more meaningful? Why?

4. Generate a variable, which is `age` squared:

```
gen age_squared=age^2
```

In what kind of applications can the new variable be used?

5. Save the resulting dataset in your computer:

```
save  
"C:\Users\Asus\Documents\AUCA\Teaching\Seminars\Seminar  
1\Data\hh1anew.dta"
```

5 Keeping Track of Things

Instead of typing commands one-by-one interactively, you can type them all in one go within a do-file and simply run the do-file once. It also helps to

keep track of the commands that you have used and save them for future reference. Always use a do-file!

1. Create a new do-file by pressing on the icon **New Do-file Editor**.
2. Copy all of the commands that have been inputted today into the do-file.
3. Execute the first five commands in the do-file by pressing on the icon **Execute selection (do)**.
4. Include your comments for each command by using the following symbols:

```
*; /* */
```
5. Save your do-file.

6 Creating Figures

1. Graph a simple histogram for the variable `age`

```
histogram age
```
2. For this exercise please use the following datafile:

```
use  
"C:\Users\Asus\Documents\AUCA\Teaching\Seminars\Seminar  
1\Data\hh1c.dta"
```

Graph a scatterplot with `h129` on the y axis and `age` on the x axis to see if there is any relationship between these variables:

```
twoway scatter (h129 age)
```

7 Further Notes

1. Whenever you do not know a command or would like to learn more about it, type `help`. For example:

```
help merge
```

2. For many operations in Stata you can also use a drop-down menu, which makes working with Stata easier and also helps you find out what the corresponding command is. However, it is not always possible to use a drop-down menu for more complicated commands and for writing programs.