QS101: Introduction to Quantitative Methods in Social Science

Week 5: STATA Session - Our University

Florian Reiche
Teaching Fellow in Quantitative Methods
Course Director BA Politics and Sociology
Deputy Director of Student Experience and Progression

30.10.2014

A Lab Session

A Lab Session
STATA Commands

Stata Commands

- Structure: *command varlist* **if/in,** *options*
- *command* is the name of the command, such as **summarize, generate** or **tabulate**
- **if** and **in** qualifiers determine what is to be included in the analysis
- *options* control what is done and how results are being presented

Example of Stata Commands

- `summarize age education if male == 1 & age > 17, detail`

Relational Operators

Symbol	Meaning
==	is or is equal to
!= or ~=	is not or is not equal to
>	is greater than
>=	is greater than or equal to
<	is less than
<=	is less than or equal to

`if` and `in`

- **`if`** – perform the command for cases which meet the following criteria
- **`in`** – specifies a subset of cases
  - Example: **`list age education sex in 1/20`**
  - This would only list the first 20 observations
  - "/" reads as "to"

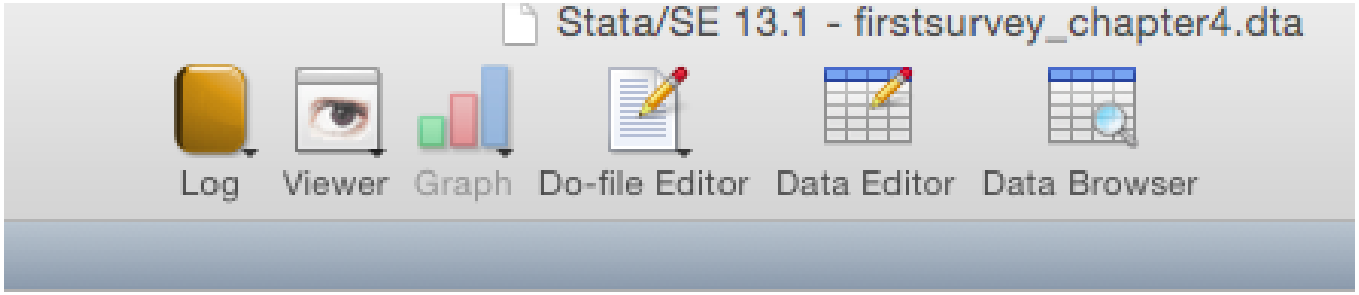
options

- Become increasingly important as we proceed through the year
- Highly relevant to create graphs

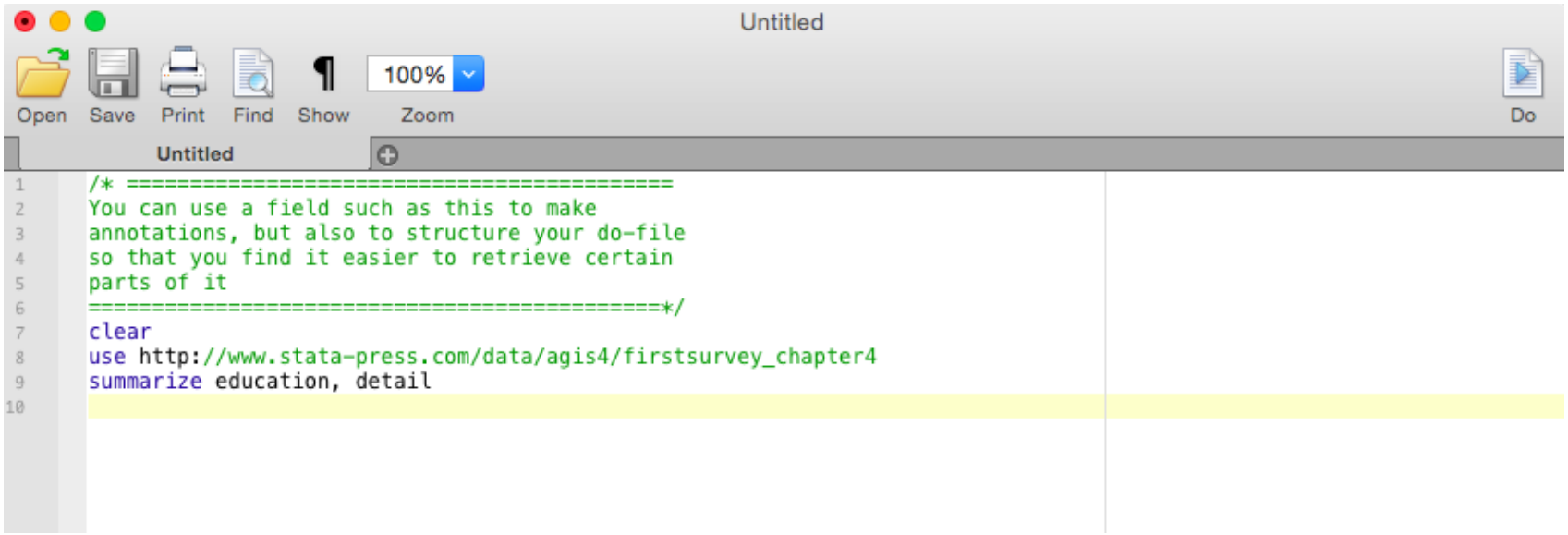
A Lab Session
do-files

Creating a do-file

- Tip: Copy and paste command output into a do-file
- This will allow you to learn using the command line quicker



The do-file window



Descriptives and Graphs for One Variable

Descriptives and Graphs for One Variable
Descriptive Statistics for One Variable

A Shortcut to Everything so far

summarize *varlist*, **detail**

Example Output

```
. summarize education, detail
```

Years of education

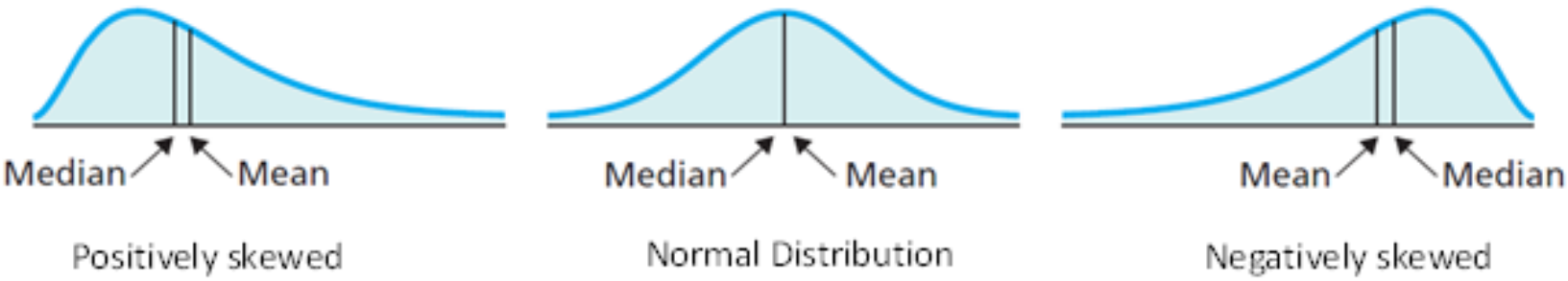
	Percentiles	Smallest		
1%	8	8		
5%	9.5	11		
10%	11.5	12	Obs	20
25%	12	12	Sum of Wgt.	20
50%	14.5		Mean	14.45
		Largest	Std. Dev.	2.946452
75%	16.5	17		
90%	18	18	Variance	8.681579
95%	19	18	Skewness	-.1636124
99%	20	20	Kurtosis	2.522208

The Mean

- There is more than one "mean"
  - Arithmetic mean
  - Harmonic mean (to average rates, such as speed, etc.)
  - Geometric mean (used when growth rates are constant, such as population size, annual income etc.)
  - **ameans** *varlist*

Skewness and Kurtosis

- Skewness
  - Third moment of a distribution
  - Positive: positively skewed
  - Negative: negatively skewed
- Kurtosis
  - Fourth moment of a distribution
  - Describes the size of the peak
  - Normal distribution has a kurtosis of 3



Frequency Tabulations

- Four variations
  - **`tabulate`** and one variable
  - **`tab1`** and several variable names
  - **`tabulate`** and two variable names
  - **`fre`** and several variable names (**`net install fre.pkg`**)

Descriptives and Graphs for One Variable
Graphs

Creating Graphs

- Use the "Graphics" menu

Homework

For Week 7

- Work through sections 5.4. - 5.6. of the Adcock book
- For week 7:
  - Prepare a short presentation (5 mins.) of the results you have obtained from the seminar in week 5, explaining how students have experienced intro week.
  - Use tabular and graphical presentation of your results.
  - Send the finished presentation AT LEAST 24h BEFORE THE SEMINAR to `F.Reiche@warwick.ac.uk`