

Introduction to Stata for MRes/PhD Students

2010/2011

Exercise 2: Data Manipulation

The second class will focus on data manipulation, combining data sets, generating new variables, some housekeeping and creating simple graphs.

1. Getting started

- a) Download the data for the first exercise from my website:
<http://personal.lse.ac.uk/lembcke/teaching.html>
the data is stored in a zip file. Unpack the data into your working folder.
- b) Open Stata and open a new do file by clicking on the appropriate button or using the keyboard shortcut (ctrl+8)
- c) At the beginning of your log file create a header which looks something like:

```
////////////////////////////////////  
// Stata exercise 2 for MRes/PhD students  
////////////////////////////////////  
clear all  
set mem 500m  
set more off  
cd "H:\EC455\data"  
cap log close  
log using "exercise02.txt", replace text
```

2. Opening the data and creating a file to work with

This week the data is split in several files: The annual information for each country, and industry and the information on a country for the whole time period. The first step is to bring all those data sets together.

- d) Use the **append** command to bring all data sets together.
- e) Merge with the country level information.

3. A first descriptive look

- f) Use the descriptive commands from last week to have a look at the data.
- g) What type of data do you have?
- h) Are the data balanced?
- i) Attach labels to all the variables in memory.
- j) Generate the natural log of the exchange rate and the square of the employment share.
- k) Generate the difference between import and export competition index.
- l) Generate a categorical variable that captures the quartiles of import competition.
- m) Create a label that assigns the appropriate label to each of the categories.
- n) Generate the growth rate of import competition
- o) Generate a new variable which is equal to 1.1 for all observations, summarize the import competition growth rate when the new variable is equal to 1.1 (using the if condition).
- p) Drop the newly generated variable from your data set.
- q) Generate a dummy that indicates the highest level of import competition in a country and industry.
- r) Create a graph that depicts the average level of import competition across time for the first 5 industries.