

Workshop 1

1. Stata introduction: Tutorials, installation, data sets, log-file, do-file, help files.

Stata tutorials are can be found on the Stata webpage:

<http://www.stata.com/links/video-tutorials/>

Stata basics

[Quick tour of Stata interface](#)

[Quick help in Stata](#)

[PDF documentation in Stata](#)

[Example data included with Stata](#)

For Workshop 1, you might find the following tutorials helpful:

[Basic scatterplots in Stata](#)

[Histograms in Stata](#)

[Descriptive statistics in Stata](#)

[Simple linear regression in Stata](#)

2. **Motivation:** We will consider the relationship between product innovation in a firm and its use of external sources of knowledge. We will also consider the potential role of internal R&D. This problem also sets up and describes the basic data related to the UK innovation data set that will be further analyzed in the lectures and workshops to come.

3. **Data:**

Use the do-file WORKSHOP1_INITIAL.do to build up your program. This file will merge three different data files and run the first part of the analysis.

4. Variables:

We are using data from the questionnaire for the UK Innovation Survey. Look up the variable definitions from the survey and make sure that you know what the variables measure.

Dependent variable (DV):

- Innovative performance (PRODNEW)

Explanatory variables:

- External sources of knowledge (EXTSOURCE). It is an index capturing the combined use of *ifsuppl ifclient ifcompet ifconslt ifrdlabs* (found in section 12.1)
- Internal R&D intensity (RDINT). Hint: We have calculated this variable as $rdint = xinterm / idbrturn$

Controls:

- Internal constraints (INCONST). It is an index capturing the level of internal constraints due to *hporgrig hplkpers hplktech hplkmkt* (found in section 8.1)
- As a measure of firm size we will use $\log(\text{employ00})$. *Hint: Calculate this variable using the generate command and call it lempl00.*

Note that the variables EXTSOURCE and INCONST have been calculated separately and merged from exercise3_extra_data.dta. They have mean=0 and sd=1 for the full sample.

The following conditions are added to remove potential outliers from your regression sample:

```
drop if expint > 1
drop if rdint > 1 & rdint != .
```

What is the effect of the last condition in these statements? Will it make a difference to the actual analysis run by Stata?

5. Descriptives:

5.1. How many valid observations do you have in your sample?

5.2. Draw up a table that summarizes the variables PRODNEW, EXTSOURCE and RDINT. Report the summary statistics that you find relevant for these variables.

5.3. Create a correlation table for the DV, the two explanatory variables, and the two controls.

5.4. Add further descriptives including graphs that you find relevant and comment on the descriptive analysis.

>>>>>>> Follow-up on questions 1-5 in class <<<<<<<<<<<<<<<<

6. Linear regression:

- 6.1. Run a linear regression with *prodnew* as DV and *extsource* and *rdint* as explanatory variables controlling for *inconst* and *lemp100*.
- 6.2. How do you interpret the coefficient of *rdint*? Is it significant?
- 6.3. How do you interpret the coefficient of *extsource*? Is it significant?
- 6.4. Is there a significant size effect?

>>>>>>> Follow-up on question 6 in class <<<<<<<<<<<<<<<<<