

Introduction to data analysis with Stata

Topic 4

Francesca Marazzi

Research question

Research question

What are the returns to schooling, i.e. what is the effect of studying more or less years on future income? Which other factors influence income?

We will use: data from the British Household Panel Survey (BHPS), 10k+ UK adults interviewed yearly from 1991 onward

Objectives

- Macros: locals and globals
- Value labels ($\neq$ variable labels)
- multivariate OLS regression & coefficient plots
- how to treat categorical data in regressions
- how to export regression results in a publication-style table
- postestimation checks (multicollinearity)
- append different datasets

Macros in Stata

- A **macro** is a string that stands for (a collection of) objects
- the latter can be strings, numbers, a list of variables, a path on your computer...
- think of a macro as the label of a container. Whenever you type that label, Stata will understand the content of the container
- Example: if we build a macro called `year` with content `"1991 1992 1994 1994 1995"`, whenever we type `year` Stata will understand `"1991 1992 1994 1994 1995"`
- Macros can be global or local

Locals Vs globals

- **Locals** are temporary and remain private to the do file they belong to
 - ▶ they will be deleted as soon as the do file stops running
 - ▶ they can be recalled via ``'` (apostrophe and quote)
- **Globals** are public and remain stored in Stata's temporary memory
 - ▶ they will be dropped only when you close Stata (or if you explicitly drop them)
 - ▶ they can be recalled via `$`

Merge Vs append

The commands `merge` and `append` are both used to combine information from different datasets. However

- `merge` is used to add variables (therefore, columns) for the same units of observation (years, people, countries...)
- `append` is used to add observations (therefore, rows) for the same variables