

Syllabus for Stata Sessions

Session 1 (Basics):

- Using Stata:
  - The different windows
  - Typing commands in the command line (as compared to point and click)
  - Command line versus do file
  - Using help (and Google)
  - Basic operators (&, !, ==, etc)
  - Importing and exporting different types of files
- Basic things to do with a dataset:
  - Describe / codebook / summarize / tabulate
  - Gen / replace
  - Keep / drop
  - Labelling
  - Other standard commands: sort, order, rename, display, count
- Merging and Appending of files
  - One to One merging
  - Many to One merging
  - One to Many merging
  - Appending

Session 2 (Basic Summary Statistics and Graphs):

- Calculating summary statistics:
  - Calculating proportions (e.g., number of girls in specific districts and in the whole dataset, mean exam scores by gender, etc)
  - Introducing tabstat to make summary tables
  - Using collapse to produce a table with province/district/school-level statistics (also including preserve and restore)
  - Exporting output to MS Word and Excel
- Data visualisation:
  - Graphs: histogram, kdensity, twoway, bar, cibar, box, pie, dot, scatter
  - Formatting graphs
  - Exporting graphs to .png files

Session 3 (Active Session – Basics):

- Active session 1: we will start with a raw student-level Grade 7 dataset, and then clean it and do some basic analysis
- Active session 2: we will start with a raw student-level Grade 9 or 12 dataset, and then clean it and do some basic analysis

Session 4 (Data Analysis):

- t-tests and comparisons of means
  - Very brief recap of theory (more detailed recap in the appendix)
  - Testing for differences within a variable (i.e., testing by a dummy)
  - Testing for differences between two variables
- ksmirnov tests to test for distributional equality (even when the means are the same)
- Regressions:
  - Interpretation of regressions and regression coefficients
  - Introduction to the regress command through a standard 'reg y x'
  - Adding controls (including i. for categoricals)
  - Brief introduction to using esttab (into Word) and making publication-quality tables
- Correlation:
  - Spearman and Pearson Correlation
  - Interpretation (and pitfalls of correlation)
  - pwcorr and associated commands

Session 5 (Intermediate Data Management):

- Introduction to egen to calculate summary statistics at different levels of data (e.g., calculating the school mean in a pupil-level dataset)
- Reshaping data
- Brief introduction to loops and locals (as well as globals)

Session 6 (Active Session – Intermediate):

- Active session: we will start with a raw *exam-level* Grade 9 dataset and need to clean it and do some basic analysis
  - Note: Unlike the grade 7 one, this involves egen to calculate subject totals, reshape to get a student-level dataset (rather than exam-level), some loops, etc

A few leftover things that we will likely will not have time to cover:

- Extracting results from r() or e()
- Working with strings (regexm, regexr, strpos, strlen, substr, etc)
- Log files