Stata Tutorial for DEAL

Datasets and Basics Analysis

Ricardo E. Miranda [1] and Xiaoqi (Jade) Peng[1]

Duke University[1]

Nov 29 - 30, 2022

Outline

Introduction

Plan for the day

Learn basics of exporting data analysis findings through excel forms and figures.

We will do this using the datasets from the previous lecture

- Data on expenditures and characteristics of households.
- Data on living place characteristics

Exporting Excels

putexcel

putexcel set sets the Excel file to create, modify, or replace in subsequent putexcel commands.
You must set the destination file before using any other putexcel commands.

- modify or replace: set options that modifies or completely replace the set file
- sheet("SheetName") writes results to a sheet with that name, or creates a new sheet if it does not exist yet

putexcel writes Stata expressions, matrices, tables, images, and returned results to an Excel file. It can also format the cells in the worksheet.

- matrix writes whole column/rows of data into excel
- locals

Freedom to create your own table format, but you must be careful with rows and columns.

outreg2

The command **outreg2** gives you the type of presentation you see in academic papers. It is not a Stata command, but a user-written procedure. Install it by typing (only the first time)

ssc install outreg2

Basic syntax:

outreg2 using desktop filepath/filename, fileformat

where the file format could be word, excel, or tex.
Other options:

- **replace** replaces any existing file that has the same name.
- **append** adds new columns to the existing ones in the file name specified.
- **addstat** augments the standard regression results with other e-class (and r or s class if they exist) statistics (scalar only).

Less flexible than putexcel but requires coding.

Drawing Figures

Common Figure Types

- **scatter** is both a command and a plottype. For example, the following command gives scatterplot overlaid with a line (the *lfit* part) showing the linear relationship between $y1$ and x

  twoway scatter y1 x || lfit y1 x

  but the "twoway" could be omitted to achieve the same effect
- **hist** could plot both discrete (use "discrete" to specify) and continuous data (use "bin" to specify the number of bins). Default options is plotting density; could be changed to *fraction*, *frequency*, or *percentage*. Use *horizontal* or *vertical* to specify the orientation of the graph.

The **twoway graphs** show the relationship between numeric data; it fits onto numeric y and x axes.

Common Figure Types - Continued

- **bar** graph bar draws vertical bar charts. In a vertical bar chart, the y axis is numerical, and the x axis is categorical.
  Syntax:

  graph bar v1, over(catvar1) by(catvar2)

  This command plots bar graph of the mean of $v1$ in separate graph areas for each level of catvar2, with bars for each level of catvar1 in each graph area
- **kdensity** Unlike the histogram, here the range is divided into **overlapping intervals** and we collect the center-point density estimates. Also, a kernel density estimator assigns a weight between 0 and 1 based on the distance from the center of the window and **sums the weighted values**.
  - **kernel**: specify kernel function; default is kernel(epanechnikov)
  - **bwidth**: half-width of kernel
  - **generate(newvar_x, newvar_d)**: store the estimation points in newvar_x and the density estimate in newvar_d

Adjusting Figures

Stylistic Adjustments

- Set both color and opacity:

  [object] color (colorstyle%#)
- Line options: **lpattern(), lwidth(), lcolor()**
- Market options: **msymbol(), mcolor(), msize()**
- legend([content], [location])
- The axises: **xscale(), xlabel(), xtitle()**
- graphregion: the background; plotregion