Session 3: Graphs in Stata

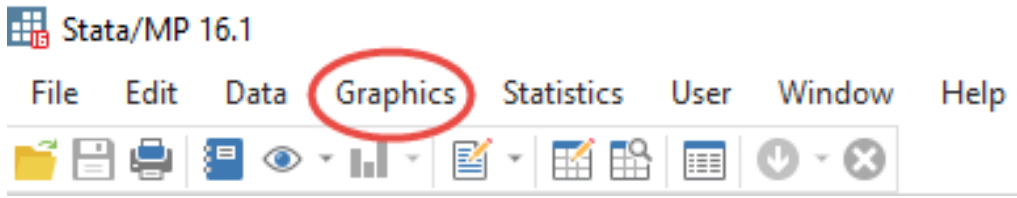
Manuel V. Montesinos

Computation Brush-Up Course
Competition and EPP Master Programs

Fall 2021

Overall Syntax for Graphs

Stata has a rich set of **graphic facilities**, accessible through the `graph` command (or the Graphics tab on the Stata menu bar).



The overall syntax of graph commands is

```
graph-command (plot-command, plot-options),
graph-options
```

Type `help graph` for an overview of Stata options for graphing.

Line Plot

For the following plots, we will use the dataset *US life expectancy*. Type `sysuse uslifeexp` and it will automatically open the dataset.

For a two-way plot, we will use the command `graph twoway`. For a line plot, we add the statement `line`, followed by the variables in the y and x axis, by this order:

`graph twoway line le year` (shortcut: `line le year`)

Scatter Plot

For scatter plots, we will use the dataset *life expectancy*. Type `sysuse lifeexp, clear` and it will automatically open the dataset.

We add the statement `scatter`, followed by the variables in the y and x axis (life expectancy against GNP per capita), by this order:

`graph twoway scatter lexp gnppc` (shortcut: `scatter lexp gnppc`)

If you want to make two scatter plots, use `||` or `()` as a separator between the two plot types.

Moreover, you can show a fitted regression line by using the `lfit` plot. Fitted line with confidence intervals: `lfitci`.

Editing and Managing Graphs

Titles, legends and captions: there are options that apply to all two-way graphs, including titles, subtitles, legends, notes and captions.

Axis scales, labels and titles: there are options that control the scaling and range of the axes, including `xscale` and `yscale`, which can be arithmetic, log, or reversed. Other options control the labeling of axis and labels, such as `xtitle` and `ytitle`.

Line styles: Stata lets you control the line pattern (`lpattern`), width (`lwidth`) and color (`lcolor`).

Editing and Managing Graphs

Labeling points (scatter plot options): Stata allows you to control the markers used for the points, including their shape and color. It is also possible to label the points using text included in another variable, using the `label` or `mlabel` options.

Managing graphs:

- Save current graph: `graph save graphname`.
- Load a saved graph: `graph graphname`.
- Check graphs in directory: `graph display`.
- You can also combine graphs using `graph combine`.

Histogram

Graphical representation of the distribution of the data: it is an **estimate** of the **probability distribution** of a continuous variable in a population.

Basic syntax for plotting a histogram: `histogram variablename`.

Options: `bin()` for changing the number of bars; `normal` to overlay the normal curve; `addlabels` to add values to the bars.

Moreover, you can change the numeric depiction of your data by choosing one of these options: `density`, `fraction`, `frequency`, `percent`.

Kernel Density

Graphical summary of the **shape** of the data. Kernel density estimation provides a **non-parametric estimate** of the probability density function of a random variable.

The **bandwidth** of the kernel is a free parameter which exhibits a strong influence on the resulting estimate. The most commonly used option for kernel densities is `bwidth`, which sets the bandwidth.

Exercise

1. Use the Stata web dataset *womenwage*: `webuse womenwage`.
2. Make a scatter plot between wage (*wage*) and schooling (*school*).
3. Plot the fitted line in the same graph.
4. Add title, subtitles, a note...
5. Make two scatter plots with their respective fitted lines, one for the subsample of women in rural areas and the other for the subsample of women in urban areas.
6. Combine the two previous graphs.