

A Very Short Introduction to Stata

Andy Grogan-Kaylor

2022-06-16

The basic philosophy of Stata—"Stata in one sentence"—is:

`do_something to_variable(s), options`

The general idea of most Stata commands is `command variable(s), options`. Often it is not necessary to use any options since the authors of Stata have done such a good job of thinking about the defaults. Commands that you actually type are represented in monospace font. `x` and `y` refer to variables in your data.

Task	Command
Open data	<code>use mydata.dta</code>
Descriptive statistics	<code>summarize x y</code>
Frequencies	<code>tabulate x</code>
Correlation	<code>corr x y</code>
Regression	<code>regress y x</code>
Logistic Regression	<code>logit y x</code> , or ¹
Multilevel Model	<code>mixed y x group:</code>
Structural Equation Modeling	<code>sem (y <- x m) (m <- x)</code>
Histogram	<code>histogram x</code> ²
Bar Graph	<code>graph bar, over(x)</code>
Bar Graph (of means)	<code>graph bar y, over(x)</code>
Pie Chart	<code>graph pie, over(x)</code>
Scatterplot	<code>twoway scatter y x</code>

¹Here we need to use the `,or` option to ask for *odds ratios* instead of *logit coefficients*.

²For graphing commands, you can often add options after a `,.` e.g. `title("title of the graph"), xtitle("title of the x axis"), ytitle("title of the y axis")`.