

# A Very Short Introduction to Stata

Andrew Grogan-Kaylor

2024-01-24

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 z</code>
Logistic Regression	<code>logit y x z, or<sup>1</sup></code>
Ordinal Logistic Regression	<code>ologit y x z, or<sup>2</sup></code>
Multinomial Logistic Regression	<code>mlogit y x z, rr<sup>3</sup></code>
Multilevel Model	<code>mixed y x z    group: x</code>
Structural Equation Modeling	<code>sem (y &lt;- x m z) (m &lt;- x z)</code>
Histogram	<code>histogram x<sup>4</sup></code>
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>

---

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

<sup>2</sup>Here again we need to use the `, or` option to ask for *odds ratios* instead of *logit coefficients*.

<sup>3</sup>Here we need to use the `, rr` option to ask for *risk ratios* instead of *logit coefficients*.

<sup>4</sup>For 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").`