

# A Very Short Introduction to Stata

The basic *philosophy* of Stata.

Andrew Grogan-Kaylor

2025-10-22

## 💡 Key Idea

The basic philosophy of Stata—“Stata in one sentence”—is:

`do_something to_variable(s), options`

Often it is not necessary to use any options since the authors of Stata have done such a good job of creating an intuitive language, and 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>
Look for or find	<code>lookfor thing</code> <sup>1</sup>
Describe the data	<code>describe x</code> <sup>2</sup>
Descriptive statistics	<code>summarize x y</code>
Frequencies	<code>tabulate x</code>
Cross-Tabulation.	<code>tabulate x y</code> <sup>3</sup>
Recode	<code>recode x (old = new)(...), generate(xR)</code> <sup>4</sup>
Rename	<code>rename x z</code>
Keep	<code>keep x y z</code>
Drop	<code>drop x y z</code>
Correlation	<code>corr x y</code>

<sup>1</sup>`lookfor thing` looks for any variable with `thing` in the variable name or variable label. `lookfor somethingelse` looks for any variable with `somethingelse` in the variable name or variable label. It is often useful to `lookfor` abbreviations e.g. `lookfor anx` instead of `lookfor anxiety`.

<sup>2</sup>`describe, short` will give you quick summary information about the data including *sample size*.

<sup>3</sup>After the `,` the `row` and `col` options can be helpful to generate *row* and *column* percentages.

<sup>4</sup>It is usually best practice, but not required, to `recode` values of a variable (e.g. `x`) into a *new* variable (e.g. `xR`), leaving the original variable untouched.

Task	Command
Regression	<code>regress y x z</code>
Logistic Regression	<code>logit y x z, or</code> <sup>5</sup>
Ordinal Logistic Regression	<code>ologit y x z, or</code> <sup>6</sup>
Multinomial Logistic Regression	<code>mlogit y x z, rr</code> <sup>7</sup>
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</code> <sup>8</sup>
Bar Graph (of categories)	<code>graph bar, over(x)</code> <sup>9</sup>
Bar Graph (of means over categories)	<code>graph bar y, over(x)</code>
Pie Chart	<code>graph pie, over(x)</code>
Scatterplot	<code>twoway scatter y x</code>

---

<sup>5</sup>Here we need to use the , `or` option to ask for *odds ratios* instead of *logit coefficients*.

<sup>6</sup>Here again we need to use the , `or` option to ask for *odds ratios* instead of *logit coefficients*.

<sup>7</sup>Here we need to use the , `rr` option to ask for *risk ratios* instead of *logit coefficients*.

<sup>8</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").`

<sup>9</sup>For bar graphs, the `asyvars` option is often helpful, as it causes the bars to be different colors.