A *Very* Short Introduction to Stata

The basic *philosophy* of Stata.

Andrew Grogan-Kaylor

2025-10-01

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	`use mydata.dta`
Look for or find	`lookfor thing` [1]
Describe the data	`describe x` [2]
Descriptive statistics	`summarize x y`
Frequencies	`tabulate x`
Cross-Tabulation.	`tabulate x y` [3]
Recode	`recode x (old = new)(...), generate(xR)` [4]
Rename	`rename x z`
Keep	`keep x y z`
Drop	`drop x y z`
Correlation	`corr x y`
Regression	`regress y x z`
Logistic Regression	`logit y x z, or` [5]

[1] `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`.

[2] `describe, short` will give you quick summary information about the data including *sample size*.

[3] After the `,` the `row` and `col` options can be helpful to generate *row* and *column* percentages.

[4] 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.

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

Task	Command
Ordinal Logistic Regression	`ologit y x z, or` [6]
Multinomial Logistic Regression	`mlogit y x z, rr` [7]
Multilevel Model	`mixed y x z || group: x`
Structural Equation Modeling	`sem (y <- x m z) (m <- x z)`
Histogram	`histogram x` [8]
Bar Graph (of categories)	`graph bar, over(x)` [9]
Bar Graph (of means over categories)	`graph bar y, over(x)`
Pie Chart	`graph pie, over(x)`
Scatterplot	`twoway scatter y x`

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

[7] Here we need to use the `, rr` option to ask for *risk ratios* instead of *logit coefficients*.

[8] 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")`.

[9] For bar graphs, the `asyvars` option is often helpful, as it causes the bars to be different colors.