A Very Short Introduction to Stata

The basic *philosophy* of Stata.

Andrew Grogan-Kaylor

2024-05-15

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	`use mydata.dta`
Descriptive statistics	`summarize x y`
Frequencies	`tabulate x`
Correlation	`corr x y`
Regression	`regress y x z`
Logistic Regression	`logit y x z, or` [1]
Ordinal Logistic Regression	`ologit y x z, or` [2]
Multinomial Logistic Regression	`mlogit y x z, rr` [3]
Multilevel Model	`mixed y x z || group: x`
Structural Equation Modeling	`sem (y <- x m z) (m <- x z)`
Histogram	`histogram x` [4]
Bar Graph	`graph bar, over(x)`
Bar Graph (of means)	`graph bar y, over(x)`
Pie Chart	`graph pie, over(x)`
Scatterplot	`twoway scatter y x`

---

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

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

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

[4] 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")`.