

Brief list of Stata commands used in Basic Biostatistics

This list is intended to aid you in identifying the command needed for a particular task. While comprehensive in the sense that any of the exercises in the course should be solvable using one or more of the commands below, it is by no means exhaustive as other commands may exist that would allow solving some exercises more elegantly. Although the list is not sorted, I have tried to present them roughly in the order you will encounter the commands in this course, but it is intended that you should work a bit on familiarizing and getting an overview of the commands yourself.

No syntax description is given here, so to learn how to actually use a command look it up with the built-in **-help-** command of Stata (for example **help use**). To search for the appropriate command use the **-search-** command. Note: You should never type the dashes next to the command name, as they are merely part of the common Stata style for referencing command names.

Main commands

- pwd-** Present working directory of Stata (where does it look for datafiles?)
- cd-** Change working directory.
- use-** Open a dataset from the hard disk.
- save-** Save current dataset to disk.
- browse-** Open spreadsheet-like window of dataset.
- list-** List of individual observations in dataset.
- log-** Open, pause, or close a log-file.
- keep-** Keep some variables OR some observations (using -if-).
- drop-** Drop some variables OR some observations (using -if-).
- generate-** Create new variable.
- replace-** Replace values in existing variable.
- recode-** Recode values in existing variable.
- egen-** Extended version of -generate-.
- label-** Attach label to variable, create labels for numeric values, assign created label to numeric values in a variable
- format-** Determine how variable values are presented in output.
- reshape-** Rearrange data into wide or long format.
- display-** Display text or results in output, may be used as calculator.
- describe-** Briefly list variables in dataset.

-codebook– Give detailed description of variables in dataset.

-count– Count number of observations, total or those fulfilling specific conditions.

-summarize– Compute mean, standard deviation, number of non-missing observations, etc. for numeric variables.

-return list– Describe computed results left behind by general commands.

-tabulate– Frequency tables of variables, and cross tabulations.

-table– Table of summary statistics.

-centile– Percentiles of variable, empirical and normal based (-normsd-option).

-ci– Confidence intervals, means, proportions, and rates.

-sdtest– Two sample comparison of standard deviations.

-ttest– One or two sample mean comparison, paired or un-paired.

-signrank– Wilcoxon matched-pairs signed-ranks test.

-signtest– Sign test of matched pairs.

-ranksum– Non-parametric two-sample test (Wilcoxon rank-sum test).

-prtest– One- and two-sample tests of proportions (approximative).

-cs– Two-sample comparison of proportions (risk difference, relative risk, odds ratio), estimates, confidence intervals, t-test, Fisher's exact test.

-cc– Case-control analysis.

-mcc– Matched case-control analysis.

-regress– Linear regression.

-ereturn list– Describe computed results left behind by estimation commands.

-predict– Predict fitted values, residuals after regression.

-lincom– Transform and combine estimates and confidence intervals.

-adjust– Table of adjusted means and proportions (in Stata 11 superseded by -margins-).

-correlate– Correlation matrix.

-pwcorr– Pairwise correlations between variables, p-values.

-spearman– Spearman's rank correlation.

-oneway– Oneway analysis of variance.

-anova– Analysis of variance.

-kwallis– Kruskal-Wallis equality-of-populations rank test

-logit– Logistic regression, shows log-odds-ratios.

- logistic**- Logistic regression, shows odds-ratios.
- stset**- Declare data to be survival time data.
- sts test**- Log-rank test.
- stcox**- Cox proportional hazards regression.
- sampsi**- Power and sample size calculations.

Graph commands

- graph**- The mother of all graph commands.
- graph combine**- Combine different graphs into one plot.
- dotplot**- Dot plot of a variable.
- histogram**- Histogram of a variable.
- graph box**- Box-plot of one or more variables.
- qnorm**- Q-Q-plot of variable.
- graph twoway**- The mother of all twoway graphs.
- scatter**- XY-plot of two or more variables.
- sts graph**- Plot survivor curve or related graphs.
- stcurve**- Plotted fitted survivor, hazard, and cumulative hazard.
- stphplot**- Log-log plot of survival curves (proportional hazards?)
- stcoxkm**- Kaplan-Meier and predicted survival plot
- graph export**- Save copy of graph to disk.

Other commands

- bysort**- Execute command for each value of variable(s).
- scalar**- Save number in named scalar.
- forvalues**- Repeat commands in a numbered loop.
- foreach**- Repeat commands for each element in a list of variables, numbers, strings, etc.
- global**- Save something in a global macro. (persists after execution of do-file).
- local**- Save something in a local macro (disappears automatically and immediately after execution of do-file).