

# Stata Basic Syntax Structure

*command* *varlist* *qualifiers*, *options*

*command* 1, 2, or 3 words specifying the task

*varlist* 0, 1, 2 or more variables

= *exp* A math or logical statement to set a value

if *exp* A logical statement to limit cases

, *options* Words and selectors to alter the command

## Using Stata Help

Type *help command*

In Stata documentation and here, the following conventions are used to describe the proper syntax for each command:

[brackets] not needed / optional element

*italics* replace with your own name(s) or value(s)

underline the shortest possible abbreviation

in this document, **bold** indicates the suggested abbreviation.

## Common Symbols

	or	==	is equal to	<	less than
&	and	!=	not equal to	<=	less than or equal to

## Challenge

What would these be if Stata syntax were used in everyday life?

order burger, meat(2) cheese(American)

order burger if (place != "Carls Jr"), nomayo

order burger if (price < 3), plain mustard

## Answers

1. Double-cheeseburger
2. From any place but Carls Jr, with all the usual toppings except mayo
3. Cheap burger with only mustard



## Page 1: Open and Manipulate Data

File Manipulation Syntax Structure: *command varlist qualifiers using file, options*

Organize	Set working directory	<b>cd</b>	<i>"directory"</i>	<i>This allows "relative" file names</i>
Log Results	Start a log file	<b>log</b>	<b>using file</b>	<i>Add a .log extension for plain text</i>
	Save a started log file	<b>log</b>	<b>close</b>	<i>You must close a log file to save it</i>
Open and Save Files	Stata (.dta)	<b>use</b>	<i>"file"</i>	<i>Opens file.dta in working directory</i>
		<b>save</b>	<i>"file"</i>	<i>Saves file.dta in working directory</i>
	Delimited File (e.g., csv)	<b>import <u>delim</u></b>	<b>using "file.csv" , <u>delim</u>(" , ")</b>	<i>Omit option if tab or comma/csv</i>
	Excel File	<b>import <u>excel</u></b>	<b>using "file.xls"</b>	<i>See also "help import"</i>
Create Variables	Use math	<b>generate</b>	<i>newvar = exp</i>	<i>exp is any math expression</i>
	Indicator/Binary Variables	<b>generate</b>	<i>newvar = (x == #) if !missing(x)</i>	<i>Dummy Code for #: True = 1, False = 0</i>
	Use Functions	<b>egen</b>	<i>newvar = function(x)</i>	<i>See "help egen" for functions*</i>
Change Values	Use math	<b>replace</b>	<i>x = exp</i>	<i>exp is any math expression</i>
	Modify individual values	<b>recode</b>	<i>x (1 2=1) (3 = 0), generate(x_r)</i>	<i>Generates a categorical variable</i>
	Group number ranges	<b>recode</b>	<i>x (1/4=1) (5/9=2), generate(x_r)</i>	<i>Generates a categorical variable</i>
	Convert string to numeric	<b>destring</b> <b>encode</b>	<i>x, generate(x_r)</i> <i>x, generate(x_r)</i>	<i>When the values are actually numbers</i> <i>Creates labels from string/text values</i>
Delete Data	Remove variables/columns	<b>drop / keep</b>	<i>x</i>	<i>Drop (remove) or Keep [only] the variables or observations indicated</i>
	Remove observations/rows	<b>drop / keep</b>	<b>if exp</b>	
Labeling	Variable Labels	<b><u>label variable</u></b>	<i>x "My label"</i>	<i>Labels are for your reference only</i>
	Value Labels	<b><u>label define</u></b> <b><u>label values</u></b>	<i>numbers 1 "One" 2 "Two"</i> <i>x numbers</i>	<i>Both commands are needed, but the defined label can be reused</i>

*\* the function **std** makes z-scores*

## Page 2: View and Summarize Data

$x$  → the name (not the label) of the variable of interest

$z$  → the name of a categorical / grouping variable

Open Windows	Look at Data	<b><u>browse</u></b>		Use <b><u>edit</u></b> to allow changes
	Variables Manager	<b><u>varmanage</u></b>		Can be used to label values
	Do Files	<b><u>doedit</u></b>		Do files hold syntax for redo-ing tasks
Display Information	about the variable	<b><u>describe</u></b>	$x$	Omit the <i>vaname</i> for info about all vars
	about the values	<b><u>codebook</u></b>	$x$	Add <b><u>, compact</u></b> for a table of variables
Confidence Intervals	for Numeric Variables	<b><u>ci</u></b>	$x$	Gives the CI around a mean
	for Proportions	<b><u>ci</u></b>	$x$ , <b><u>binomial</u></b>	Gives the CI around a proportion
Univariate Descriptive Statistics	for a Numeric Variable	<b><u>summarize</u></b>	$x$	<i>N/Obs.</i> , Mean and Standard Deviation
		<b><u>summarize</u></b>	$x$ , <b><u>detail</u></b>	Also Median*, Skewness, Kurtosis, etc.
Bivariate Descriptive Statistics	for a Categorical Variable	<b><u>tabulate</u></b>	$x$	Frequency table of all values
	for 2 Categorical Variables	<b><u>tabulate</u></b>	$x_1 x_2$ , <b><u>column row</u></b>	Contingency table with row & col percents
	for Numeric + Categorical	<b><u>tabulate</u></b>	$z$ , <b><u>summarize(x)</u></b>	Statistics for $x$ by grouping variable $z$
		<b><u>tabulate</u></b>	$z$ , <b><u>sum(x) nofreq nost</u></b>	Pivot Table, displays only means by group
Univariate Graphs	Histogram	<b><u>histogram</u></b>	$x$ , <b><u>normal</u></b>	Omit the option for a clean graph
		<b><u>histogram</u></b>	$x$ , <b><u>width(#)</u></b>	Specify the width of each bin
Bivariate Graphs	Box-and-Whiskers Plot	<b><u>graph box</u></b>	$x$	One numeric variable
	Box Plot by Groups	<b><u>graph box</u></b>	$x$ , <b><u>over(z)</u></b>	Separated by grouping variable $z$
	Scatter Plot	<b><u>scatter</u></b>	$x_1 x_2$	For two numeric variables
Graph Labeling	Add Text Labels	<i>any graph</i>	$x$ , <b><u>title(My Title)</u></b>	Appears over the graph
		<i>any graph</i>	$x$ , <b><u>note(My Notes)</u></b>	Appears under the graph
Compare Groups	“Prefix” Command	<b><u>bysort</u></b>	$z$ : <i>any command</i>	Use with any above to do for each value of $z$

\* the median is labeled “50%”

### Page 3: Inferential Statistics

$x \rightarrow$  Independent Variable (IV), Criterion, or Grouping Variable

$y \rightarrow$  Dependent Variable (DV) or Response Variable

Examine Associations	Nominal	<b>Chi-Square</b>	<b>tabulate</b>	$x_1 x_2$ , <b>chi2 expect</b>	<b>expect</b> includes expected values
	Ordinal	Kendall's Rank <b>Correlation</b> Coef.	<b>tabulate</b>	$x_1 x_2$ , <b>taub</b>	<i>PRE alternative, see also <b>ktau</b></i>
		Spearman's Rank <b>Correlation</b>	<b>spearman</b>	$x_1 x_2$	<i>Comparable to Pearson's Correlation</i>
	Numeric	Pearson's <b>Correlation</b> Coefficient	<b>pwcorr</b>	$x_1 x_2$ , <b>sig listwise</b>	<i>Uses pairwise deletion by default (<b>listwise</b> does the same as <b>correlate</b>)</i>
Compare a statistic to a specific value: #		One-Sample <b>t-Test</b>	<b>ttest</b>	$x == \#$	<i>For a mean or a mean difference</i>
		One-Sample <b>Proportions Test</b>	<b>prtest</b>	$x == \#$	<i>For a proportion or a proportion difference</i>
Compare specific values to each other (e.g., before vs after)		Test of Equal Variances	<b>sdtest</b>	$x_1 == x_2$	<i>If significant, add <b>unequal</b> to <b>ttest</b></i>
		Paired <b>t-Test</b>	<b>ttest</b>	$x_1 == x_2$	<i>Compares paired numeric values</i>
		Paired <b>Proportions Test</b>	<b>prtest</b>	$x_1 == x_2$	<i>Compares paired proportions</i>
		Wilcoxon Paired <b>Sign-Rank Test</b>	<b>signrank</b>	$x_1 == x_2$	<i>Compares paired ranks (non-parametric)</i>
Compare Groups	2 Groups	Test of Equal Variances	<b>robvar</b>	$y$ , <b>by(x)</b>	<i>If significant, add <b>unequal</b> to <b>ttest</b></i>
		Independent <b>t-Test</b>	<b>ttest</b>	$y$ , <b>by(x)</b>	<i>Compares 2 means</i>
		Independent <b>Proportions Test</b>	<b>prtest</b>	$y$ , <b>by(x)</b>	<i>Compares 2 proportions</i>
		Wilcoxon <b>Rank-Sum Test</b>	<b>ranksum</b>	$y$ , <b>by(x)</b>	<i>Compares 2 ranks (non-parametric)</i>
	3+ Groups	Kruskal-Wallis <b>Rank Test</b>	<b>kwallis</b>	$y$ , <b>by(x)</b>	<i>Compares 3+ ranks (non-parametric)</i>
		<b>ANalysis Of VAriance</b>	<b>oneway</b>	$y x$	<i>Compares 3+ means (1 IV)</i>
Predict Values	Numeric	[Multiple] Linear <b>Regression</b>	<b>regress</b>	$y x_1 x_2 \dots$	<i>Numeric <math>y</math>, Numeric or binary/dummy <math>x</math>'s</i>
		Plot Residuals vs Fitted Values	<b>rvfplot</b>		<i>Check homoskedasticity after running reg</i>
	Binary	Logistic <b>Regression</b>	<b>logit</b>	$y x_1 x_2 \dots$	<i>Reports coefficients (log-odds)</i>
			<b>logistic</b>	$y x_1 x_2 \dots$	<i>Reports odds-ratios</i>