Useful Stata Commands

Command	Short	Menu Path	Sub-Commands	Description
use	N/A.	File → Open	N/A.	Opens a file, though the full file location is needed (don't forget to include double quotation marks if any part of the file location or name has a space in it).
mvdecode	N/A.	Data → Create or change data → Other variable-transformation commands → Change numeric values to missing	mv	Recodes specified values into specified missing values for specified variables (or all variables if _all is specified instead of a particular variable name). The mv sub-command precedes that list the old and new values. e.g. mv(-9=.a).
recode	N/A.	Data → Create or change data → Other variable-transformation commands → Recode categorical variable	generate	Recodes the specified values into newly specified values for specified variables, e.g. to reverse a five-point scale (4=0) (3=1) (2=2) (2=3) (0=4). The generate sub-command creates a new variable recoded as specified, rather than replacing the values in the original variable, which is much safer.
generate	N/A.	Data → Create or change data → Create new variable	normal	Generates a new variable with the same data as a specified existing variable, but without any of the value labels, e.g. generate[new variable] = [existing variable]. The normal sub-command creates a variable expressing the proportion of cases that fall below each case in turn.
replace	N/A.	Data → Create or change data → Other variable-transformation commands → Recode categorical variable	N/A.	Replaces the values of a variable (usually used straight after the generate command, i.e. to create the values on a new variable) based on an if function.
clonevar	N/A.	Data → Create or change data → Other variable-transformation commands → Clone existing variables	N/A.	Clones the specified variable, including all the value labels, but with a new name, e.g. clonevar [new variable] = [existing variable]. This then gives a new variable to be changed as necessary whilst leaving the original variable unaltered.
codebook	N/A.	Data → Describe data → Describe data contents (codebook)	compact	Provides a description of the variables specified (or all the variables in the dataset if none is specified) including mean, standard deviation, and percentiles. The compact sub-command provides a briefer output.
describe	N/A.	Data → Describe data → Describe data in memory or in a file	N/A.	Provides a description of the variables specified (or all the variables in the dataset if none is specified) including the value labels.
list	N/A.	N/A.	N/A.	Provides a list of the values on the specified variable(s) of each case in the dataset.
summarize	sum	Statistics → Summaries, tables, and tests → Summary and descriptive statistics → Summary statistics	detail	Provides summary statistics for the specified variables or all the variables in the dataset if none is specified). The detail sub-command provides more detailed summaries.

tabulate	tab	N/A.	chi2 V taub	Provides a table of the variable specified summarizing how many cases fall into each category and including percentages and cumulative percentage. Can use the command with two variables, in which case it will cross tabulate them. If a crosstab is created then the chi2, V, and taub sub-commands will, respectively, return the Chi-square, Cramer's V, and Kendall's tau-b statistics for the relationship.
tabs1	N/A.	Statistics → Summaries, tables, and tests → Frequency tables → Multiple one-way tables	N/A.	Provides individual tables for multiple variables rather than crosstabs (as would be the case if two variables were listed after the tabulate command).
tabstat	N/A.	Statistics → Summaries, tables, and tests → Other tables → Compact table of summary statistics	statistics, by	Provides a table for the specified variable listing the specified statistics and, if the by sub-command is used, can sub-divide by another specified variable.
mean	N/A.	Statistics → Summaries, tables, and tests → Summary and descriptive statistics → Means	N/A.	Provides the mean(s) of for the specified variable(s).
bysort	N/A.	N/A.	sum	Sorts by the specified variable and then runs the subsequent command across each of the categories of that variable.
numlabel	N/A.	Data → Data utilities → Label utilities → Prepend values to value labels	add, remove	Adds or removes (depending on the specified sub-command) the numerical values alongside the value labels in a variable.
histogram	N/A.	Graphics → Histogram	Normal, width, start, discrete, frequency, percent, title, subtitle, xtitle, note, legend, scheme	Provides a histogram of the specified variable. The normal sub-command draws a normal distribution curve on the histogram. The width sub-command specified how wide (on the scale of the variable in question) the columns will be (e.g. how many years they will cover). The start sub-command specifies the lowest value that will be included in the histogram. The discrete sub-command specifies that the values in the variable are discrete rather than continuous. The frequency and percent sub-commands specify whether the y axis should display those things respectively. The various title, note, and legend sub-commands allow the addition of text to various parts of the histogram. The scheme sub-command allows the colour scheme of the histogram to be specified.
kdensity	N/A.	Statistics → Nonparametric analysis → Kernel density estimation	addplot	Provides a density plot of the specified variable; this is the same as the histogram function but provides a smooth line rather than columns.
graph box	N/A.	Graphics → Box plot	over	Provides a box plot of the specified variable. The over sub-command provides a separate box plot for each of the

				categories in a second specified variable.
save	N/A.	File → Save	N/A.	Saves the file, though the full file location is needed (the programme will return an error message if there is already a file in that location with the same name).
pwcorr	N/A.	Statistics → Summaries, tables, and tests → Summary and descriptive statistics → Pairwise correlations	sig obs	Creates a correlation matrix for the specified variables. The sig and obs sub-commands, respectively, provide the significance for and number of observations included in each correlation.
regress	reg	Statistics → Linear models and related → Linear regression	beta	Regresses any number of independent (interval or dummy) variables onto a dependent (interval) variable, which must be the first variable in the list of all variables. The resulting table automatically includes the values of b, their significance, and the R-square value. The beta sub-command adds a column showing the value of beta for each independent variable.
scatter	N/A.	Graphics → Twoway graph (scatter, line, etc.)	N/A.	Plots each case with data on two variables on a graph with each axis representing one of those variables.
lfit	N/A.	Graphics → Twoway graph (scatter, line, etc.)	N/A.	Plots a line of best fit for all cases that have data on two variables, with those two variables represented by each axis of the graph.
twoway	N/A.	Graphics → Twoway graph (scatter, line, etc.)	N/A.	Allows two (or more) things (e.g. a scatter plot and a line of best fit) to be displayed on the same graph. The initial command must be followed by a set of brackets containing the command for the first type of graph (as it would be written if it was being produced on its own) then a second set of brackets containing the command for the second kind of graph to be plotted), and so on.