Minitab 17 commands

Prepared by Jeffrey S. Simonoff

Data entry and manipulation

To enter data by hand, click on the `Worksheet` window, and enter the values in as you would in any spreadsheet. To then save the data as `Minitab` file, click on the `Session` window, and then click on the "diskette" icon and enter a file name in the appropriate box. This creates a `Minitab` "project" file in .mpj format. These files also can include statistical output, graphs, and multiple data files. To read in a previously saved data file, click on the "open folder" icon and enter the file name in the appropriate box. Data files also can be inputted to `Minitab` as worksheets. Click on `File` → `Open Worksheet` and choose the appropriate file type (possibilities include Excel, Extensible Markup Language (.xml), and text) and file name.

An inherently continuous variable can be converted to a discrete one (which can then be summarized using a frequency distribution) by clicking on `Data` → `Code`. If the code is to a set of numbers, click `Numeric to Numeric`; if it is to a set of labels, click `Numeric to Text`. Enter the original variable under `Code data from columns:`, and a new variable name or column number under `Store coded data in columns:` (you can put the same variable name in the latter box, but then the original values will be overwritten). Enter a set of original ranges in the boxes under `Original values:`, and the corresponding new values in the boxes under `New:`. If more than eight recodings are needed, the dialog box can be called repeatedly until all values are recoded.

To create new variables or transform old ones, click on `Calc` → `Calculator`. Enter the new variable name under `Store result in variable` and the transformation desired under `Expression`. If you put a current variable name under `Store result in variable` you will overwrite the old values of the variable. So, for example, to take logs (base 10), use *LOGT( )* under `Expression` (`Log base 10` under `Functions:`).

Although it is possible to omit observations in a sample by simply highlighting them in the data worksheet and pressing the delete key, this is generally not advisable, since then the observation cannot be recovered without reopening the original file (and if you save the data before doing that, the observation is gone completely). A better approach is to create a subset of the worksheet that has the observations you want; this will create a new worksheet that can be analyzed, but the original worksheet will still be there as well. Click on `Data` → `Subset Worksheet`. You can give the new worksheet an identifying name if you like under `Name:`. Click the radio button next to `Specify which rows to exclude`, click the radio button next to `Specify rows:`, and enter the row numbers of the

outliers in the associated box. Note that there is a good deal of flexibility in the subsetting; you can identify rows to include or exclude, identify them by some condition (for example, observations with values of a predictor greater than 10), or brush them on a scatter plot and identify them that way, in addition to specifying them by row number(s).

Summary statistics and graphics

Frequency distributions for qualitative variables are obtained by clicking on `Stat` → `Tables` → `Tally Individual Variables`. Enter the variable names in the box under `Variables:` (note that you can double click on variable names to "pick them up"). You'll probably want to click on the box next to `Percents` to get the percent of the sample with each value. If there is a natural ordering to the categories, clicking on the boxes next to `Cumulative counts` and `Cumulative percents` is also appropriate.

To construct a histogram, click on `Graph` → `Histogram`. Clicking on `Simple` will give a histogram, while clicking on `With Fit` will superimpose the best-fitting normal curve. A technically more correct version of the plot would be obtained by first clicking on `Tools` → `Options` → `Individual graphs` → `Histograms` (double clicking on `Individual graphs`) and clicking the radio button next to `CutPoint`. Right-clicking on a constructed graph allows you to change the appearance of the histogram (this is true for most plots). You can add a frequency polygon to the histogram by right-clicking on it, clicking on `Add` → `Smoother`, and changing the `Degree of smoothing` value to 0. You can then omit the bars (leaving only the frequency polygon) by right-clicking on the graph, clicking `Select item` → `Bars`, and then pressing the delete key (you can select other parts of this, and most other, graphical displays this way, and delete them if you wish the same way).

To construct a stem–and–leaf display, click on `Graph` → `Stem-and-Leaf`. Enter the variable name under `Graph variables`. Note that the stem–and–leaf display appears in the Session Window, not in a Graphics Window. To get stem–and–leaf displays separated by groups, click the box next to `By variable:`, and enter the variable that defines the groups (the variable that defines the groups must use different integer values to define the groups).

To construct a boxplot, click on `Graph` → `Boxplot`. To construct separate boxplots of different variables, click on `Simple` and enter the variable name(s) under `Graph variables`. To construct side–by–side boxplots, click on `Graph` → `Boxplot` → `With Groups`. Enter the variable name(s) under `Graph variables` and the variable(s) that determine the groups under `Categorical variables for grouping`. Note that these grouping variables can be either numerical or categorical.

To obtain descriptive statistics, click on `Stat` → `Basic Statistics` → `Display Descriptive Statistics`. Enter the variable name(s) under `Variables`. To get descriptive statistics separated by groups, click the box next to `By variables:`, and enter the variable that defines the groups (either a numerical or categorical grouping variable can be used here).

To construct a scatter plot, click on `Graph` → `Scatterplot` → `Simple`. Enter the variable name for the vertical axis under `Y variables` and the variable name for the horizontal axis under `X variables`. You can construct multiple scatter plots at the same time by adding variable names. To construct a plot with different symbols for observations from different groups, click on `Graph` → `Scatterplot` → `With Groups`. Enter the variable name for the vertical axis under `Y variables`, the variable name for the horizontal axis under `X variables`, and the name of the variable that defines the groups under `Categorical variables for grouping`. To superimpose a lowess curve on a plot, right-click on the plot, and click `Add` → `Smoother`, and click `OK` when the dialog box pops up. To superimpose separate lowess lines for different groups on the same plot, you would click in the box next to `Apply same groups of current displays to lowess smoother` (this only appears in the multiple group scatter plot situation).

To construct a scatter plot with a least squares regression line superimposed, click on `Stat` → `Regression` → `Fitted Line Plot`. Enter the variable name for the vertical axis under `Response (Y):`, and the variable name for the horizontal axis under `Predictor (X):`. Much more extensive regression output, including diagnostic statistics and graphics, is obtained by clicking `Stat` → `Regression` → `Regression` → `Fit Regression Model`. Enter the target variable (Y) under `Responses:` and the predictor (X) under `Continuous predictors:`.

To obtain the correlation coefficient between two variables, or among a set of variables, click on `Stat` → `Basic Statistics` → `Correlation`. Enter the variable names under `Variables`.

To obtain the covariance between two variables, or among a set of variables, click on `Stat` → `Basic Statistics` → `Covariance`. Enter the variable names under `Variables`. Entries along the diagonal of the output matrix are the variances of each variable, while entries along the off-diagonal are the covariances between the pairs of variables.

To construct a normal plot, click on `Graph` → `Probability Plot` (single). Enter the variable you are plotting under `Graph variables:` and click `OK`.

Interval estimation

A z–based confidence interval for the mean is obtained by clicking `Stat` → `Basic`

Statistics → 1-Sample z. Enter the variable of interest in the box to the right, and the true population standard deviation in the box next to Known standard deviation:. The coverage level of the interval can be changed by changing the number in the box next to Confidence level: after clicking Options. Plots of the variable can be obtained by clicking on Graphs.

A t–based confidence interval for the mean is obtained by clicking Stat → Basic Statistics → 1-Sample t. Enter the variable of interest in the box to the right. The coverage level of the interval can be changed by changing the number in the box next to Confidence level: after clicking Options. Plots of the variable can be obtained by clicking on Graphs.

To obtain confidence intervals for a Binomial proportion, click on Stat → Basic Statistics → 1 Proportion. If you already have the number of events (successes) and the number of trials, change the drop-down box to say Summarized data, enter in the number of events (x) and the number of trials (n) in the appropriate dialog boxes, and click OK. The interval that comes out is the exact version (based on the Binomial distribution). To get the standard Central Limit Theorem–based interval, click on Options, and then change the drop-down box next to Method to Normal approximation. If you have a variable that corresponds to successes and failures (ones and zeroes), keep the setting One or more samples, each in a column and enter the variable name in the box below.

Hypothesis testing

A z–test is obtained by clicking Stat → Basic Statistics → 1-Sample z. Enter the variable of interest in the box to the right, and the true population standard deviation in the box next to Known standard deviation:. Click the box next to Perform hypothesis test, and enter the null mean in the box next to Hypothesized mean:. The form of the alternative hypothesis can be changed by clicking Options and changing the drop-down box next to Alternative hypothesis:. To obtain a graphical display of the data, with null mean value and confidence interval for the mean, click on Graphs.

A t–test is obtained by clicking Stat → Basic Statistics → 1-Sample t. Enter the variable of interest in the box to the right. Click the box next to Perform hypothesis test, and enter the null mean in the box next to Hypothesized mean:. The form of the alternative hypothesis can be changed by clicking Options and changing the drop-down box next to Alternative hypothesis:. To obtain a graphical display of the data, with null mean value and confidence interval for the mean, click on Graphs.

A paired t–test is obtained by clicking Stat → Basic Statistics → Paired t. Enter the variables of interest under Sample 1: and Sample 2:. Clicking Options allows

you to change the null mean value (if necessary); the form of the alternative hypothesis can be changed by changing the box next to `Alternative hypothesis:`. To obtain a graphical display of the differences, with null mean value and confidence interval for the mean difference, click on `Graphs`.

A sign test is obtained by clicking `Stat` → `Nonparametrics` → `1-Sample Sign`. Enter the variable of interest under `Variables:`. Click the radio button next to `Test median:`, and change the entry in the box to the actual null median value (if necessary). The form of the alternative hypothesis can be changed by changing the box next to `Alternative:`.

A signed rank test is obtained by clicking `Stat` → `Nonparametrics` → `1-Sample Wilcoxon`. Enter the variable of interest under `Variables:`. Click the radio button next to `Test median:`, and change the entry in the box to the actual null median value (if necessary). The form of the alternative hypothesis can be changed by changing the box next to `Alternative:`.

Two–sample t-tests are obtained by clicking on `Stat` → `Basic Statistics` → `2-Sample t`. There are two possible forms for the data: with the variable in one column, with a second column containing codes for the two groups (the *stacked* form), or with the variable separated into two columns, one for each group (the *unstacked* form. If the data are in stacked form, keep the drop-down box entry as `Both samples in one column`, enter the variable name under `Samples:`, and the variable that defines the groups under `Sample IDs:`. The subscript variable can be either numerical or text. If the data are in unstacked form, change the drop-down box to `Each sample is in its own column`, and enter the two variables in the boxes next to `Sample 1:` and `Sample 2:`, respectively. If you want the t-test that assumes equal variances in the two groups, click `Options`, and then the box next to `Assume equal variances`.

You can convert from stacked to unstacked form, and vice versa. To convert from stacked to unstacked, click on `Data` → `Unstack Columns`. Enter the variable(s) to be split up in the box next to `Unstack the data in:`. Enter the variable that defines the groups in the box next to `Using subscripts in:`. You can then choose where to put the new variables, and whether `Minitab` should name them for you. To convert from unstacked to stacked, click on `Data` → `Stack` → `Columns`. Enter the variables to be combined under `Stack the following columns:`. You can then choose where to put the stacked variable and associated variable of subscripts, and whether you want the subscripts to be the names of the variables (if you uncheck that box, the subscripts are the integers 1, 2, etc.).

Normal plots of more than one variable on the same plot are obtained by clicking on `Graph` → `Probability Plot`, then `Multiple`, and entering the variable names under `Variables:`. Note that this corresponds to the data being in *unstacked* form. If the data are in *stacked* form the different curves can be obtained by entering the grouping variable

under Categorical variables for grouping.

Tests of homogeneity of variance, and confidence intervals for standard deviations, are obtained by clicking on Stat → ANOVA → Test for Equal Variances. Enter the variable of interest in the box next to Response:, and the variable that defines the groups under Factors:.

The median test is obtained by clicking on Stat → Nonparametrics → Mood's Median Test. The data can only be treated if they are in *stacked* form. Enter the variable of interest in the box next to Response:, and the variable that defines the groups in the box next to Factor:.

To obtain a Mann–Whitney test, click on Stat → Nonparametrics → Mann-Whitney. The data can only be treated if they are in *unstacked* form. Enter the two variables that have the observations for the two groups in the boxes next to First Sample: and Second Sample:, respectively.

A confidence interval and hypothesis test for the equality of two Binomial proportions can be obtained by clicking Stat → Basic Statistics → 2 Proportions. Data can be given in stacked form (zeroes and ones in one column, with a second column identifying groups), unstacked form (two columns of zeroes and ones), or as summarized values of the number of trials and number of successes in each group.

Regression

To fit a linear least squares regression line, click on Stat → Regression → Regression → Fit Regression Model. Enter the target variable under Responses: and the predicting variable under Continuous predictors:. To perform a multiple regression, just enter the variables desired as predictors under Continuous predictors:.

A prediction interval and a confidence interval for average y for a given predictor value are obtained as a followup to the regression fit. Click on Stat → Regression → Regression → Predict after you have fit your model. Enter the appropriate value(s) under Enter individual values to get the intervals. If you want to get intervals for more than one observation, you can put all of the variable values needed in columns, and put in the column names rather than individual value(s) after changing the drop-down menu to Enter columns of values.

To obtain regression residual plots, while in the regression dialog box click on Graphs. Under Residual Plots click Normal plot of residuals, Residuals versus fits, and Residuals versus order (if you have data with a natural ordering, like time series data). To get a four-in-one residual plot, click on Graphs while fitting a regression, and click the radio button next to Four in one. To get plots of residuals versus different variables, just

enter the names of the variables you want under `Residuals versus the variables:`. To get plots of standardized residuals when constructing graphs in a regression, change the drop-down box next to `Residuals for plots:` to say `Standardized`.

To create a regression plot with pointwise confidence and prediction intervals superimposed, click on `Stat` → `Regression` → `Fitted Line Plot`. Enter the target variable under `Response (Y):` and the predicting variable under `Predictor`. Click on `Options`, and click on `Display confidence bands` and `Display prediction bands` under `Display Options`.

To save regression diagnostics when performing a regression, click on `Storage`, and then click on `Standardized residuals`, `Leverages`, and `Cook's distance`.

Best subsets regression is performed by clicking on `Stat` → `Regression` → `Regression` → `Best Subsets`. Enter the target variable under `Response:` and the predicting variable(s) under `Free predictors:`. If there are any variables that you want to be in all regression models, enter them instead under `Predictors in all models:`. The output will list the two models of each size (one predictor, two predictors, etc.) with highest R^2, which allows you to see how many predictors it takes to account for most (or all) of the potential predictive power in the predictors.