

STUDENT GUIDE TO USE OF MENUS FOR RESAMPLING

About this manual

This manual is intended for use with the *Bootstrap Methods and Permutation Tests* by Hesterberg, Monaghan, Epstein, and Clipson, W. H. Freeman, which may be used independently or as a supplement to *The Practice of Business Statistics* by Moore, McCabe, Duckworth and Schlove or *Introduction to the Practice of Statistics* by Moore and McCabe. Note that there are different versions of *BMPT* for *PBS* and *IPS*.

This manual explains how to use the dropdown menus to perform resampling. It is not meant to be a substantive introduction to using S-PLUS. Please refer to the *S-PLUS Manual* by Snow & Chihara (2003) for a more complete introduction to S-PLUS; this is available from www.whfreeman.com, as a supplement to *Introduction to the Practice of Statistics* by Moore and McCabe.

There is additional documentation on using S-PLUS for the exercises in *BMPT* contained in the *PBSdata* and *IPSdata* libraries available from www.insightful.com/Hesterberg/bootstrap

---Caveat---

When this version of the manual was written the version of *BMPT* for *IPS* and the *IPSdata* library, do not yet exist.

---Acknowledgements---

This manual was written by Ashley Clipson and Rachel Epstein. The authors would like to acknowledge the contributions of Tim Hesterberg and Shaun Monaghan, and support from the National Science Foundation under grant DMI-0078706.

**Table of
Contents**

About this manual	1
Table of Contents	2
How to obtain the software and data	3
How to load the software and data	3
Windows you can view in S-PLUS	3
Looking at data	4
Plotting data	5
Bootstrapping a statistic and calculating standard error	6
Plotting a bootstrap object	8
Percentile confidence intervals, BCa limits, and tilting limits	9
Calculating a t interval	10
Bootstrap the difference in means	11
Permutation test for the difference in means	13
Bootstrapping correlation	14
Bootstrap a single proportion	15
Bootstrap the difference in proportions	16
Permutation test for the difference in proportions	17
Bootstrap the slope and intercept of a regression line	18
Bootstrapping complex statistics	20
Finite Sampling	22
Stratified Sampling	23
Statistics from the Command line	24

How to obtain the software and data

To bootstrap using S-PLUS, you must first download the free S+Resample library. To do exercises in *The Practice of Business Statistics* it is easiest to use the PBSdata library. To do exercises in *Introduction to the Practice of Statistics* it is easiest to use the IPSdata library.

Information on obtaining and installing these is available from:

www.insightful.com/Hesterberg/bootstrap

How to load the software and data

At the beginning of every S-PLUS session, the S+Resample library must be loaded to enable resampling procedures on your computer.

1. Open S-PLUS on your computer.
2. Load the resample library from the menu **File > Load Library**
3. Choose resample from the list.
4. Then click on “Attach at top of search list”

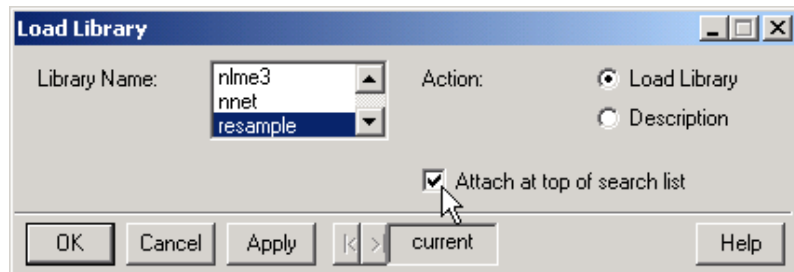


Figure 0.1: Loading the resample library.

5. Click OK. **This must be done at the beginning of every startup.**
6. To load either the PBSdata or IPSdata libraries, repeat steps 2, 3, and 5, but with either PBSdata or IPSdata in place of resample. The data libraries need not be loaded in position 1.

Windows you can view in S-PLUS

There are three windows you are most likely to use:

1. The *Object Explorer* window, which contains all of the objects you have saved.
2. The *Report window*, which is where S-PLUS prints output. You may type in this window and save the report for later reference.

3. The *Commands* window, which enables you to specify more advanced or detailed instructions for the computer.

Other windows include:

4. The *History* window, which holds a record of the commands you have input in both the drop-down menus and the Commands window.

5. The *Commands History* window, which holds a record of only the instructions you specified in the Commands window.

6. The *Script* window, which allows you to edit and save your commands.

Looking at data

To look at data, from the main menu bar, go to **Data > Select Data**.

1. Check “Existing Data” in the Source region.
2. Type the name of the data set in the Existing Data region.

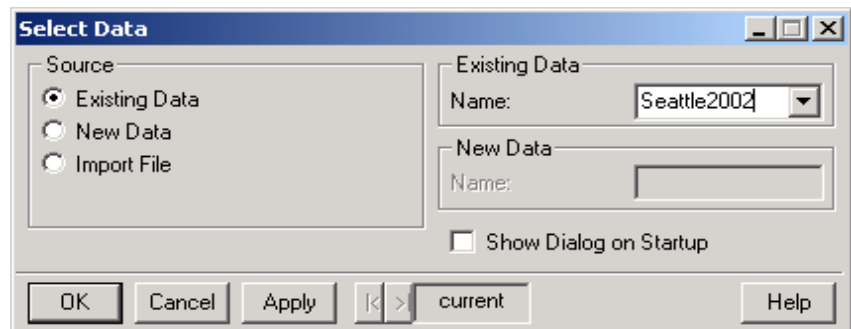


Figure 0.2: *Importing data.*

3. Click OK. This loads the data set into the *Object Explorer* window.

****Note:** You may want to import data from a file or disk. To do this, select “Import File” instead of “Existing Data” and do not name the data. Select “OK” and locate and name the file you are importing in the screen that appears.

4. To view a data set, open the *Object Explorer* window.



Figure 0.3: Location of the *Object Explorer* window icon.

5. Then double-click on the name of the data set. This opens the data.

Object	Pos	Data Class	Dimension
sdData	1	numeric	1
Seattle2002	1	data.frame	2669x1
seDist	1	numeric	1
sept	1	numeric	5
set	1	integer	5

Figure 0.4: Opening data from the *Object Explorer* window.

Plotting data

In order to make a plot, you must first open your data in the manner described in the *Looking at data* section.

1. Open the 2D plot menu by clicking on the 2D plot icon in the main menu bar.



Figure 0.5: Location of the *2D plot* icon.

2. Highlight the data you wish to plot.

		1	2
		Col1 ↓	
1		142.00	
2		232.00	
3		132.50	
4		200.00	
5		362.00	

Figure 0.6: Where to click to highlight a column.

3. Click on the plot icon for the type of graph you wish to create.

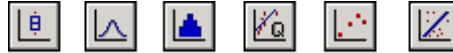


Figure 0.7: *Examples of plot icons. (Pictured here: Boxplot, Density, Histogram, Normal probability plot, Scatterplot, and Linear regression).*

4. You can edit the graphs once they are produced by right-clicking on the graphs and choosing different options.

Bootstrapping a statistic and calculating standard error

To bootstrap the mean, quartiles, median, variance, standard deviation, skewness, and/or kurtosis, open **Statistics > Data Summaries > Summary Statistics with Bootstrap....**

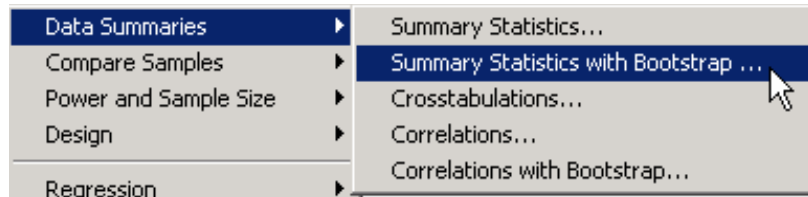


Figure 0.8: *Location of Summary Statistics menu.*

1. In the Data tab, enter the name of your data set in the “Data Set” section of the Data region. Then, if appropriate, select the variable(s) you wish to analyze. Variable names will be the column names in your data set.

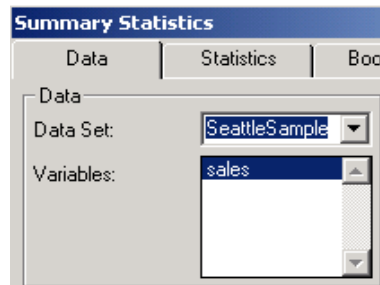


Figure 0.9: *Selecting data to bootstrap.*

2. You can choose to save your results in the “Save As” section of the Results region.

3. On the Statistics tab, select the statistic(s) that you wish to calculate. We choose the mean and median for this as an example.

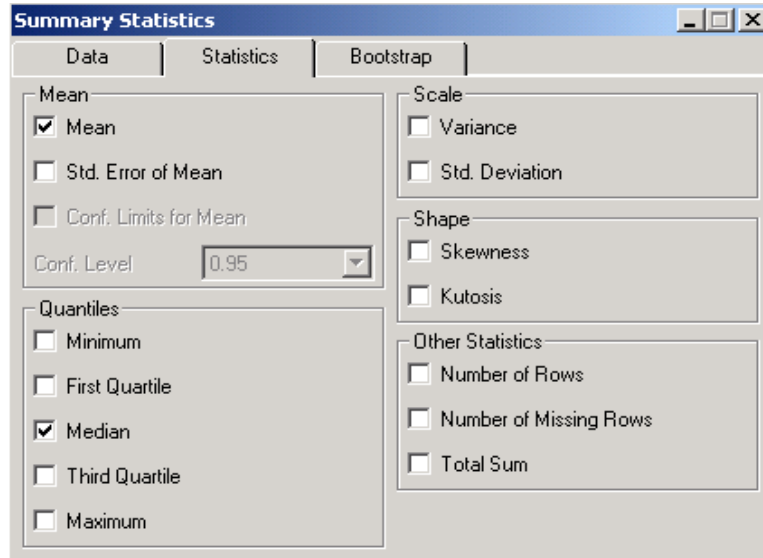


Figure 0.10: *Choosing the statistic to bootstrap.*

4. On the Bootstrap tab, select Perform Bootstrap in the Bootstrap region.

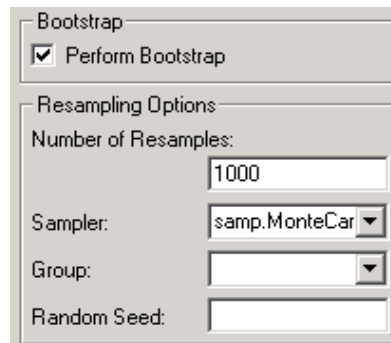


Figure 0.11: *The Bootstrap menu.*

5. In the Resampling Options region, specify the number of resamples you wish to generate (typically 1000) in the “Number of Resamples” box. Here you can also specify the type of resampling you wish to do in the “Sampler” box, the group argument in the “Group” box, and set the seed in the “Random Seed” box.

6. You can save your bootstrap object to the Object Explorer window by filling in the “Save As” box in the Save Bootstrap Object region.

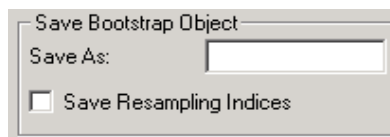


Figure 0.12: *Saving a bootstrap object and supplying additional arguments.*

7. Click OK. The summary of the bootstrap object (including the standard error of the statistic) will print in a *Results* window.

Your output will look similar to the following:

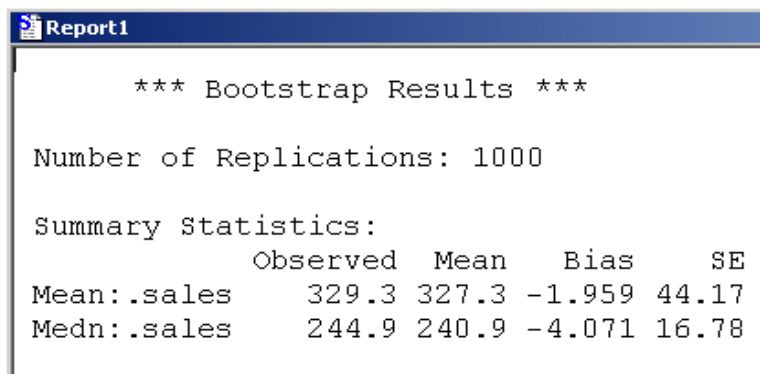


Figure 0.13: *Example output from a Report window.*

“Observed” = statistic from the original sample

“Mean” = mean of the bootstrap replicates

“Bias” = difference between observed statistic and the mean of the bootstrap replicates

“SE” = standard error of the observed statistic

Plotting a bootstrap object

To plot a bootstrap object’s distribution of replicates or its normal probability plot, you must first create the bootstrap object as outlined in the *Bootstrapping a statistic and calculating standard error* section.

1. In the Bootstrap tab, select “Distribution of Replicates” and/or “Normal Quantile-Quantile” in the Plots region.

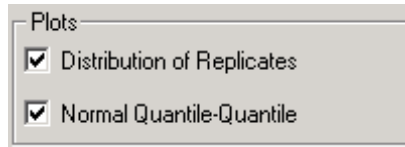


Figure 0.14: *Plotting a bootstrap object.*

2. Click OK. The graph(s) you selected will print in a graph sheet.

Percentile confidence intervals, BCa limits, and tilting limits

Calculating Percentile confidence intervals, BCa limits, and tilting limits requires creating the bootstrap object in the same manner as was described in the *Bootstrapping a statistic and calculating standard error* section.

1. Select the intervals you wish to create in the Summary Results region of the Bootstrap tab.

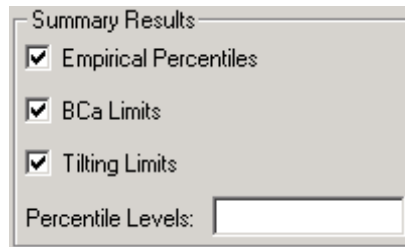


Figure 0.15: *Selecting intervals for the bootstrap object.*

2. Specify the percentile levels in the “Percentile Levels” box. Note that the limits for .025 and .975 produce a 95% interval, whereas the limits for .05 and .95 produce a 90% interval.

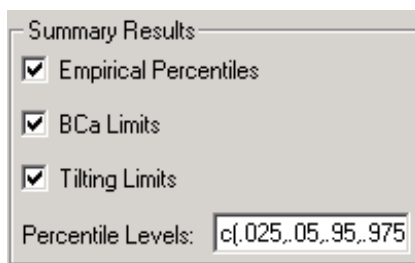


Figure 0.16: *Selecting confidence intervals to display. The defaults are BCa Limits and Tilting Limits. The percentile levels apply to all intervals selected.*

3. Bootstrap tilting produces two types of limits, using exponential tilting and maximum-likelihood tilting. In the text, only maximum-likelihood tilting is shown because it is generally a bit more accurate.

Calculating a t interval

For a non-bootstrap t interval, you can find the standard error of the mean by using the **Summary Statistics** menu. For a bootstrap t interval, you must find the bootstrap standard error using the methods in the *Bootstrapping a statistic and calculating standard error* section.

1. Bootstrap the mean and its standard error using the **Summary Statistics** menu. Be sure to save the bootstrap object.

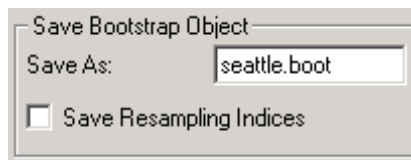
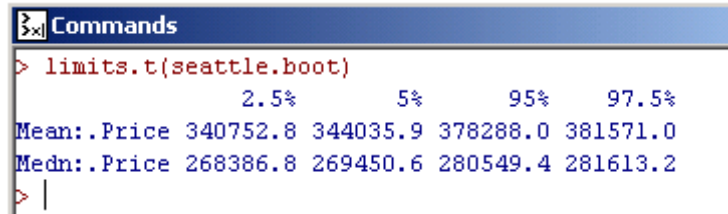


Figure 0.17: *Where to save a bootstrap object.*

2. In the *Commands* window, type “limits.t(your bootstrap object name)”



```
> limits.t(seattle.boot)
                2.5%      5%      95%      97.5%
Mean:.Price 340752.8 344035.9 378288.0 381571.0
Medn:.Price 268386.8 269450.6 280549.4 281613.2
> |
```

Figure 0.18: Example input and output of finding the t intervals using the *Commands window*.

3. The endpoints of your confidence interval will print in a *Report* window.

****Note:** You can always find the t statistic using t tables, and plugging your results by hand into the formulas you learned in the text.

Bootstrap the difference in means

From the menu, select **Statistics > Compare Samples > Two Samples > 2t Test with Bootstrap** to bootstrap the difference in means. This menu also allows you to perform a traditional studentized t Test.

1. In the “Data Set” box, type the name of the data set.
2. There are two different ways your data may be expressed:

a) If the data set consists of two columns, each containing a different set of data, type the names of each of the columns containing the variables in the boxes labeled “Variable 1” and “Variable 2”.

The image shows a 'Two-sample t Test' dialog box with the following settings:

- Data:**
 - Data Set:
 - Variable 1:
 - Variable 2:
 - ☐ Variable 2 is a Grouping Variable
- Test:**
 - Type of t Test: ☒ Paired t, ☒ Two-sample t
 - ☒ Assume Equal Variances
- Hypotheses:**
 - Mean Under Null Hypothesis:
 - Alternative Hypothesis:
- Confidence Interval:**
 - Confidence Level:
- Results:**
 - Save As:
 - ☒ Print Results

Figure 0.19: *Entering the data on the Main tab for bootstrapping the difference of the means, when the two datasets are in separate columns.*

b) If the data consists of one column containing the data in two groups and one column containing a grouping variable (such as a list of 1's and 2's), type the name of the column containing the data points in the box labeled “Variable 1”. Then type the name

of the column containing the grouping variable in the box labeled “Variable 2” and check the box labeled “Variable 2 is a Grouping Variable”.

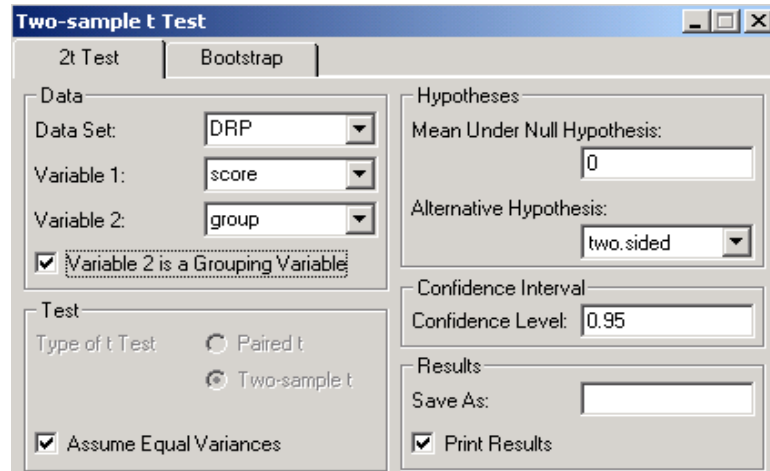


Figure 0.20: Entering the data on the Main tab for bootstrapping the difference of the means, when the two datasets are in one column (e.g. Height) with a grouping variable in another column (e.g. Sex).

3. On the Bootstrap tab, check “Perform Bootstrap”.
4. If confidence intervals are desired, check the box(es) next to the type of confidence intervals you want to create.
5. The results will be printed in the Report window. This may take a while because a permutation test is also being performed to give a P-value.

Permutation test for the difference in means

To perform a permutation test for the difference in means, select **Statistics > Compare Samples > Two Samples > 2t Test with Bootstrap** from the main menu bar and then follow steps 1-6 in the *Bootstrap the difference in means* section.

1. On the Main tab in the t-Test menu, choose an alternative hypothesis, either **two.sided**, **greater**, or **less**.

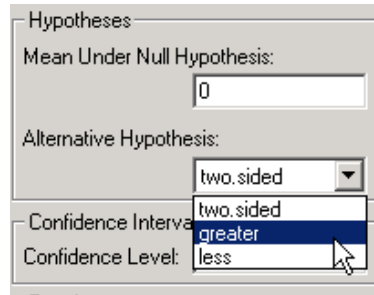


Figure 0.21: *Selecting an alternative hypothesis for a permutation test.*

2. On the Bootstrap tab, select “Perform Bootstrap”.

****Note** that it is possible to print confidence intervals for the bootstrap along with your permutation test simply by selecting a confidence interval from the Bootstrap tab.

3. Click OK. The *P*-value for the permutation test will automatically be printed in the Report window.

Bootstrapping correlation

From the menu, you must select **Statistics > Data Summaries > Correlations with Bootstrap**

1. Type the name of your data set in the “Data Set” box.
2. Select the variables for which you want to bootstrap the correlation, or choose **<ALL>** to bootstrap the correlation between all the variables. To select more than one variable, hold down the “Control” key while you click on the variables you are interested in.

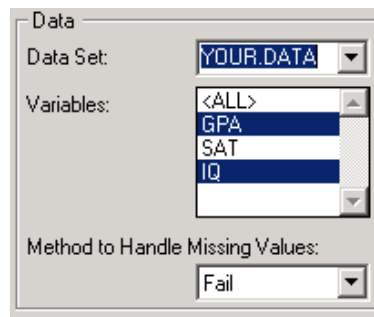


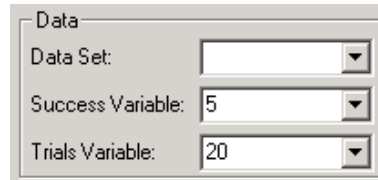
Figure 0.22: *Selecting the variables for bootstrapping correlation.*

3. On the Bootstrap tab, select “Perform Bootstrap”.
4. If confidence intervals are desired, check the appropriate boxes in the “Summary Results” section of the menu.
5. Click OK. The results will be printed in the Report window.

Bootstrap a single proportion

From the menu, select **Statistics > Compare Samples > Counts and Proportions > Proportions Parameters**.

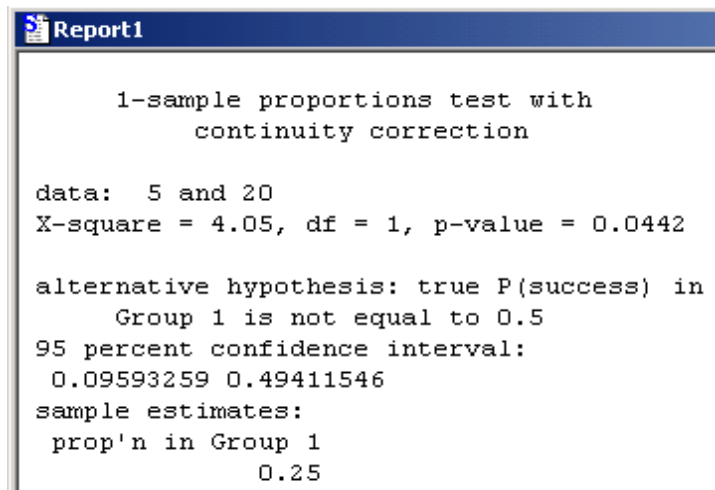
1. Leave the “Data Set” box blank.
2. In the “Success Variable” box, type the number of successes. For example, if there were 5 successes in 20 trials, type 5.
3. In the “Trials Variable” box, type the number of trials. For example, if there were 5 successes in 20 trials, type 20.



The screenshot shows a dialog box titled "Data". It contains three input fields, each with a dropdown arrow on the right. The first field is labeled "Data Set:" and is empty. The second field is labeled "Success Variable:" and contains the number "5". The third field is labeled "Trials Variable:" and contains the number "20".

Figure 0.23: *Inputting data for bootstrapping a single proportion.*

4. On the Bootstrap tab, select “Perform Bootstrap”.
5. Click OK. The results will be printed in a Report window.



The screenshot shows a window titled "Report1" with a blue header bar. The main area contains the following text:

```
1-sample proportions test with
continuity correction

data:  5 and 20
X-square = 4.05, df = 1, p-value = 0.0442

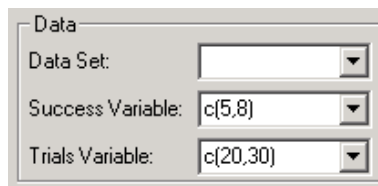
alternative hypothesis: true P(success) in
Group 1 is not equal to 0.5
95 percent confidence interval:
 0.09593259 0.49411546
sample estimates:
prop'n in Group 1
      0.25
```

Figure 0.24: *Result window for a one sample proportion test.*

Bootstrap the difference in proportions

From the menu, select **Statistics > Compare Samples > Counts and Proportions > Proportions Parameters with Bootstrap**.

1. Leave the “Data Set” box blank.
2. In the “Success Variable” box, type $c(a, b)$, where **a** and **b** are the number of successes in each of the cases. For example, if the first case consists of 5 successes in 20 trials, and the second case consists of 8 successes in 30 trials, you would type $c(5, 8)$.
3. In the “Trials Variable” box, type $c(d, e)$, where **d** and **e** are the number of trials in each of the cases. For example, if the first case consists of 5 successes in 20 trials, and the second case consists of 8 successes in 30 trials, you would type $c(20, 30)$.



The image shows a screenshot of the 'Data' dialog box in Minitab. It has three input fields, each with a dropdown arrow on the right. The 'Data Set' field is empty. The 'Success Variable' field contains the text 'c(5,8)'. The 'Trials Variable' field contains the text 'c(20,30)'.

Figure 0.25: *Inputting data for bootstrapping the difference in proportions. The figure shows how to enter the difference between the proportions 5/20 and 8/30.*

4. On the Bootstrap tab, select “Perform Bootstrap”.

5. Click OK. The results will be printed in a Report window. This may take a while because a permutation test is also being performed to give you a P-value.

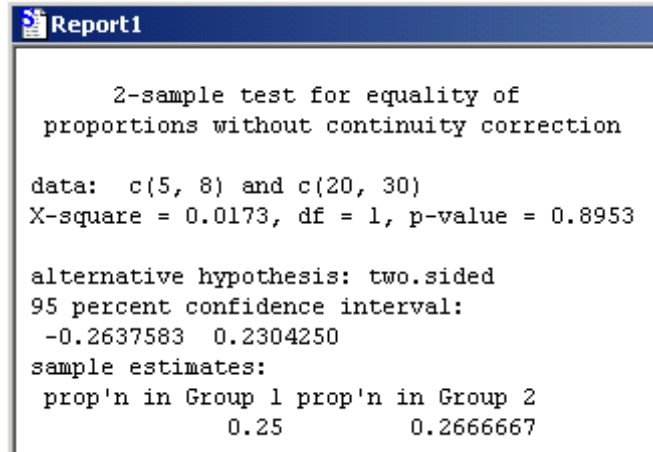


Figure 0.26: Result window for a two sample proportion.

Permutation test for the difference in proportions

To perform a permutation test for the difference in proportions, follow steps 1-3 in the *Bootstrap the difference in proportions* section.

4. On the Main tab, choose an alternative hypothesis, either **two.sided**, **greater**, or **less**.

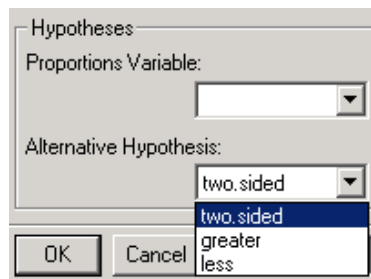


Figure 0.27: Selecting an alternative hypothesis for a permutation test.

5. On the Bootstrap tab, select “Perform Bootstrap”.

****Note** that it is possible to print confidence intervals for the bootstrap along with your permutation test, simply by selecting a confidence interval from the Bootstrap tab.

6. Click OK. The P -value for the permutation test will automatically be printed in the Report window.

Bootstrap the slope and intercept of a regression line

From the menu, select **Statistics > Regression > Linear with Bootstrap**.

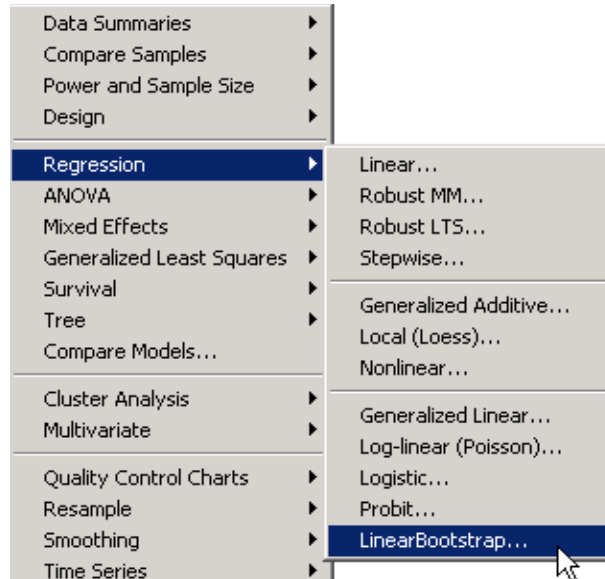


Figure 0.28: *Where to find the Linear with Bootstrap menu.*

1. Type the name of your data set in the “Data Set” box.
2. In the “Dependent” box, select the column name of the dependent variable.
3. In the “Independent” box, select the column name of the independent variable.

4. On the Bootstrap tab, select “Perform Bootstrap”.

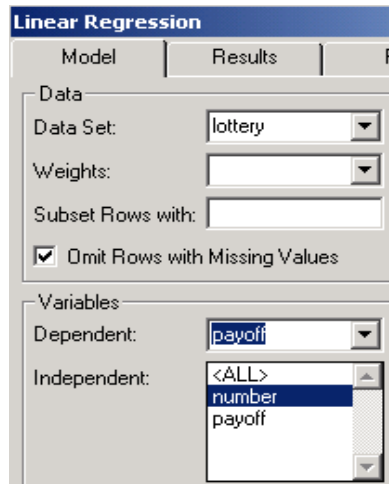


Figure 0.29: *Specifying dependent and independent variables.*

5. Leave the “Statistic” box as the default **coef**. This will bootstrap both the slope and intercept of the regression line.

6. If confidence intervals are desired, check the appropriate boxes in the “Summary Results” section of the menu.

7. Click OK. The results will be printed in the Report window. The top part of the results are for the linear object without bootstrapping. The lower part gives the bootstrap results for the intercept and slope of the regression line. The coefficient of the slope is referred to in the results by the name of the independent variable. The “Correlation of Replicates” table gives the correlation between the resample intercepts and the resample slopes.

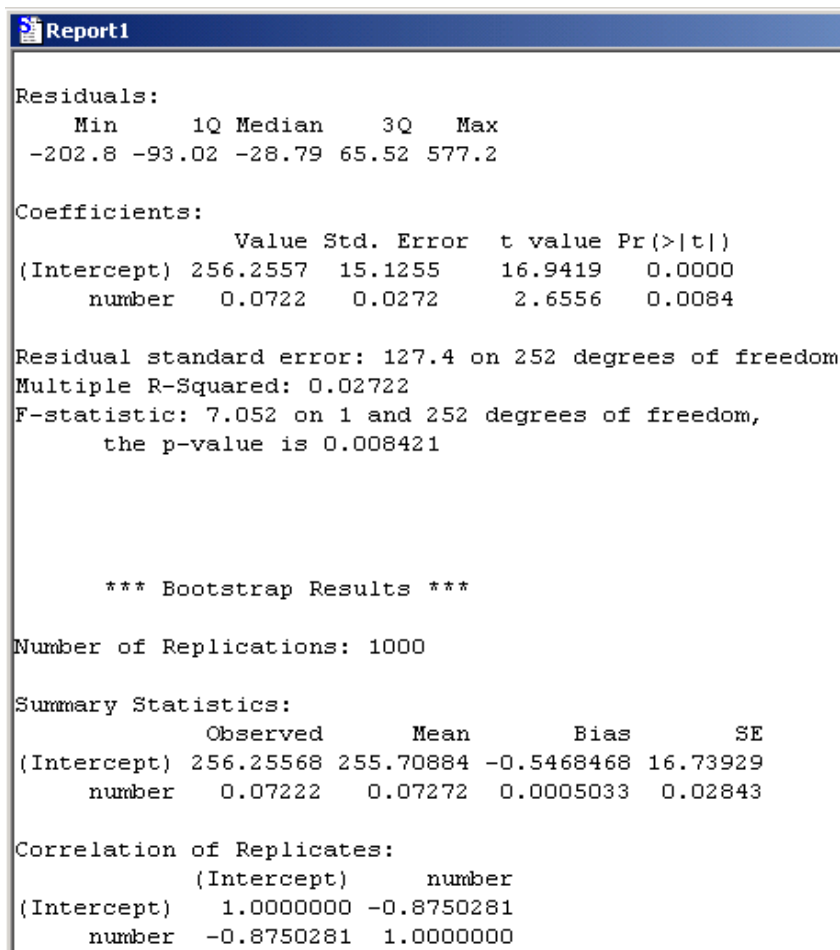


Figure 0.30: Results of bootstrapping the slope and intercept of a regression line. Tilting intervals are shown here.

Bootstrapping complex statistics

One of the strengths of the bootstrap is its ability to handle complex statistics and to find standard errors statistics other than, and including, the mean. These complex statistics include bootstrapping

trimmed means, ratios of proportions, percent increases, etc. To calculate such statistics, open **Statistics > Resample > BootstrapNEW**.

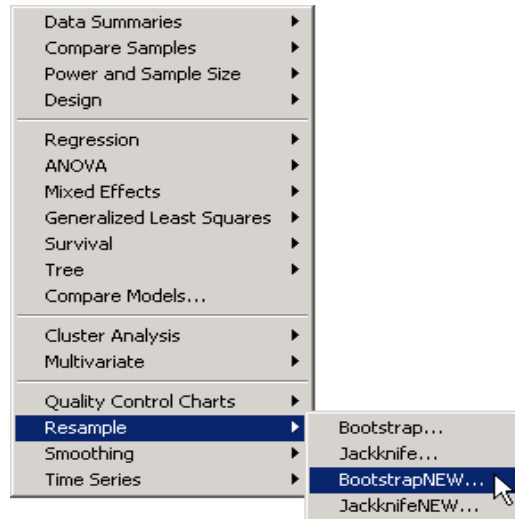


Figure 0.31: *Where to bootstrap complex statistics.*

1. The tabs in this menu offer the same options as does the Bootstrap tab in the Summary Statistics menu. Make specifications in the same manner as described in the *Bootstrapping a statistic and calculating standard error* section.
2. On the Model tab, select the name of your data. In the Statistic to Estimate region, specify the statistic you want to estimate in the “Expression” box. The formulas for complex statistics often look complex themselves.

For the following statistics, your expression would look like...

10% trimmed mean: `mean(name of your variable, trim = .1)`

ratio of variances: `var(name of 1st variable)/var(name of 2nd variable)`

difference of medians: $\text{median}(\text{1st variable}) - \text{median}(\text{2nd variable})$

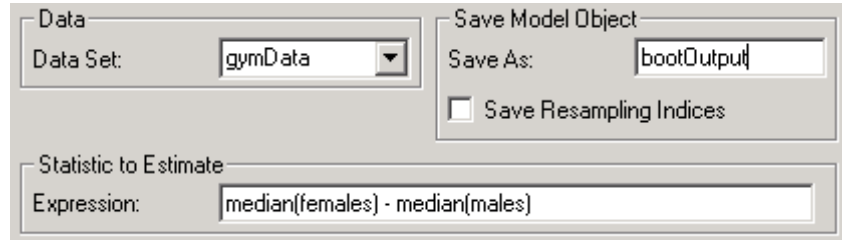


Figure 0.32: *Example of bootstrapping a complex statistic.*

You can refer to the S-PLUS help function to learn how to express other complex statistics.

****Note:** Remember to consider factors such as whether your data is paired or unpaired. For example, if you are performing a two-sample t test, you must select “paired t ” in the Test region of the **2t Test with Bootstrap** menu.

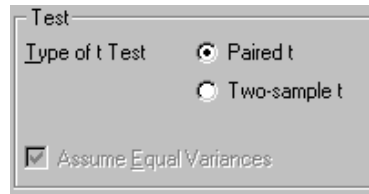


Figure 0.33: *Performing a two-sample t test with paired data.*

Finite Sampling

The process of finite sampling is nearly the same as the bootstrap sampling explained in the *Bootstrapping a statistic and calculating standard error* section, but with two more steps.

1. In the “Sampler” box in the Resampling Options region on the Bootstrap tab, select `samp.finite(size=, bootknife=, N=?)` from the dropdown menu.

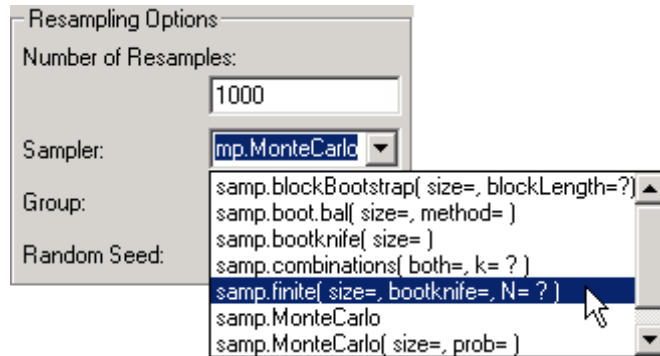


Figure 0.34: *Finite resampling.*

2. Delete the “?” and replace it with the size of the whole population.

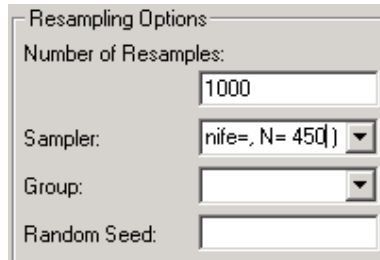


Figure 0.35: *Where to specify the superpopulation size.*

3. Click “OK.” Your results will appear in a *Report* window.

Stratified Sampling

This is similar to the method explained in the *Bootstrapping a Statistic and calculating standard error* section, except there is one extra step.

1. Since stratified sampling takes resamples from groups that exist within the original sample, the “Group” argument on the bootstrap tab must be specified.

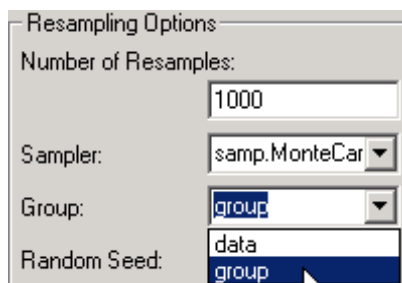


Figure 0.36: *Specifying a group on the bootstrap tab.*

2. Click “OK.” Your results will appear in a *Report* window.

Statistics from the Command line

There are some tests which must be ordered from the command line, such as F tests and permutations tests for correlations or differences in medians. We do not recommend performing tests such as this without further knowledge of S-PLUS.