APPENDIX 2
USING JMP 8.0

I. **Introduction**

These instructions assume that you have a basic familiarity with the use of a Macintosh computer (e.g., using the mouse, pulling down menus, moving windows). If you don't know these things please ask your laboratory instructor for a basic introduction to the computer. In addition, these instructions only introduce the mechanics of operating the elements of the JMP statistics software that you will need for Biology 2 -- you should refer to the JMP Help menu for a more complete description of the graphing and statistical options that are available to you through this program. The Biology 2 Statistical Appendix (Appendix I) describes the rationale for choosing which statistical test to use, how the statistics are calculated, and how to interpret the results.

II. **Getting into the Program**

These instructions are written for use of the Macintosh computers installed in the Biology 2 laboratory. They may vary depending on the type of computer that you are using.
For information regarding how to download this or other available programs, please contact ITS at HELP (4357).

1. Find the folder entitled JMP 8 icon on your computer.
2. To enter the program, use the mouse to move the cursor onto the JMP icon, then double click on it.
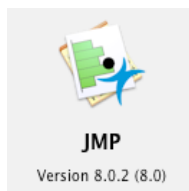
3.The "**JMP Starter Window**" and **"Tips of the Day Window"** should be open when you enter the program (figure 1). Close the "**Tips of the Day Window."** The starter window is the starting point for creating new data files, editing data or beginning any analysis. You may return to the Starter Window at any time by pulling down the **Window** menu and choosing **JMP Starter**.
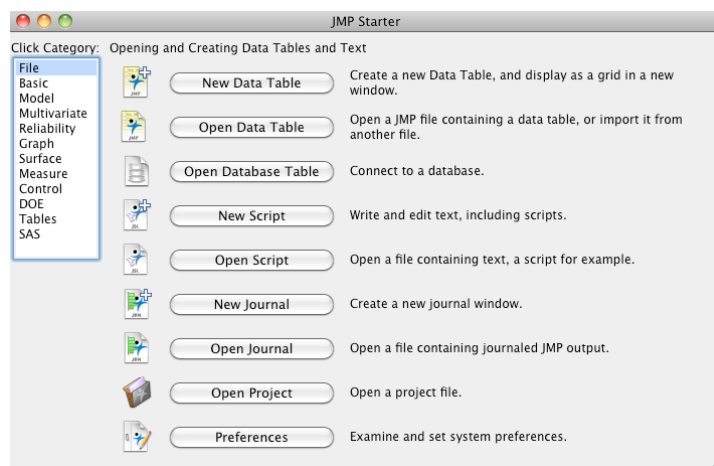

Figure 1. The JMP Starter Window is the starting point for making data files or performing analyses using JMP.

III. Making Data Files

1. Click on the **FILE** tab on the left hand side of the JMP Starter Window to see the options for opening and creating data tables.

2. Click on the **New Data Table** button (figure 1). The program will present you with an empty dataset window (figure 2)**.**
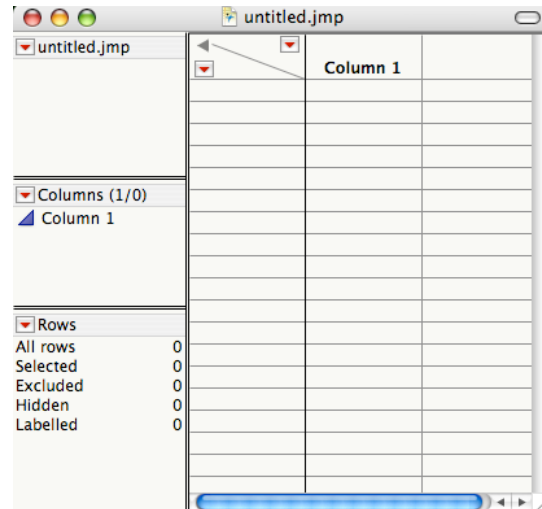

Figure 2. The data table as it appears before columns have been defined.

A. The **Columns** box at the center left of the window defines the columns of the data table. Name your first column of data by clicking on the column name (now labeled "Column 1") and typing in the new column name. You may also change column names by clicking the column heading.

B. Now you must decide what type of data will be entered into the data file. Generally you will use real or integer values:
   - Real numbers include those that have decimal places while integer values are whole numbers without decimals. Data that you measure on a continuous scale like temperature or weight are expressed as real numbers.
   - Data that come as discrete units like elephants or seeds should be expressed as integers.
   - Category data describe states: gender (female versus male), handedness (right versus left), or species.
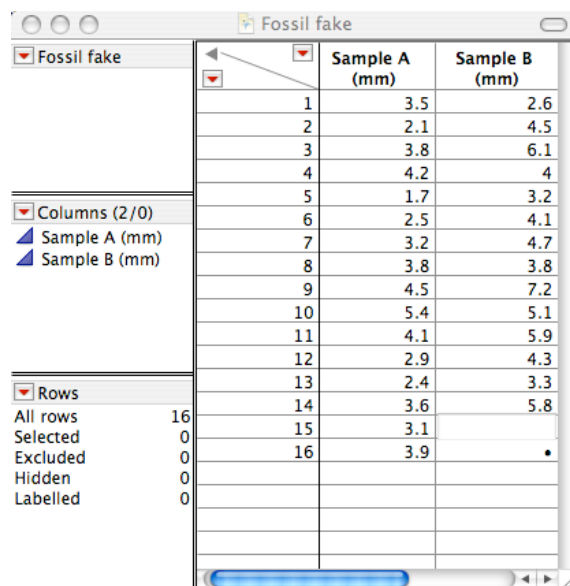
   Click in the purple triangle to the left of the column name (Column 1). A menu will appear to allow you to choose "**Continuous**", "**Ordinal**" (integer) or "**Nominal**" (category) data.

C. Add data columns by choosing **Columns>New Column**. Click on the small red triangle at the top of the box to see the menu. This menu will also allow you to

change the order of the columns in the data table, or to delete columns. You may also add a new column by double-clicking on the empty header of the next column in the data table.

IV. **Entering Data**

1. JMP stores each piece of data in a cell. Click on the cell in which you wish to enter data. It should turn white with a blue border (figure 3).

Figure 3. The white cell with blue border indicates the cell into which data will be entered.

2. Type in the datum for that cell and press "**return**". Continue until you have entered all your data for that column. Your data set may be so large that the screen is not big enough to view it all at once; you can then use the scroll bars at the side and bottom to move the data file within the window. To enter data in the next column, click on the first cell.

3. At any time you may save the contents of your files by pulling down the **File** menu and selecting **Save** (or **Save as** if you wish to rename the file). Remember to give your files clear, descriptive names so that you will be able to easily identify them later as your data.

4. To edit data within the data file, simply move the cursor to the cell you wish to alter and click once, turning the cell white with a blue border and allowing you to type in the new data (figure 4). When you have finished with a cell, you can use the "return" button, the arrow buttons or move the cursor via the mouse to the next location you wish to change. *Remember to save your work frequently!*

Figure 4. Add data to any cell of the table by clicking that cell with the cursor.

V. **Viewing Data: Frequency Distributions**

1. Before you can legitimately use your data in any statistical analyses you must first be sure that they meet the assumptions that the particular statistical tests require. For the statistics we expect you to use in Biology 2, the first step is to examine the distribution of the data. In addition to helping you determine whether the data conform to the assumptions of the statistical test, visualizing the distribution allows you to review the data and may highlight mistakes in data entry. ***Refer to the first few pages of the Biology 2 appendix on statistics (Appendix I) to ensure that you understand how distributions and summary statistics inform your selection of a statistical test <u>before</u> proceeding.***

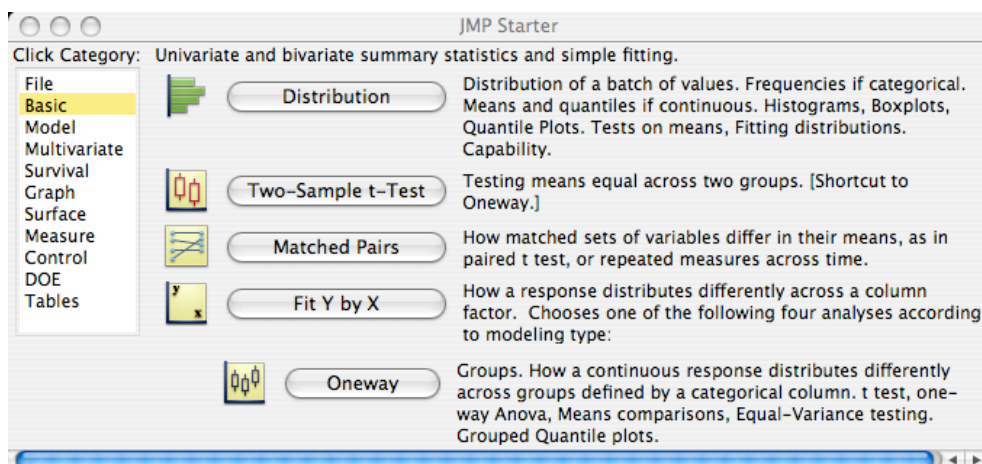2. From the JMP Starter Window click on the **Basic** tab (figure 5).

Figure 5. The Basic tab in the JMP Starter is the page from which you begin analysis.

3. The **Distribution** button opens the "Distribution" window (figure 6).
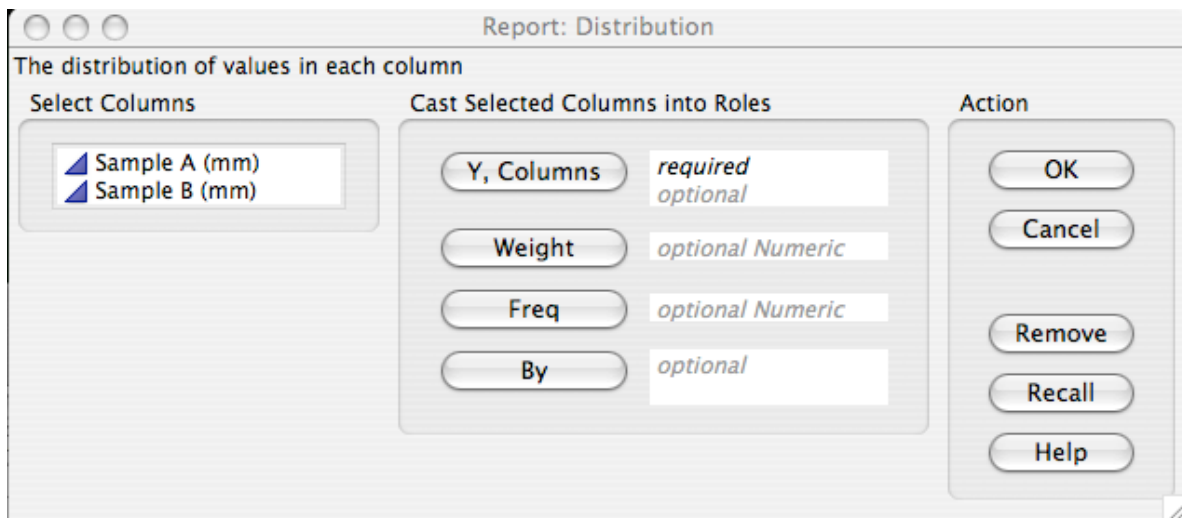
Figure 6. Choose the column of data to examine from the Distribution window.

A. The "**Select Columns**" box at the left of the window shows all of the variables you have entered (column names) and will let you choose which frequency distribution to view.

B. Click on the column name of the desired column and drag it to the **Y, Columns** space in the "**Cast Selected Columns into Roles**" box. You may select more than one column of data if you want to examine the frequency distribution of more than one group of observations.

C. Click **OK** to view the frequency distribution.

4. A report window will appear (figure 7) showing a frequency distribution of your data, with the data grouped into intervals. Each of these "intervals" is shown as a bar in a frequency histogram (***refer to the Biology 2 statistics appendix, appendix I***). The data will also be summarized in a table of "Quantiles" and a table of "Moments" below the histogram.
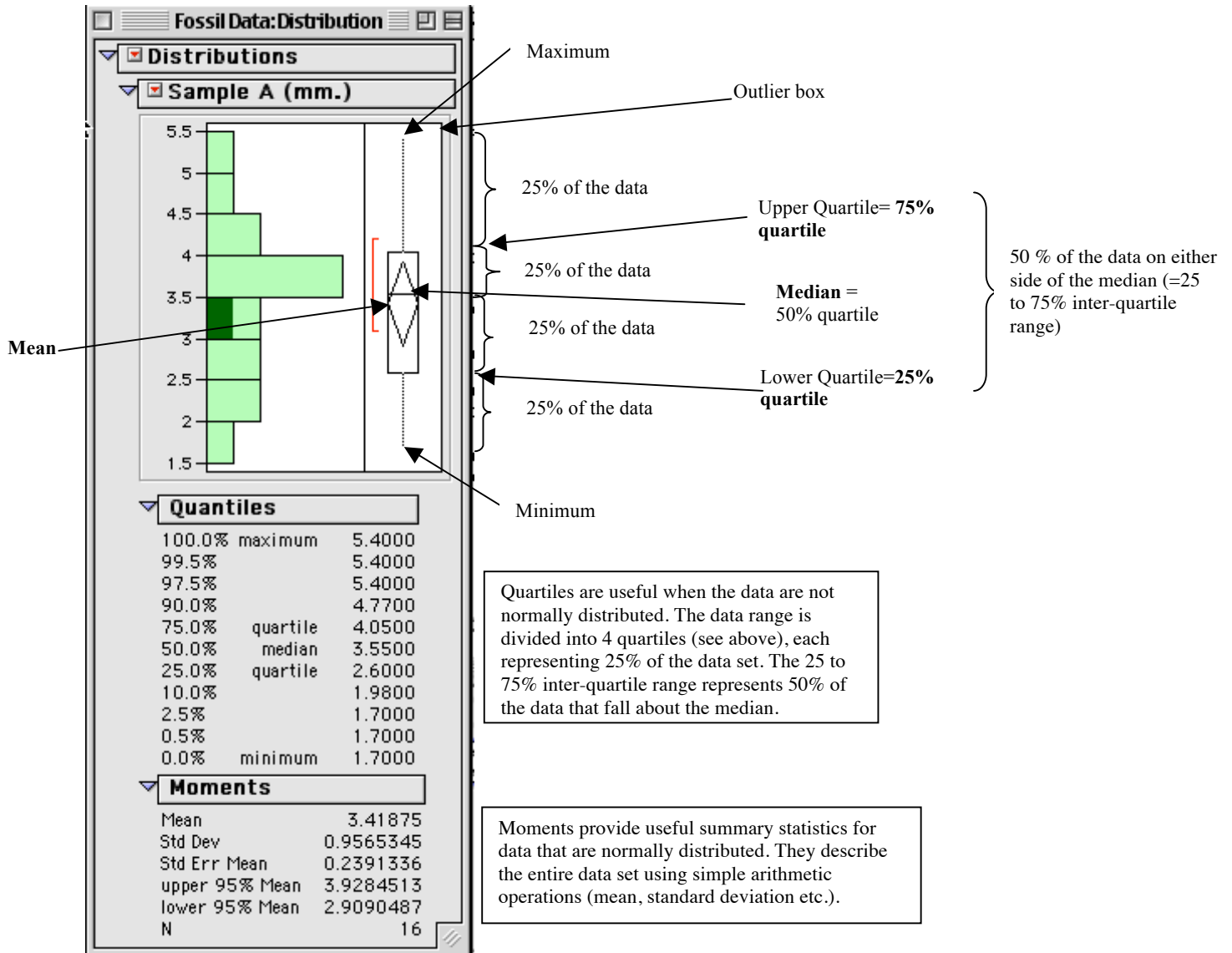

Figure 7. The distribution report as it appears automatically.

5. If you have selected several variables (columns), each will be displayed in a separate bar chart. The window may be resized for more convenient viewing by pulling on the lower right corner of the report.

6. Examine the distribution of the data. Are there any points that don't seem to make sense and might indicate that you have mistyped the data when you entered them into the file? Do the data form a humped pattern, with the high point being somewhere near the center of the distribution and tapering towards the ends? ***Review the Biology 2 statistics appendix, appendix I, to understand the significance of the shapes you see.***

   A. The small red triangle to the left of the column name at the top of the window opens a menu of data summaries. Select **Display Options** to add/remove the tables of quartiles or moments so that you have the descriptive statistics most useful for summarizing your data.

   B. From the pull-down menu you can also remove the **outlier box plot** at the right of the histogram to view only the frequency distribution.

   C. Select **Histogram Options** from the pull-down menu and click on **Show Counts** to add a count axis (x-axis) showing the number of data included in each bar of the histogram.

   D. Double-click on the y-axis to open the "**Y-Axis Specification Window**". From this window you can set minimum and maximum values for the axis and specify the interval width (increment) and number of ticks to be shown on the axis. Another way to specify the interval width is to choose the grabber from the tools menu (figure 8).

Figure 8. The "grabber" tool.

   E. Change the number of intervals in the histogram by grabbing the right side of an interval bar. Move the grabber along the count (X) axis to increase or decrease the number of observations contained in that interval. The intervals will automatically change in the values contained within as your data is grouped into fewer, larger intervals, or more, smaller intervals.

   F. Change the starting value for the intervals by grabbing the top or bottom of an interval bar. Move the grabber over the side of the bar and click the mouse. Drag the grabber to the desired position up or down the value (Y) axis, then release the mouse.

8. If you have more than one histogram, it is useful to have the frequency distributions displayed on the same scale. Click on the small red triangle to the left of "**Distributions**" on the bar at the top of the report. From the menu that appears, choose **Uniform Scaling**.

9. You may save the histograms you have created by pulling down the **Distributions** menu and selecting **Script**. Select **Save Script to Data Table**. The histogram will be saved

with the data file containing the data that are described. To view a saved report, click on the small red triangle to the left of **Distribution** in the upper left box of the data window. From the menu that opens, choose **Run Script**. *NOTE: If you edit or alter the data in your data table, your saved histograms and summary statistics will automatically be updated to include those changes. Save your entire JMP file under another name ("save as") as necessary to avoid such problems.*

To split a column of data

If all of your data are entered directly into a single column of measurements, with another column to indicate the sample to which each observation belongs (as you would set up data to perform an unpaired *t*-test), then you will want to examine separately the frequency distributions of *each* of the two groups you wish to compare.

1. Choose the **Basic** tab in the "**JMP Starter.**" Click on the **Distribution** button.

2. Click and drag the name of the data column into the **Y, Columns** box of the "**Distribution**" window. Click and drag the name of the indicator column into the **By** box. Click on **OK**.

3. The report should show multiple histograms, one for each group of data. The indicator(s) that you used to identify each group will be shown above each histogram.

VI. **Are your data normally distributed? Testing for Normality**

Refer to the statistics appendix (p-I-3) for the definition of a normal distribution. Before beginning data analysis, always test your data against a normal distribution using the **Continuous Fit** command.

1. The small red triangle to the left of the column name at the top of the window opens a menu of data summaries. Select **Continuous Fit.**

2. From the menu that appears, choose **Normal.** A normal curve will be overlaid on the frequency distribution of your data. The normal curve will be calculated using estimates of the mean and standard deviation based on the sample data (figure 10).
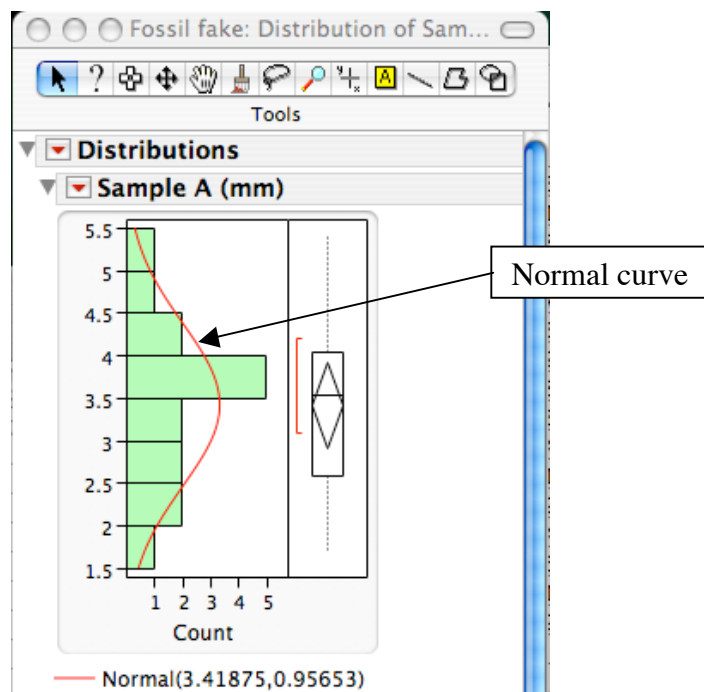
Figure 10. A normal curve, drawn using the estimates of the mean and standard deviation calculated from your sample, can be overlaid on the frequency distribution of your sample for comparison.

3. Click on the small red triangle to the left of the "Fitted Normal" bar that is added to the report below the "Moments" report. From the menu that opens, choose **Goodness of Fit.**

4. **JMP** will calculate a Shapiro-Wilk W test to compare between the distribution of your data and the calculated normal distribution.

5. A **Goodness-of-Fit** table is added to your report (figure 11). Two values are given that should be noted: the **W-value** and the **p-Value** (Prob<W). These statistics quantify the comparison between your data and the calculated normal distribution. **If p > 0.05,** then your data **are** normally distributed. *Review the section on interpreting the results of statistical tests in the Biology 2 statistics appendix (Appendix I-5). **If p < 0.05,** then your data **are not** normally distributed.

Figure 11. The "Fitted Normal" table added to the report compares the sample distribution to the calculated normal distribution.

VII. **Directions for a Small Sample of Useful Statistics**

COMPARING TWO SAMPLES

Student's *t*-Test - Unpaired (two-sample) *t*-Test

If you want to test to see if two samples represent populations that are different from each other (e.g. are the fossils from Sample A of different length than those from Sample B), a common test that is used is the *t*-test. ***Refer to the Biology 2 statistics primer (Appendix I-9) for a description of this statistical test and the circumstances under which it can be used correctly.***

Data setup for the unpaired (two-sample) *t*-Test

These directions assume that you have already named your columns and entered your data. You also need a column containing information that will make it possible for the program to distinguish between the data sets in the column of measurements. You should pick a name that indicates this, say something like "A, B indicator". The grouping variable must be a nominal variable: click on the small box to the left of the column name in the "Columns" box and select "**Nominal.**"

You may use "1" and "2" to indicate the 2 categories. Using integers as a grouping variable is slightly simpler to enter into the computer, but because this notation will appear on the analysis that you generate, be sure to note which sample is indicated by "1" and which by "2". You may also use letters such as "a" and "d" or category names such "alive" and "dead" to label the data. When you type in a letter, a window will open asking if you want the column to contain character data rather than numeric data (figure 12). Click on the change button.
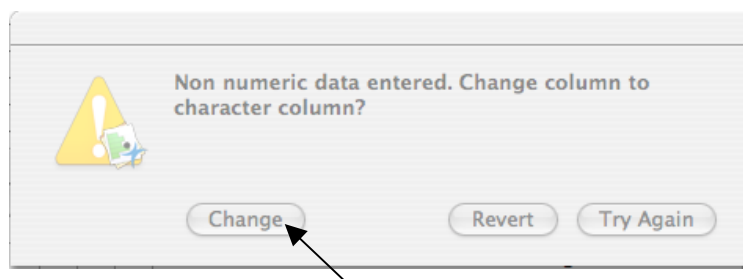

Figure 12. Change the column to a character column if you want to use categories to indicate the samples from which data were taken.

From the Basic tab of the JMP Starter

1. All measurements to be compared must be in a single column of data, with a grouping (Nominal) variable to indicate the sample to which each datum belongs.

2. In the JMP starter window (Basic category) click on the **Oneway** button to open the window allowing you to select columns for the analysis (figure 13).
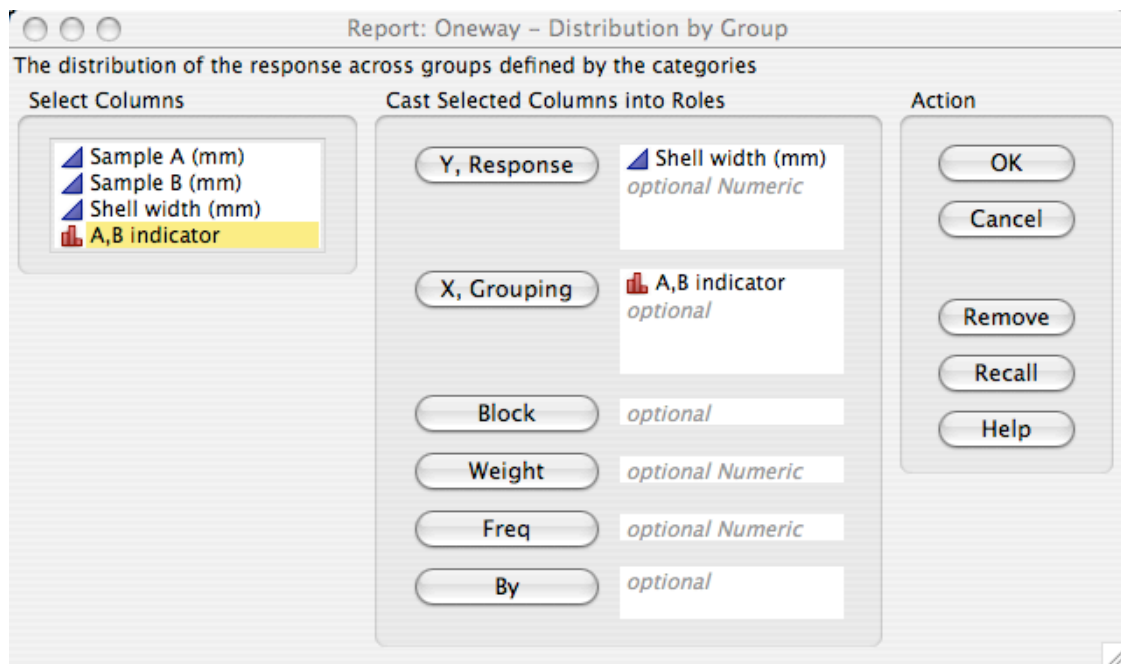
Figure 13. The Oneway analysis window.

A. Click on the name of the column containing your measurements and drag it to the **Y, Response** space in the "**Cast Selected Columns into Roles**" box.

B. Click and drag the name of the column with the indicator variable into the **X, Grouping** box. This variable should have a series of small red bars to the left indicating that it is nominal variable.

C. Click on **OK**.

3. The window that appears shows your data as a scatterplot (figure 14). Click on the red triangle to the left of the title bar ("**Oneway Analysis of Shell width (mm) By A, B indicator**") to see the analysis menu.
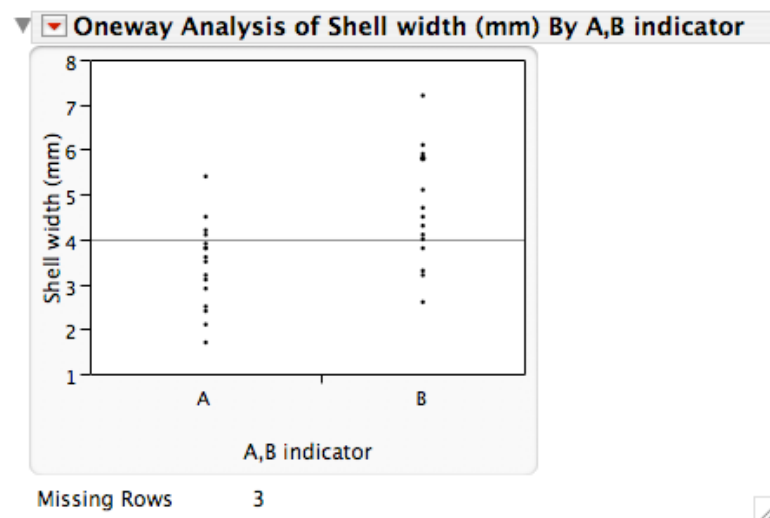
Figure 14. The default oneway analysis: a scatterplot of the data.

If your data are normally distributed:

   A. From the pull-down menu, choose **Means/ANOVA/Pooled t**.

   B. The results of the *t*-Test will be shown along with other statistics (figure 15).

       Note: JMP calculates two t-Tests: assuming equal variances and assuming unequal variances. **Review the meaning of "variance" in the Biology 2 statistics appendix (I-4).** *The Means/ANOVA/Pooled t option will show the calculations assuming equal variances. You can display the calculations assuming unequal variances by selecting the "t test" option from the one-way analysis pull-down menu. The P-values for the two t-Tests will be similar if the variances of your two samples are equal. If the P-values are dissimilar, then choose the more conservative assumption that the variances are not equal.*

   C. The values that you will need to report are the ***t*-value (t Ratio)**, the degrees of freedom (**DF**) and the probability (**Prob > |t|**). ***Refer to the Biology 2 statistics appendix, appendix I, or contact your laboratory instructor in order to interpret these statistics.***
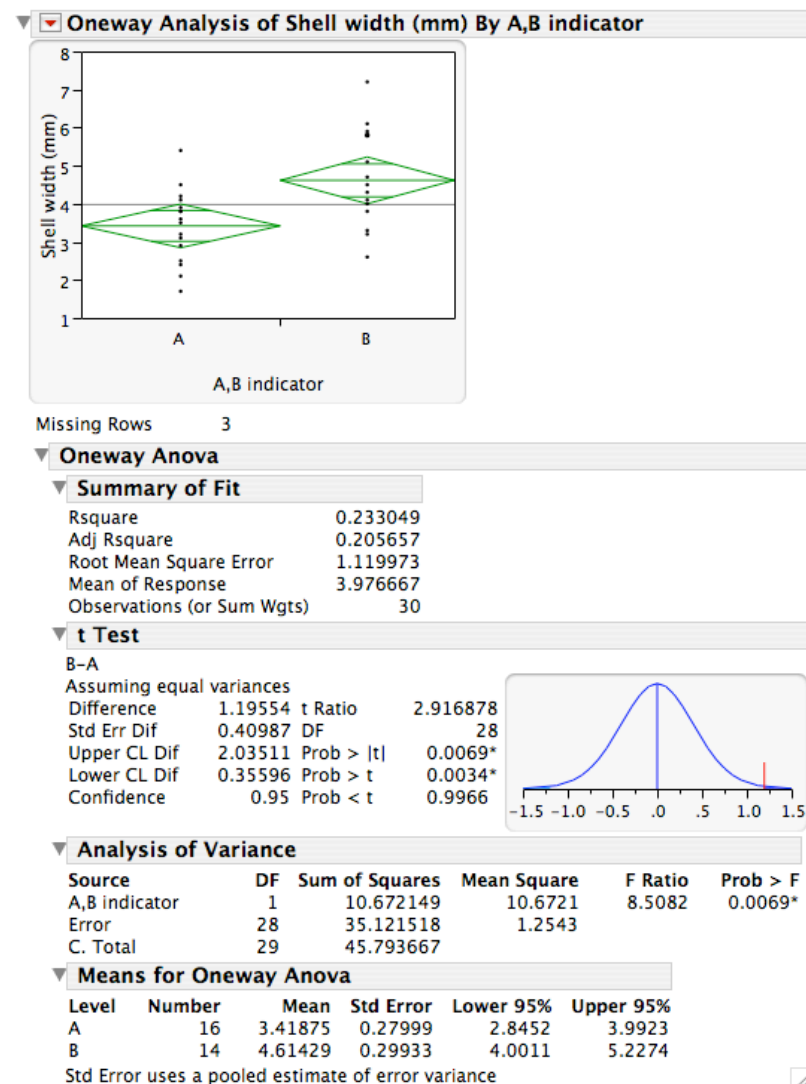
Figure 15. The results of the unpaired *t*-Test.

4. You may save the report you have created by pulling down the **Oneway Analysis...** menu and selecting **Script**. Select **Save Script to Data Table**. The analysis will be saved with the data file containing the data that are compared. To view a saved report, click on the small box to the left of **Oneway** in the upper left box of the data file window (Figure 16). From the menu that opens, choose **Run Script**. NOTE: *NOTE: If you edit or alter the data in your data table, your saved data analysis will automatically be updated to include those changes.*

Figure 16. To view a saved report, click on the small box to the left of **Oneway** in the upper left box of the data file window.

Non-parametric comparison – Wilcoxon rank-sum test

If your data are **not normally distributed**, then a non-parametric test is the appropriate way to compare between groups. *Refer to the Biology 2 statistics appendix (I-14) for a brief explanation of the way these tests compare groups.*

Data setup for the Wilcoxon rank-sum test

All measurements to be compared must be in a single column of data, with a grouping (Nominal) variable to indicate the sample to which each datum belongs (see p. II-10 for instructions to edit data into a single column).

From the Basic tab of the JMP Starter

1. Click on the **Oneway** button to open the window allowing you to select columns for the analysis (figure 17).
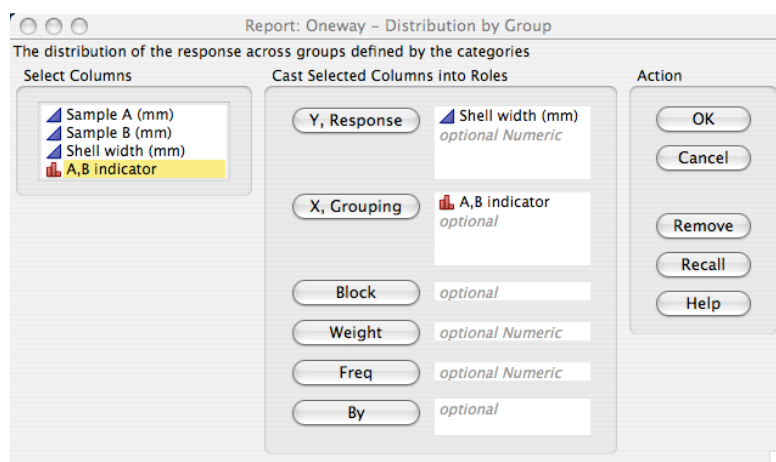

Figure 17. The Oneway analysis window.

A. Click on the name of the column containing your numerical measurements and drag it to the **Y, Response** space in the "**Cast Selected Columns into Roles**" box.

B. Click and drag the name of the column with the indicator variable into the **X, Grouping** box. This variable should have a series of small red bars to the left indicating that it is nominal variable.

C. Click on **OK**.

2. The window that appears shows your data as a scatterplot (figure 18). Click on the red triangle to the left of the title bar ("Oneway Analysis of shell width (mm.) By A, B indicator") to see the analysis menu.
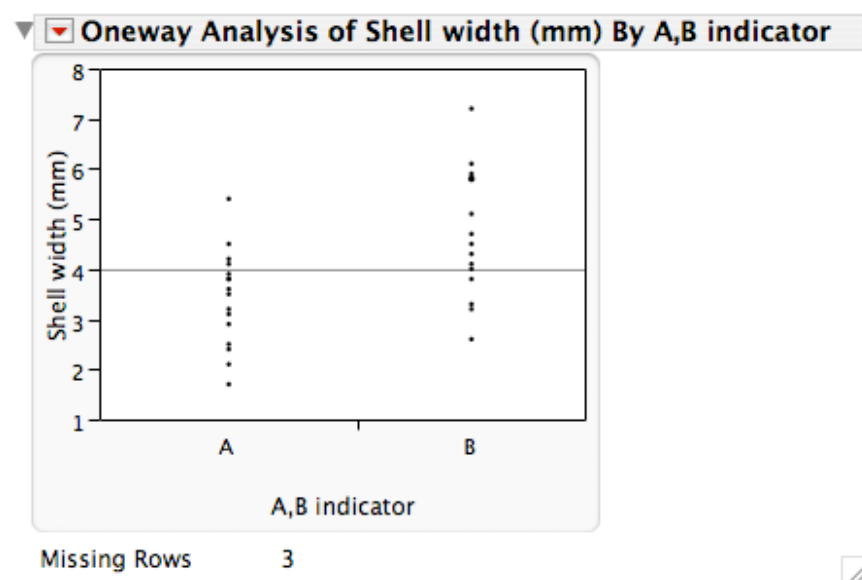

Figure 18. The default Oneway analysis: a scatterplot of the data.

A. From the pull-down menu, choose **Nonparametric**.

B. From the menu that opens, choose **Wilcoxon Test**.

C. The results of the Wilcoxon Rank-Sums test will be shown along with other statistics (figure 19). Because you are comparing two samples, refer to the results bar titled "2-Sample test, Normal Approximation". The values that you will need to report are the **Z-value** and the probability (**Prob > |Z|**). *Refer to the Biology 2 statistics appendix, appendix I, or contact your laboratory instructor in order to interpret these statistics.*
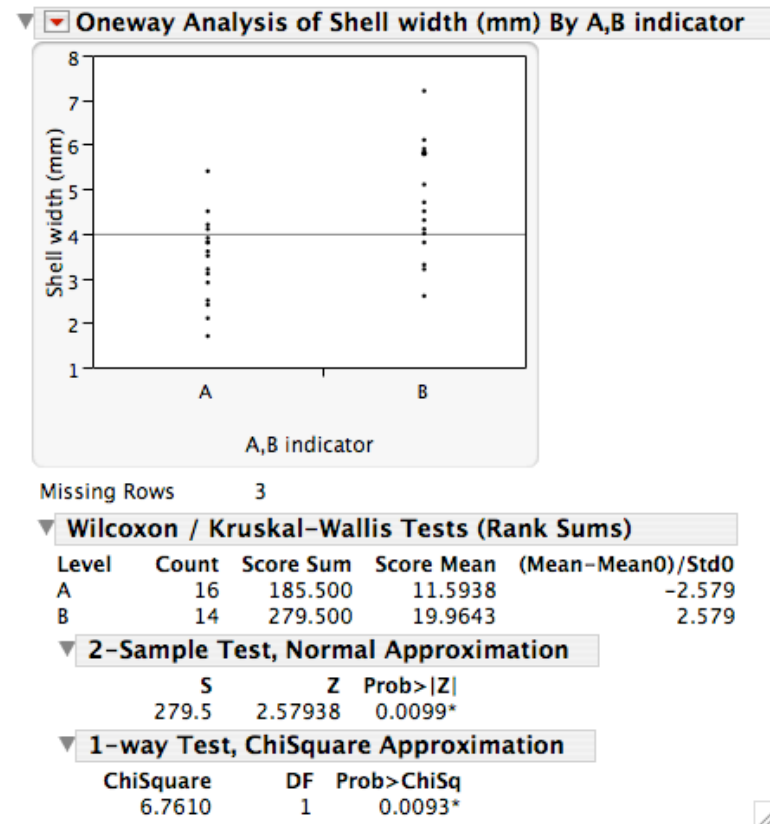
```
▼ ▼ Oneway Analysis of Shell width (mm) By A,B indicator
```

Figure 19. The results of a non-parametric comparison between two samples.

Student's *t*-Test - Paired *t*-Test

If you want to test to see if two treatments applied to the same sample give results that are different from each other (e.g., the running speeds of lizards at two different temperatures), a paired *t*-test is the appropriate analysis. ***Refer to the Biology 2 statistics appendix (Appendix I-13) for a description of this statistical test and the circumstances under which it can be correctly used.***

Data setup for the paired *t*-test

1. Enter your data in columns. Each row should correspond to one individual, and each column to a treatment (figure 20).

Figure 20. Data in each row of the table are repeated measurements on individual test animals.

2. Create another column that calculates the difference between the two responses that you wish to compare.

   A. Double click on the empty column header at the end of the data table. Give the new column a clear descriptive name.

   B. Pull down the **Cols** menu and choose "**Formula...**"

   C. A dialog box will open (figure 21). In the "**Table Columns**" box at the left, click on the name of the column containing the values of the first response measured. Then click on the minus sign in the block of arithmetic functions in the center. Click on the name of the column containing the values of the second response. Click the "**Apply**" button to calculate the differences between the responses.

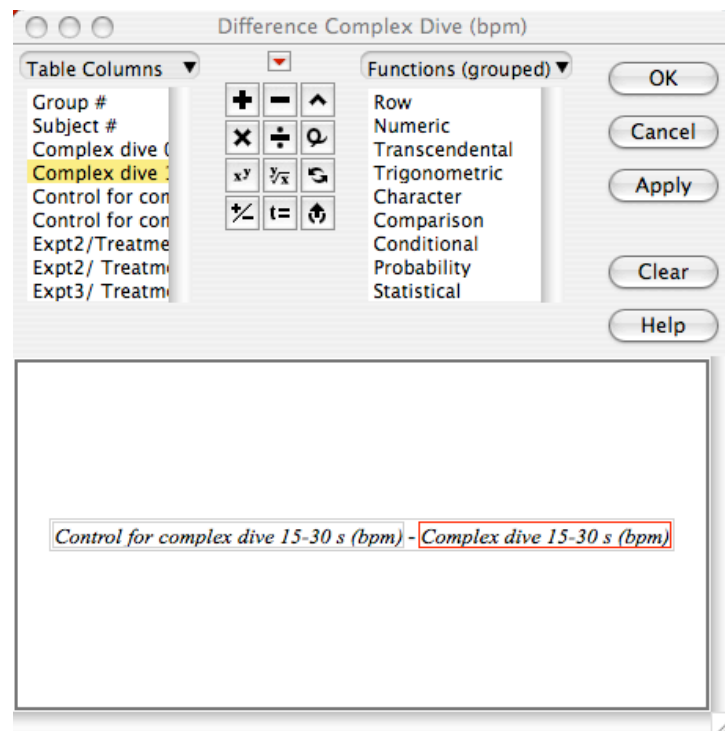Figure 21. The dialog box in which formulas may be specified to calculate from columns of data.

   D. Click on "**OK**" to leave the formula window.

From the Basic tab of the JMP Starter

1. Examine the distribution of the differences using the **Distribution** platform of the **Basic** tab of the JMP Starter *(see "Viewing Data -- Frequency Distributions", page 5, for detailed instructions).*

A. Cast the column containing the calculated differences between measurements into the **Y, Columns** box (figure 22). Click on **OK**.
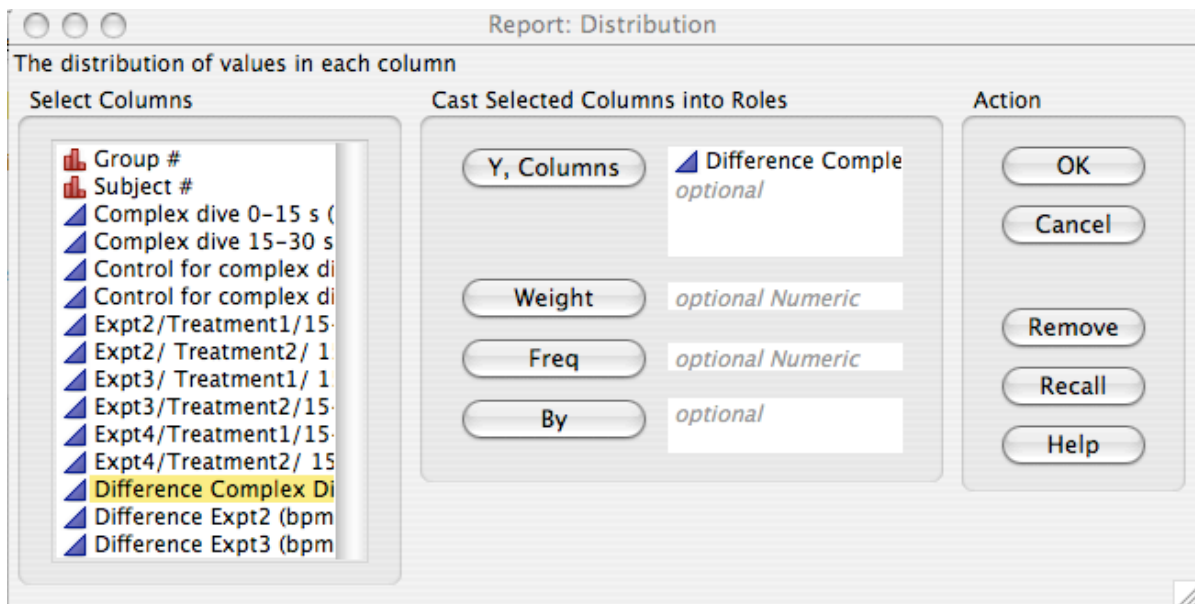

Figure 22. To view the frequency distribution of the calculated differences between pairs of measurements, choose the column of calculated values as the response.

B. The report that appears will show the frequency distribution of the differences (and also the "quantiles" and "moments" of the distribution). The paired *t*-test assumes that the differences are normally distributed. ***Refer to the Biology 2 statistics appendix, appendix I, to review the normal distribution concept.***

2. Test the hypothesis that the mean of the differences between measurements on your subjects is zero, from the report of the distribution.

A. Click on the small red triangle on the bar containing the name of your column of differences.

B. From the menu that appears, choose "**Test Mean**".

C. In the window that opens, specify the hypothesized mean to which you wish to compare the sample mean. If the null hypothesis that you are testing is that there is *no difference* between the paired measurements, then you may accept the default that the hypothesized mean is zero (figure 23).
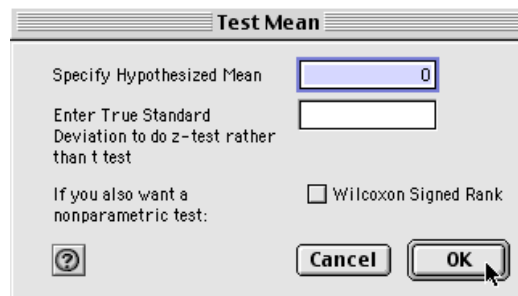
Figure 23. Accept the default value of zero for the hypothesized mean to test the hypothesis that there is no difference between paired measurements.

D. Click "OK".

3. The results of the paired *t*-test will be added to the end of the report under the **Test Mean = value** bar (figure 24). The **df** (degrees of freedom) should be the number of complete data pairs (i.e., the number of individuals)-1. If any data are missing (that is, some of the pairs are not complete), the analysis will omit the pair entirely. The "**Actual Estimate**" is the difference between the two treatments, averaged for all of the pairs.

JMP supplies a table of P-values for the paired *t*-test. Use **Prob > |t|** as the P-value representing the testing of the hypothesis that the sample mean is not different from zero.

Figure 24. The table added to the report contains the descriptive statistics (mean difference and standard deviation), and the comparative statistics (*t*-value, df, P-value) for the paired *t*-test.

Non-parametric comparison -- Wilcoxon signed rank test

If, after examining the frequency distribution of the *differences* between pairs of measurements, you find that the paired *t*-test is not an appropriate comparison, you may run a

non-parametric test called the **Wilcoxon signed rank test**. *Refer to the Biology 2 statistics appendix (I-14) for a brief explanation of the way this test differs from a **t**-test.*

1. Examine the frequency distribution of the differences (refer to instructions above for the paired *t*-test). Click on the small red triangle on the bar containing the name of your column of differences (*e.g.*, "**Difference B-A**"). From the menu that appears, choose "**Test Mean**".

2. From the "**Test Mean**" window, click on the box next to "**Wilcoxon Signed Rank**" (figure 25).
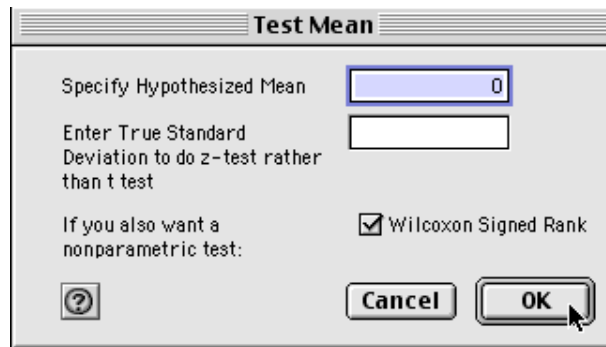
Figure 25. Accept the default value of zero for the hypothesized mean to test the hypothesis that there is no difference between paired measurements and select the Wilcoxon Signed Rank option.

3. Click "OK".

The results of a paired *t*-test *and the Wilcoxon Signed Rank test* will be added to the end of the report under the **Test Mean = value** bar (figure 26). The **df** (degrees of freedom) should be the number of complete data pairs (i.e., the number of individuals)-1. If any data are missing (that is, some of the pairs are not complete), the analysis will omit the pair entirely. The "**Actual Estimate**" of the mean difference between the two treatments will also be supplied, but remember that if the differences are *not* normally distributed, this estimate may not accurately describe the data.

Use **Prob > |t|** as the P-value for the test of the hypothesis that the sample mean is not different from zero.

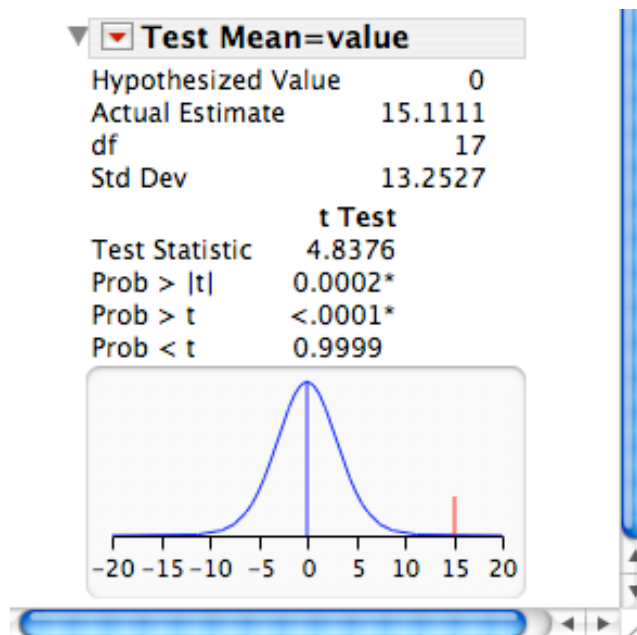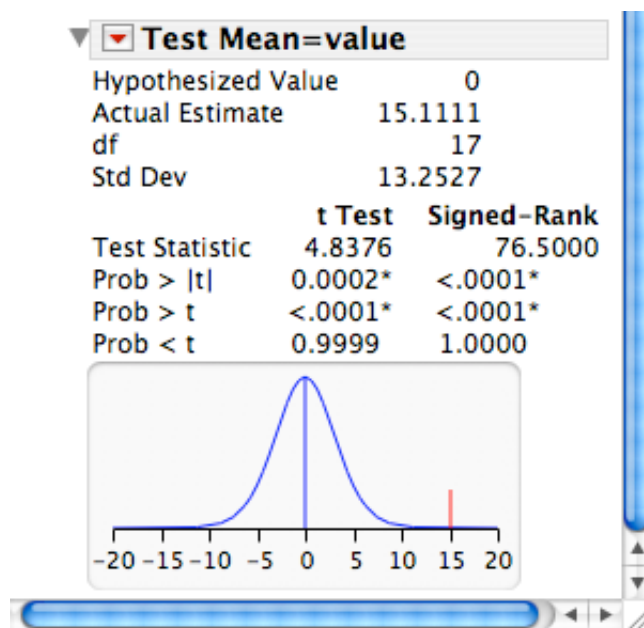Figure 26. The table added to the report contains the test statistic, the degrees of freedom, and the P-value for the Wilcoxon signed rank test, along with other values.

CHI-SQUARE TESTS

This type of statistical test is used when the data are of the "frequency type". The Chi-square $(=\chi^2$) is computed in two different statistical tests: "goodness of fit" which is used when you want to compare one sample with a hypothetical distribution and "contingency table" that is used to test whether a set of variables are related to or associated with each other. ***Refer to the Biology 2 statistics appendix (I-7) for description of this statistical test and the circumstances under which it can be correctly used.***
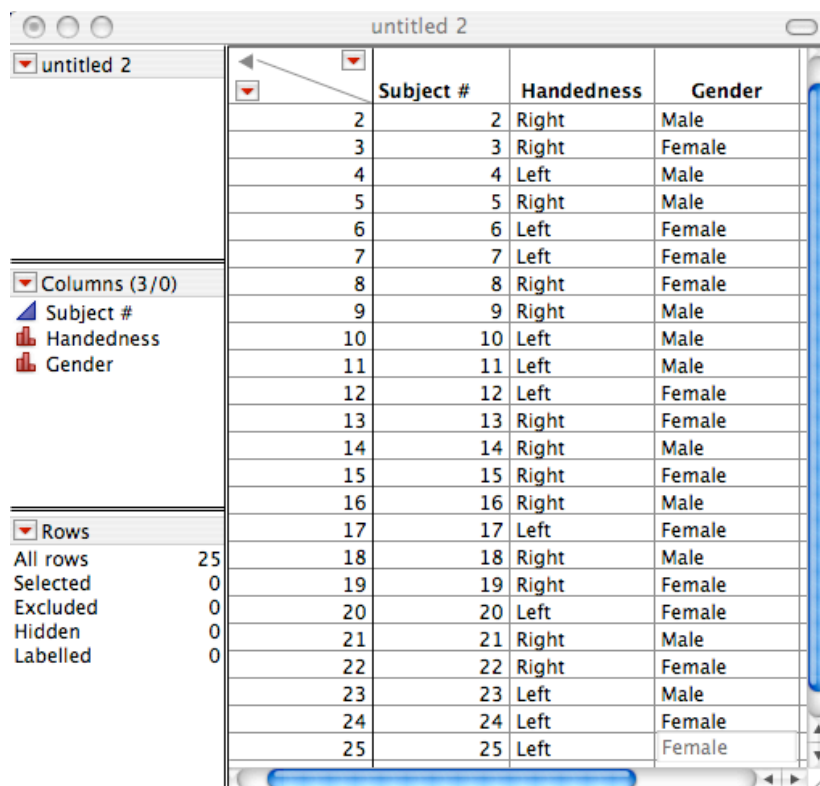
Goodness-of-fit

You may calculate the Chi-square statistic for the goodness-of-fit test using a calculator, as described in the Biology 2 statistics appendix (I-7). JMP will not calculate this simple test.

Contingency Table - Chi-square Test

Data setup for the Contingency Table

Data should be entered as a table of observations (figure 27). To test a hypothesis that there is an association between two categorical variables (e.g., an association between gender and handedness), there must be *two* nominal or ordinal grouping variable columns for each observation.

Figure 27. The dataset recording handedness and gender of students in a laboratory section of Biology 2.

From the Basic tab of the JMP Starter

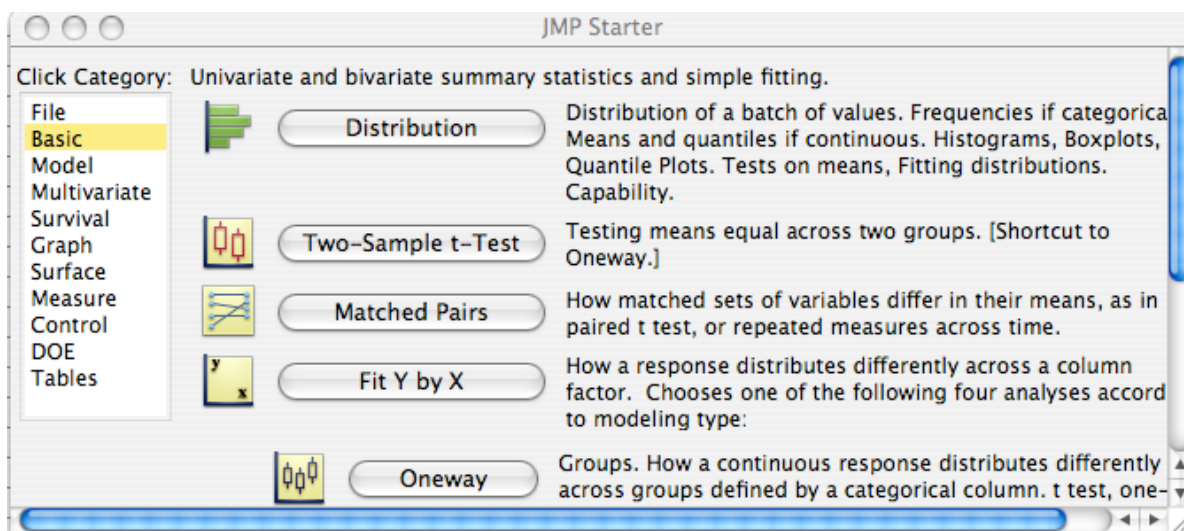1. From the JMP Starter Window click on the **Basic** tab (figure 28).

Figure 28. The Basic tab in the JMP Starter is the page from which you begin analysis.

2. The **Fit Y by X** command allows four options, listed below the command on the **Basic** tab. Click on the **Contingency** button. You may get to the same window by pulling down the **Analyze** menu and selecting **Fit Y by X**.

3. The data columns recording the grouping variables for which you wish to test for association must be designated as *nominal* or *ordinal* variables (figure 29). Cast one column as **Y, Response** and the other as **X, Factor**. Click on **OK**.
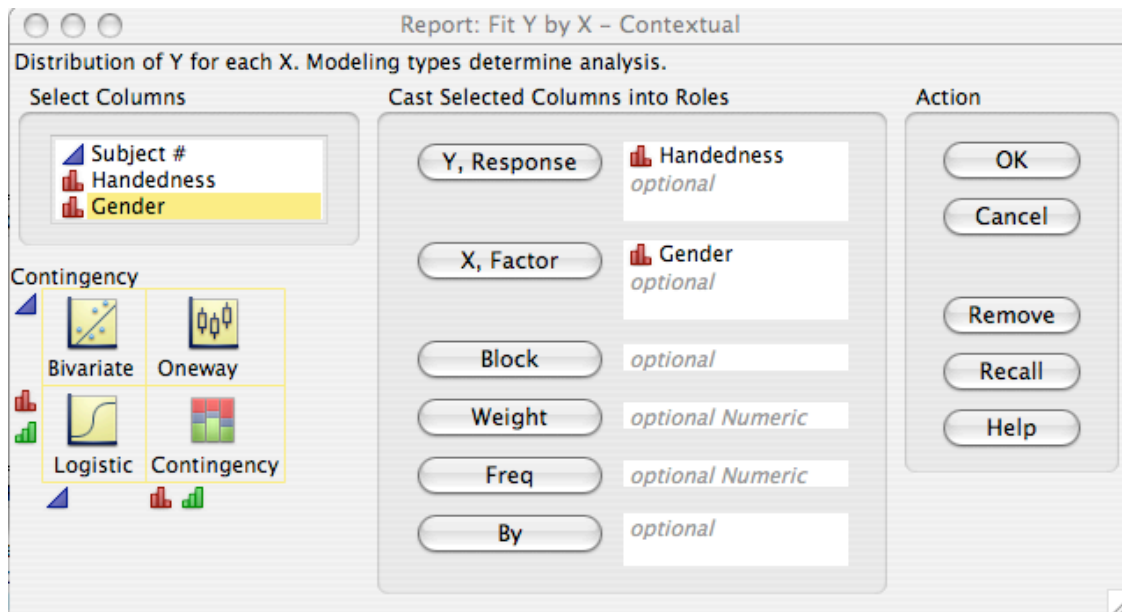
Figure 29. Choosing two nominal variables as factor and response in the **Fit Y by X** command will lead JMP to test for association between those variables.

4. The window that appears shows by default a mosaic plot of the proportion of observations across the two categorical variables, and a contingency table containing the frequency of observations in each cell of the table and also proportions (figure 30). The pull down menu of the **Contingency Analysis of ___ By ___"** bar will allow you remove the plot.

Figure 30. The default analysis for the contingency table. The mosaic plot shows
the observations as proportions.

5. Pull down the menu on the **Contingency Table** title bar by clicking on the small red triangle. From the menu, choose the statistics that will be most useful to summarize and interpret your observations: choose **Count** to display the frequency of observations, choose **Expected** to see the calculated expected values for the frequency, and ch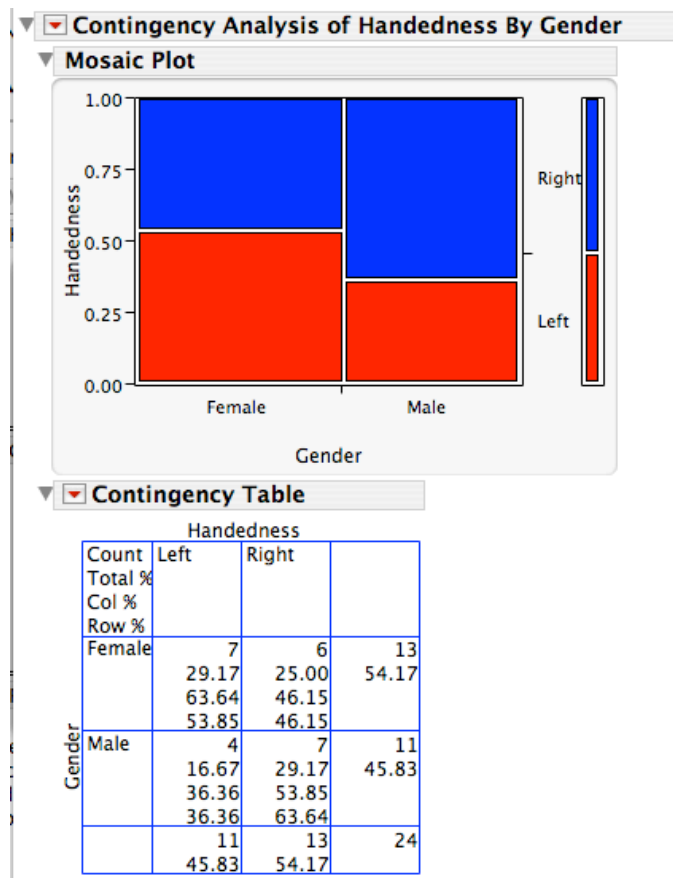oose **Cell Chi Sq** to see the contribution to the total chi square value from each cell. Remove other statistics from the table. *Refer to the Biology 2 statistics appendix, appendix I, to help you interpret this table.*

6. Check that the expected values in each cell are five or greater before proceeding.

7. Choose **Tests** from the pull down menu of the **Contingency Analysis…** bar. The Pearson Chi Square test is the test for association described in the Biology 2 statistics appendix, appendix I. The values that you will need to report are the Pearson **Chi Square** value and the Pearson probability (**Prob > Chi Sq**). The degrees of freedom for the test may be found in the table shown below, as the model degrees of freedom (for example, in figure 31, df=1). *Refer to the statistics appendix, appendix I, or contact your laboratory instructor in order to interpret these statistics.*
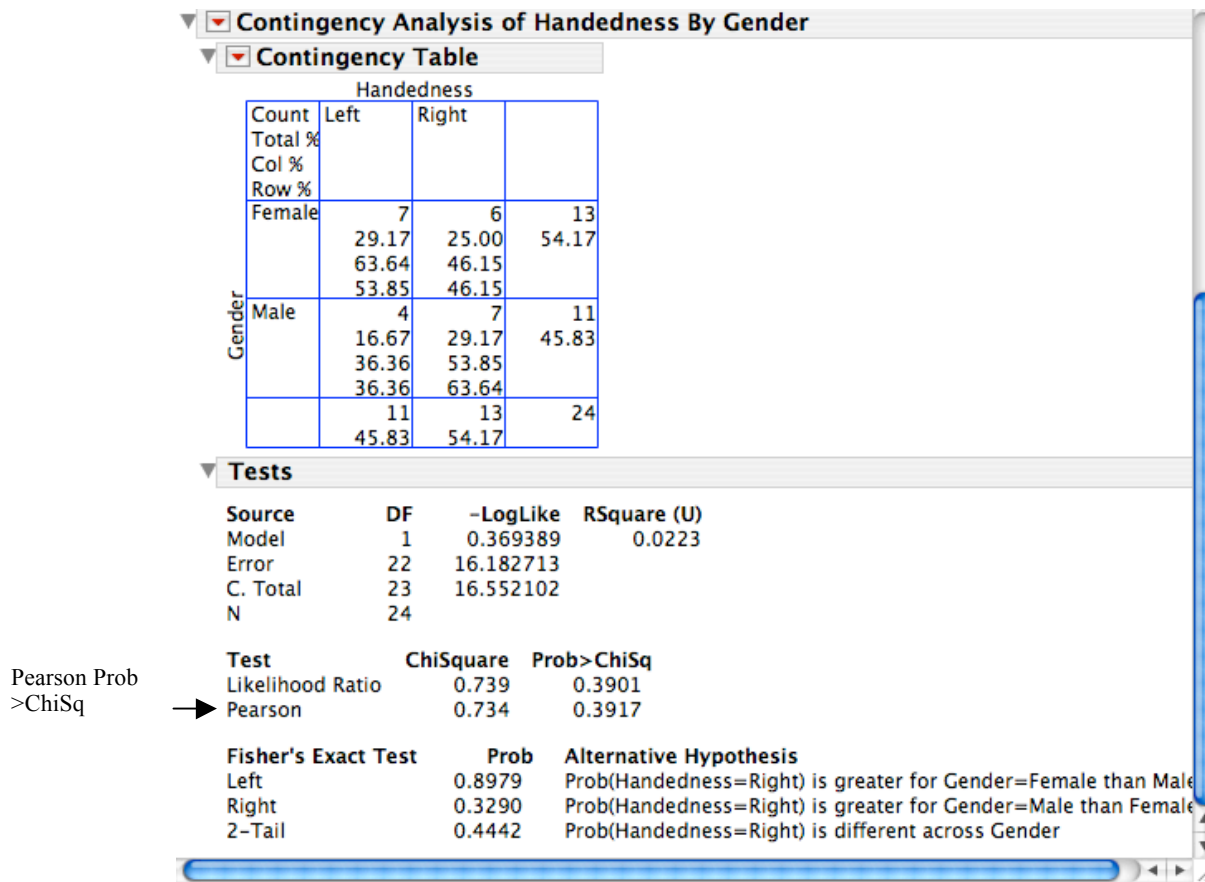
▼ ☑ **Contingency Analysis of Handedness By Gender**
　▼ ☑ **Contingency Table**

Count Total % Col % Row %	Handedness		
	Left	Right	
Female	7 29.17 63.64 53.85	6 25.00 46.15 46.15	13 54.17
Male	4 16.67 36.36 36.36	7 29.17 53.85 63.64	11 45.83
	11 45.83	13 54.17	24

(Row label "Gender" appears vertically beside Female/Male.)

▼ **Tests**

Source	DF	−LogLike	RSquare (U)
Model	1	0.369389	0.0223
Error	22	16.182713	
C. Total	23	16.552102	
N	24		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.739	0.3901
Pearson	0.734	0.3917

Pearson Prob >ChiSq → Pearson

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.8979	Prob(Handedness=Right) is greater for Gender=Female than Male
Right	0.3290	Prob(Handedness=Right) is greater for Gender=Male than Female
2-Tail	0.4442	Prob(Handedness=Right) is different across Gender

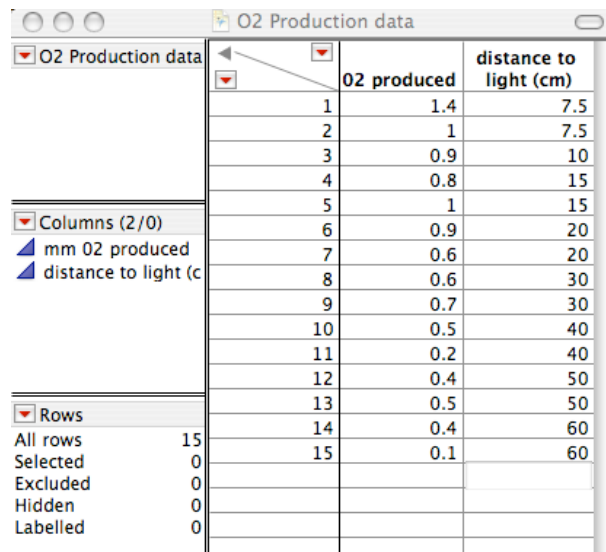Figure 31. The completed contingency table analysis.

CORRELATION ANALYSIS

Correlation Analysis

This test allows you to calculate the degree of linear relationship between two variables. ***Review the Correlation analysis section in the statistics appendix (I-19) for a more complete description of this statistic.***

Data setup for Correlation

Data are entered in columns. You must designate the data as continuous variables (figure 32).

Figure 32. Two continuous variables are necessary to test for correlation.

From the Multivariate tab of the JMP Starter

1. Although "Correlation" appears on the list of options available from the **Bivariate** command on the **Basic** tab, you must start from the **Multivariate** tab to correctly perform a correlation analysis.

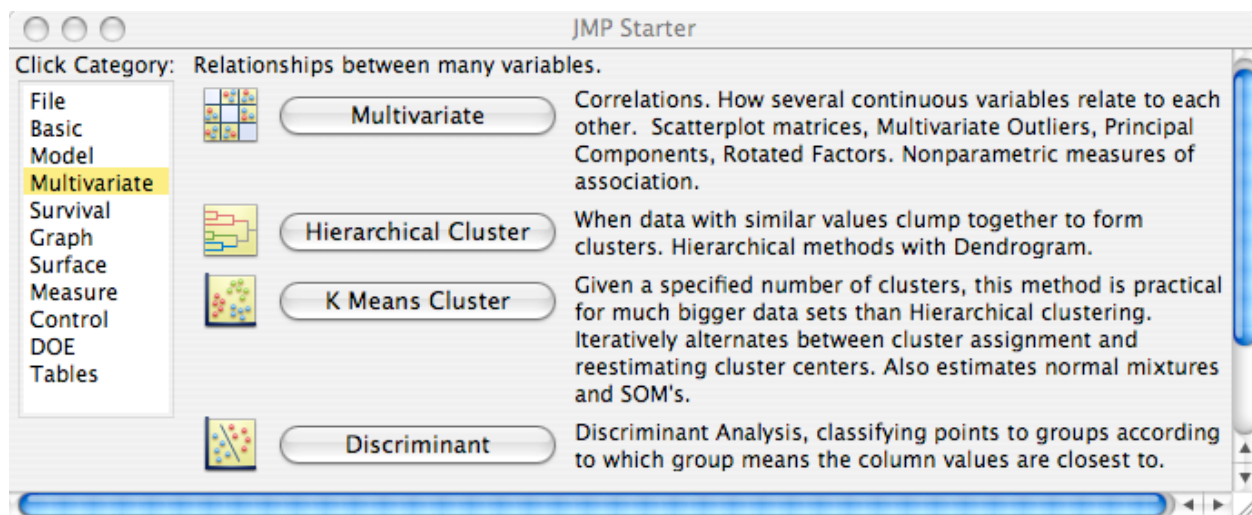2. Click on the **Multivariate** button (figure 33).

Figure 33. The Multivariate tab of the JMP Starter.

3. In the **Multivariate and Correlations** window that appears, cast the two continuous data columns that you wish to test for correlation into the **Y, Columns** box (figure 34). Click on **OK**.
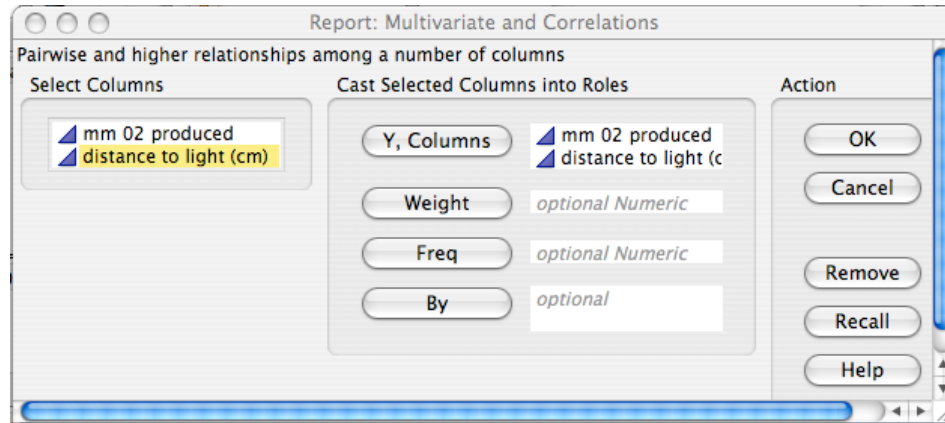
Figure 34. To test for correlation of two continuous variables, choose both variable columns as **Y**.

4. A new window will appear (figure 35) showing a matrix of correlation coefficients and showing your data as two scatterplots (with both variables displayed as the abscissa and again as the ordinate). More easily manipulated plots of your data may be created using other software. These plots are useful mainly in testing the assumption of bivariate normal distribution of the data. ***Refer to the statistics appendix, appendix I, for a more complete explanation of this assumption.*** The red ellipses imposed on the scatterplots will enclose 95% of the points if the distribution of the points is bivariate normal.

If the data are bivariate normally distributed, continue with the correlation analysis as described below. If the data are not bivariate normally distributed, then you should perform a nonparametric correlation as described in the following section of this appendix.
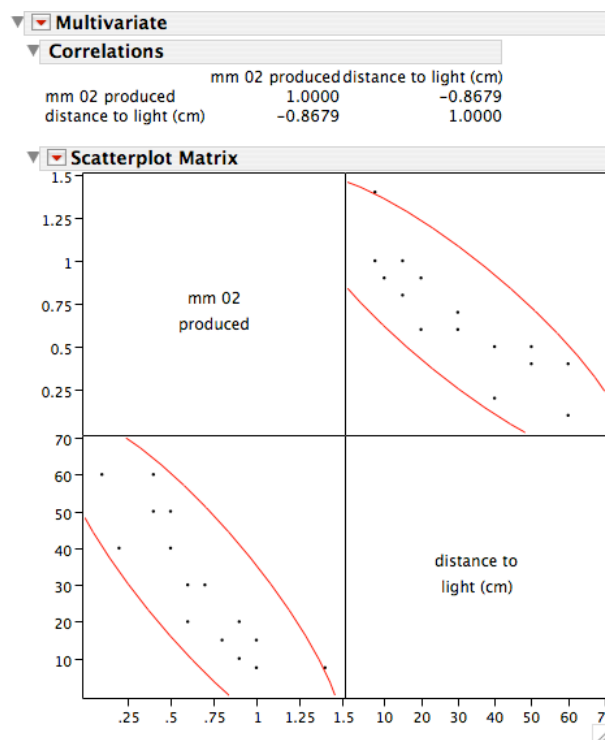
Figure 35. The default matrix of correlation coefficients and the scatterplots with 95% bivariate normal ellipses.

5. To perform the analysis, pull down the menu from the **Multivariate** bar by clicking on the small red triangle. Choose **Pairwise Correlations** to add the correlation analysis to the results window (figure 36). The correlation coefficient (r) is displayed under **Correlation**. The degrees of freedom for this test are calculated as the count minus two ($v = $ n-2). The probability for the example below would be recorded as P< 0.0001, v=13, r=-0.87. ***Refer to the statistics appendix, appendix I, for details about interpreting this information.***
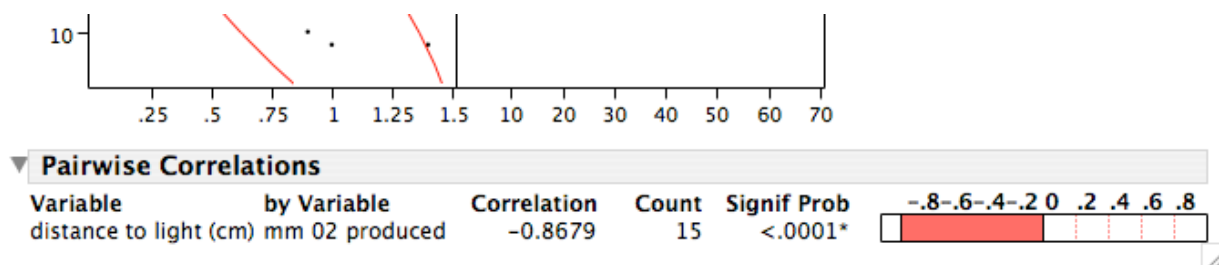

Figure 36. The results of a correlation analysis.

Non-parametric correlation -- Spearman's Rho test

1. If the scatterplot of the data does not show a bivariate normal distribution, then Spearman's rho test is a more appropriate analysis than choosing pairwise correlation from the correlation menu. Spearman's rho is calculated from the ranks of the data instead of from the values.

2. To perform the analysis, pull down the menu from the **Multivariate** bar by clicking on the small red triangle. Choose **Nonparametric Correlations**. From the submenu, choose **Spearman's ρ (Rho)** to add the correlation analysis to the results window (figure 37). Record the value of **ρ (rho)** and the probability (**Prob > |ρ|**).

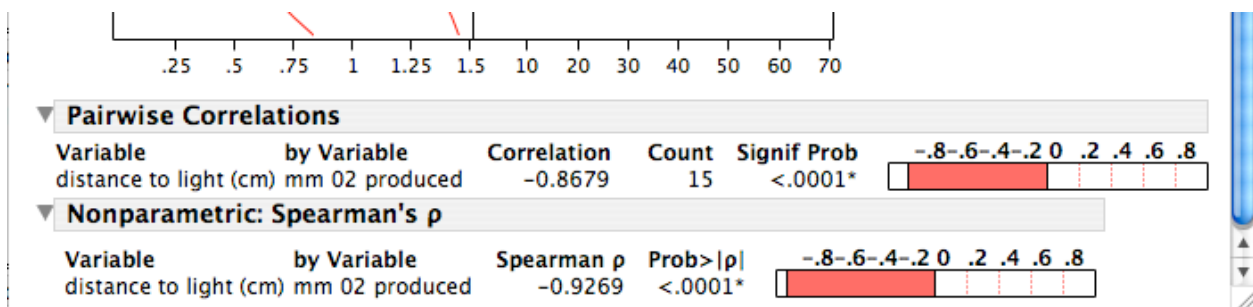Figure 37. The results of a nonparametric correlation analysis.

COMPARING MORE THAN TWO SAMPLES

Analysis of Variance (ANOVA)

ANOVA can be used to test a hypothesis that three or more samples represent populations that are different from one another. *Refer to the Biology 2 statistics appendix for a description of this statistical test.*

Data setup for ANOVA

1. Data should be entered in the same way that they would be arranged for an unpaired two-sample comparison:
   - All of the measurements will be in one column.
   - A grouping variable that clearly indicates the categories or treatments to be compared should be in the other (figure 38).

Figure 38. To compare three groups of measurements, for example, the running speeds of three size
            classes of animals, set up data in one column of continuous data and one grouping variable.

The column containing your measurements should be defined as **Continuous** data. The
grouping variable column must be a nominal or ordinal variable.

2. ANOVA assumes that samples are drawn from normally distributed populations. Check
the frequency distributions of the groups of measurements to be compared before
proceeding with the analysis (*Refer to p. II-9 for instructions*).

From the Basic tab of the JMP Starter

1. From the JMP Starter Window click on the **Basic** tab (figure 39).
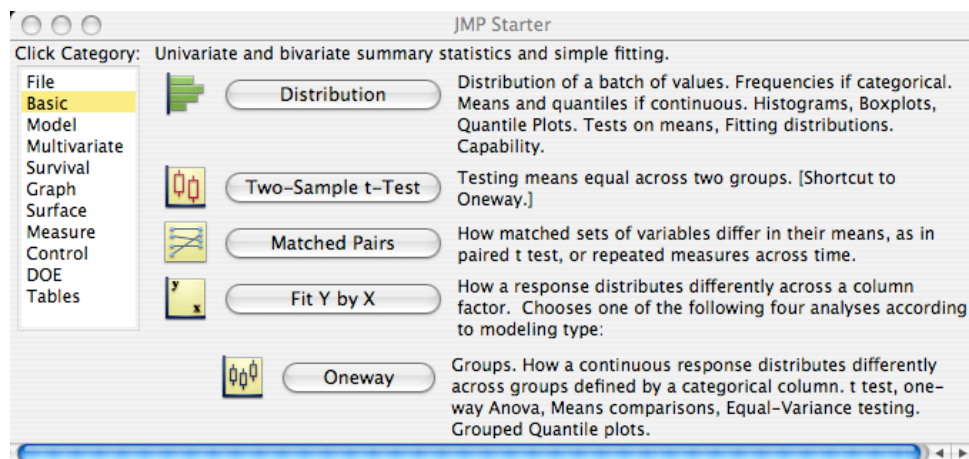
Figure 39. The Oneway command of the Basic tab begins an Analysis of Variance.

2. Click on the **Oneway** button. The **Oneway – Distribution by Group** window will open (figure 40).

    A. Cast the column of measurements into the **Y, Response** box.

    B. Cast the nominal grouping variable into the **X, Grouping** box.

    C. Click on **OK**.

Figure 40. Select columns for the Analysis of Variance.

3. The window that appears shows your data as a scatterplot (figure 41). Click on the red triangle to the left of the title bar ("Oneway Analysis of Running Speed (cm/sec) By Size Class") to see the analysis menu.
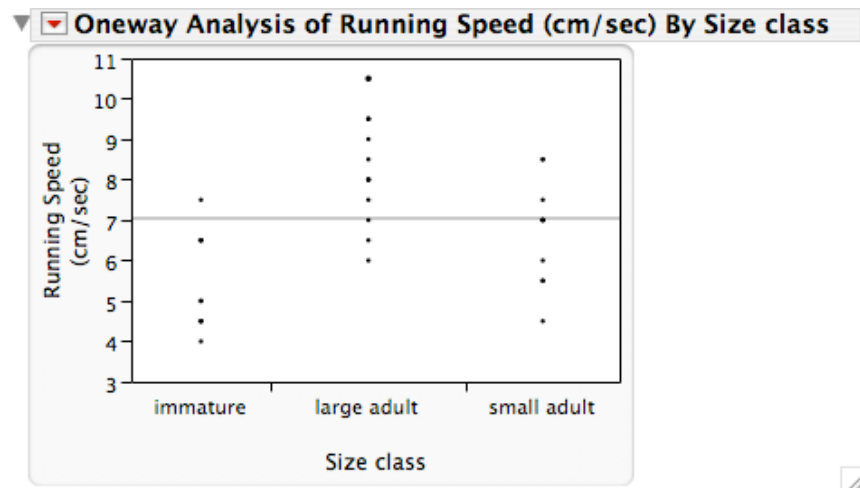
Figure 41. The default oneway analysis: a scatterplot of the data.
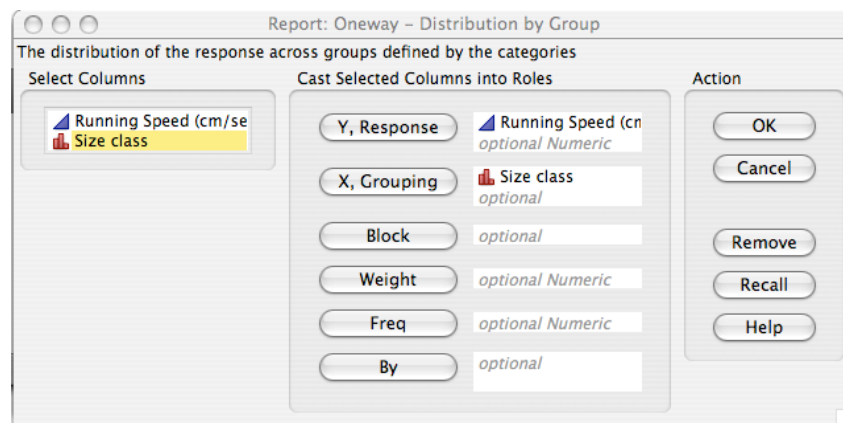
A. From the pull-down menu, choose **Means/ANOVA**.

B. The results of the ANOVA will be shown along with other statistics (figure 42).
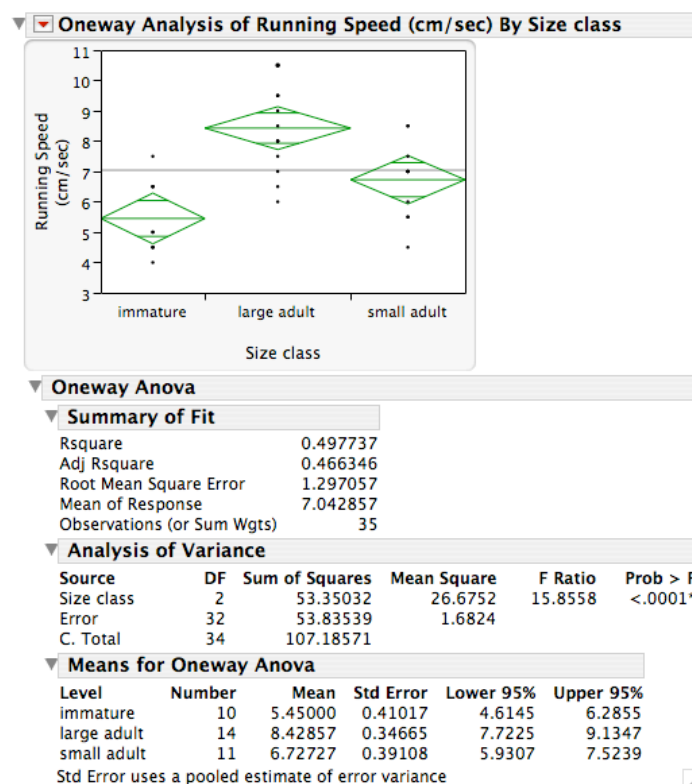
Figure 42. The results window of the Analysis of Variance.

C. *Refer to the Biology 2 statistics appendix, appendix I, or contact your laboratory instructor in order to interpret the ANOVA table.* In addition to the ANOVA table, JMP also supplies a means table giving the means and confidence intervals around

those means, for each of your groups. These data summaries can be very helpful in interpreting the results of the ANOVA analysis, but you should consider carefully what information is most appropriate or informative for the presentation of your data and what figure or figures will best support that information.

4. **If the ANOVA indicates that there *are* significant differences** between the means of different groups of measurements, you will want to know *which* groups are different. JMP also supplies post-hoc comparisons between pairs of categories.

   A. Click on the red triangle to the left of the title bar ("Oneway Analysis of Running Speed (cm/sec) By Size Class") to see the analysis menu.

   B. Choose **Compare Means**.

   C. From the pull down menu, choose from the options for post-hoc comparisons between groups. **Each Pair, Student's t** will supply *t*-tests between pairs of categories (figure 43). Each pair wise comparison involves the same probability of error (shown in figure 43 as α = 0.05). If you make many pair wise comparisons, then the probability of error across all tests is higher than that for individual tests.

▼ ▢ Oneway Analysis of Running Speed (cm/sec) By Size class

▶ Oneway Anova

▼ Means Comparisons

▼ ▢ Comparisons for each pair using Student's t

t	Alpha
2.03693	0.05

Abs(Dif)-LSD

	large adult	small adult	immature
large adult	-0.9986	0.6368	1.8847
small adult	0.6368	-1.1266	0.1229
immature	1.8847	0.1229	-1.1815

Positive values show pairs of means that are significantly different.

Level		Mean
large adult	A	8.4285714
small adult	B	6.7272727
immature	C	5.4500000

Levels not connected by same letter are significantly different.

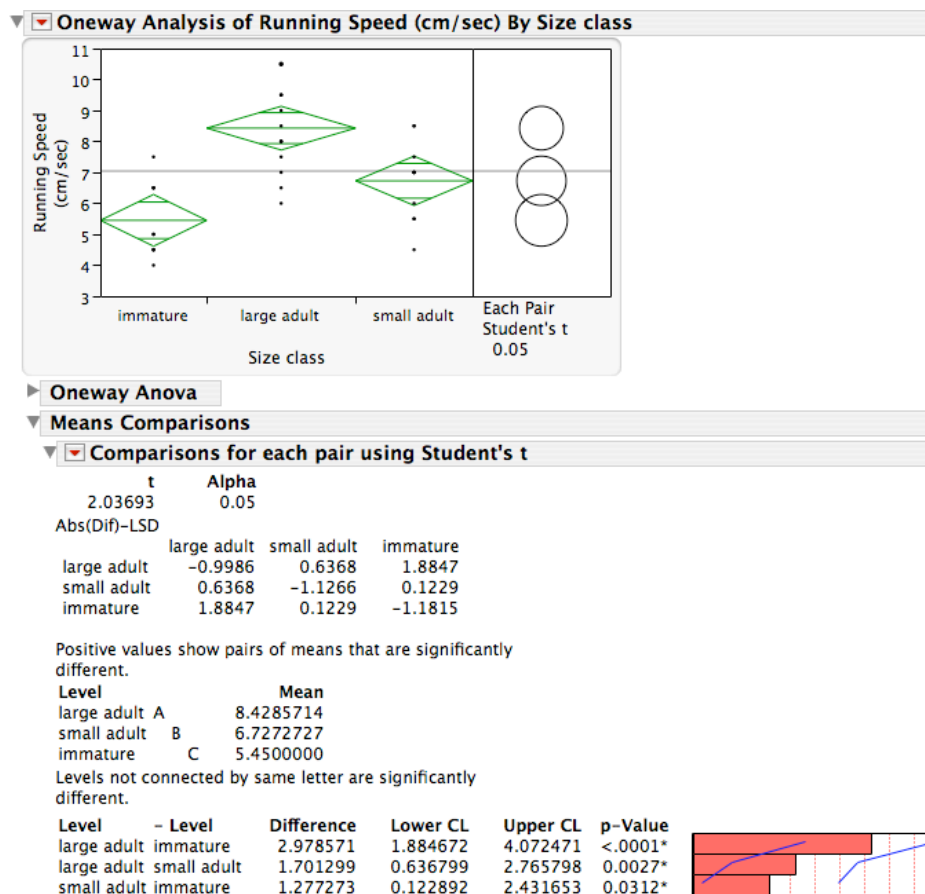| Level	– Level	Difference	Lower CL	Upper CL	p-Value
large adult	immature	2.978571	1.884672	4.072471	<.0001*
large adult	small adult	1.701299	0.636799	2.765798	0.0027*
small adult	immature	1.277273	0.122892	2.431653	0.0312*

Figure 43. If the ANOVA indicates significant differences between means, *t*-tests between pairs of means will indicate which groups of measurements are different

Non-parametric comparison – Kruskal-Wallis test

If, after examining the frequency distributions of each of the groups of measurements you wish to compare, you find that ANOVA would not be an appropriate statistical test, you may run a nonparametric test called the **Kruskal-Wallis** Test.

1. Data should be set up exactly as they would be for the ANOVA.

2. From the JMP Starter Window click on the **Basic** tab.

3. Click on the **Oneway** button. The **Oneway – Distribution by Group** window will open.

   A. Cast the column of measurements into the **Y, Response** box.

   B. Cast the nominal grouping variable into the **X, Grouping** box.

   C. Click on **OK**.

4. The window that appears shows your data as a scatterplot. Click on the red triangle to the left of the title bar ("Oneway Analysis of Running Speed (cm/sec) By Size Class") to see the analysis menu.

   A. Choose **Nonparametric**.

   B. From the pull-down menu, choose **Wilcoxon**. The Kruskal-Wallis test is automatically supplied if there are more than two groups in your dataset.

   D. The results of the Kruskal-Wallis rank sums test will be appear (figure 44). The test statistics that you should report are the Chi square value, the degrees of freedom, and the P-value (**Prob > ChiSq**).
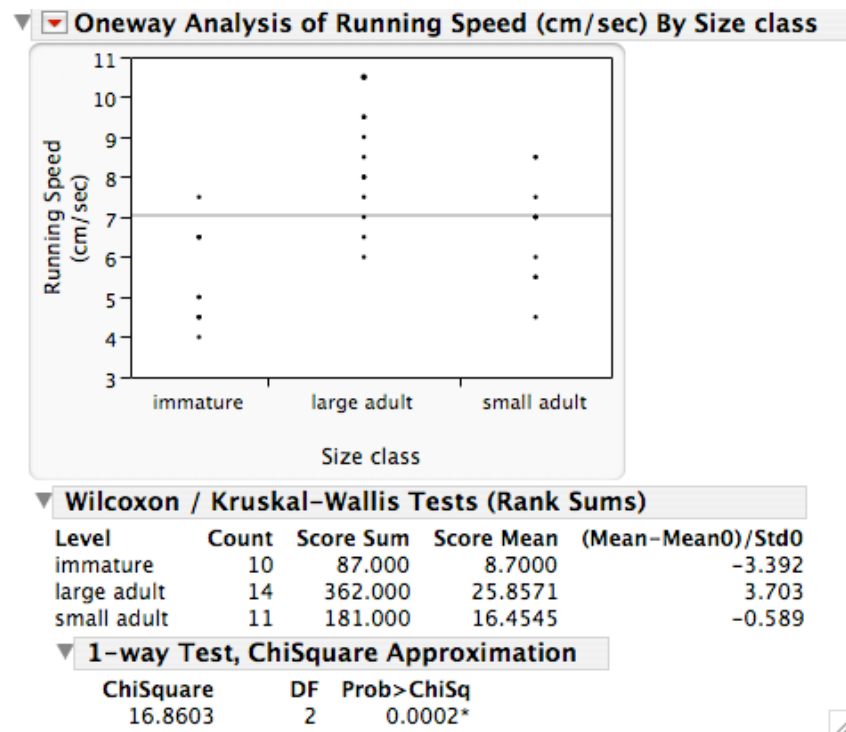
Figure 44. The results of a Kruskal-Wallis rank sums test (nonparametric comparison between three groups).