

**Experimental  
Design &  
Data Analysis  
Handbook**

**2.0**

## **OBJECTIVES**

**The objectives of this appendix are to outline some how-to's when it comes to statistics and experimental design. The table of contents will guide you through the appendix. It is written in logical order of how to proceed with an experiment.**

## **HOW TO USE THIS HANDBOOK**

**Students are NOT expected to memorize all the statistical tests and their applications. However, terms described in pages 2-15 should be well understood in order to be able to use this appendix effectively.**

**Students who are already familiar with the terminology can proceed directly to DECISION TREE 1 to determine the appropriate test to use for their research question. Please note, that several important and useful analyses have been left out of this handbook, namely correlation and regression analysis. If your data do not fit into the categories given in DECISION TREE 1, consult with a faculty advisor. It may be that you need to use a different analysis.**

## TABLE OF CONTENTS

How to Set Up an Experiment or Prepare for a Laboratory Experiment.....	2
What to Do Before Lab.....	2
How to Collect Data .....	2
Working to Understand Your Data.....	3
How to Use Inferential Statistics .....	3
The Null Hypothesis.....	4
Making Decisions.....	4
About the 0.05 level of significance.....	5
Two-tailed versus one-tailed tests: Choosing the right null hypothesis.....	5
How to Choose the Right Test.....	7
Types of data .....	7
Paired vs. non-paired data .....	7
A word about data transformations .....	8
Common data transformations.....	8
DECISION TREE 1 → START HERE! .....	9
DECISION TREE 2.....	9
DECISION TREE 3.....	10
DECISION TREE 4.....	11
How to Use Statistix For Windows .....	12
How to Run Descriptive Statistics and Graphing in Statistix.....	14
How to Get to Statistical Tests in Statistix.....	14
How to Run Descriptive Statistics in Statistix.....	15
How to Make a Histograms in Statistix .....	15
What is a Histogram? .....	16
How to Make a Box and Whisker Plot in Statistix.....	18
Testing Data for Normality in Statistix.....	20
Testing for Equal Variances in Statistix .....	21
How to Interpret Results of a One-Sample t-Test.....	22
How to Interpret Results of a Two-Sample t-Test.....	23
Two tailed example .....	23
One-tailed example.....	24
How to Interpret the Results of a Rank Sum (Mann-Whitney U) Test .....	25
How to Interpret the Results of the Kruskal-Wallis One-Way AOV .....	26
How to Interpret the Results of a Paired t-Test .....	26
How to Interpret the Results of a Wilcoxon Signed-Rank Test .....	27
How to Interpret the Results of a Sign Test.....	27
How to Interpret the Results of a One-Way Analysis of Variance.....	28
How to Calculate the Chi Square Statistic.....	30
How to Calculate the G Statistic.....	31
What Information You Need From a Statistical Analysis .....	31
How to Present Results of a Statistical Analysis in the Text of a Report.....	32
How to Present Results of a Statistical Analysis by using Graphs.....	33
Data Used in Examples of the Handbook.....	35
References.....	35

## How to Set Up an Experiment or Prepare for a Laboratory Experiment

1. Select a topic of interest (if you have that option).
2. Do background reading on the topic (you may have to do more serious research depending on the expectations of your assignment). At a minimum, read the appropriate sections of the text and your lab manual.
3. Decide on a methodology (if already assigned, read through it carefully) and consider what your experimental and null hypotheses are going to be.
4. Decide on the statistics you will use to analyze the data gathered using that methodology.
5. Set up your experiment.
6. Collect your data. Do this accurately and precisely (see section below on data collection).
7. Use descriptive analyses to understand your data. Do the tests you have chosen to use allow for you to use them (have you met all the assumptions of the tests)?
8. Use the appropriate statistics on your data.
9. Draw conclusions.
10. Write a report or answer questions relating to the lab material.

Notice that only parts 5 and 6 may be done during the laboratory period. The other steps are crucial to your understanding and accurate completion of a lab or experiment. This appendix will help you with some how-to's for laboratory exercises in Biology 101 and 102. Please consult an appropriate writing manual to complete step 10. You should also consult an appropriate statistics text for details on terms, calculations, and to find values from statistical tables.

## What to Do Before Lab

The most important thing you can do for an experiment or lab experience is to determine and understand the experimental and null hypotheses. Your experimental hypotheses will often set up what you expect to find in a particular experiment. Your null hypothesis is what is used in a statistical test to test your data against. It can vary depending upon the test and type of experiment that has been set up. Therefore, it is crucial to understand the statistics before you seek to design an experiment.

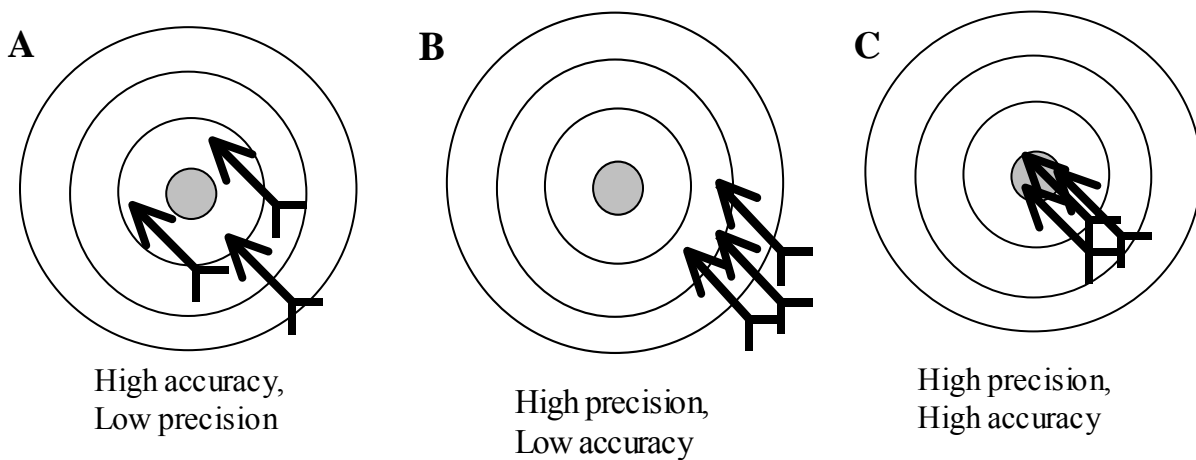
Before walking into lab, you should be able to answer the following questions:

- (1) What are the results expected from this exercise and why?
- (2) What types of data will be collected (continuous, discrete, or categorical)?
- (3) What types of descriptive statistics will be valuable to calculate?
- (4) What types of statistical tests might be run on the data?

## How to Collect Data

When variables are measured such as length, weight, and time, these are considered imprecise measurements because the **accuracy** (degree of closeness of the measurement to the actual value) of the measurement is dependent on the device used and how well the measurer uses it. The degree of **precision** (repeatability, or closeness of measurements to each other) must also be considered when determining the means of taking measurements. Take the example of trying to hit the bull's eye in darts. If you have a high accuracy and low precision, you may hit near the bull's eye, but not hit it (Figure 1A). On the other hand, if you have high precision and low accuracy, you may miss the

bull's eye, but all darts will be close to each other (Figure 1B). When taking measurements, we strive to be both accurate and precise (Figure 1C).



**Figure 1.** Examples of the difference between precision and accuracy. Note that low accuracy and low precision would result in darts all over the board, none of which hit the bull's eye.

### Working to Understand Your Data

Once you have your data, the single most important thing you can do is seek to understand it. The first step towards understanding it is to use descriptive statistics, histograms, graphs, etc. LOOK at your data using various means available to you: graph your data in a histogram, scatterplot, line graph or other appropriate way. This will help you understand what your observations are and how best to interpret them. Quite often, if it looks like there's no difference between groups, then there probably isn't any statistically significant difference.

An important concept to understand and keep in all your work is to ALWAYS include some measure of dispersion with your data. Means by themselves are not very useful. You must always try to understand the amount of variance in the data (measured using standard deviation, variance, or standard error). Many statistical tests use these measures to see how similar or different your groups of data are. (Consult a statistics text for a thorough description of descriptive statistics and see the section "How to Present Results of a Statistical Analysis by Graphs" for a comparison of standard deviation and standard error).

One of the first things you should do with your data is look at it. There are several tools you can use to visualize your data. In Statistix, if you go to "Summary Statistics" under the "Statistics" menu, you will find many tools you can use. The first and perhaps most basic step is to calculate descriptive statistics for your data.

### How to Use Inferential Statistics

Scientists ask questions about populations, but must make measurements on a subsamples from entire populations. We must therefore decide if the samples that we have measured differ enough for us to state that there is a difference between the populations. We then use **inferential** statistics because we use a sample to infer the properties of a larger population. If we have a complete census of the population, then these would not be needed. The types of questions we often

ask are things such as: What effect does temperature have on the running speed of lizards? Does species A have longer tails than species B? Are red oaks more variable than pin oaks?

There are two types of inferential statistics: parametric and non-parametric tests. **Parametric statistical analyses** are very powerful inferential tests. They are powerful, because they are able to detect small differences between subsamples. However, these tests require samples to be randomly drawn from the population and that the data follow a normal distribution. A **normal distribution** is symmetrical about its center (the mean and median) and bell shaped. Most parametric statistical tests will also assume that the samples being compared will have equal variances. This is why you may have to test your data for normality and homoscedasticity (equal variances).

**Non-parametric analyses** are much less powerful than parametric tests. These tests make fewer assumptions about the population from which the sample was taken but you must still understand their assumptions to use them properly. None assume data are normally distributed, although some assume that the distributions of more than one sample are from the same distribution. Some also assume variances are equal. You can first try to transform your data to see if you can meet the assumptions of normality and homoscedasticity (see “A Word About Data Transformations” in the section entitled “How to Choose the Right Test”). That way, you can still use the more powerful test. You may, however, be forced to use a non-parametric test from time to time.

### *The Null Hypothesis*

Whenever you do a statistical test, there is always a null and an alternative hypothesis being tested. You must state the null and experimental hypotheses before doing an experiment. This will require you to figure out what statistics you are most likely to use to analyze your data. The **null hypothesis ( $H_0$ )** will state that there is no significant difference between the studied groups of measurements. The **alternate or experimental hypothesis ( $H_A$ )**, would be that there is a difference. An experiment must be designed to be able to accurately differentiate between the null and experimental hypotheses. Here again, it is important to understand the underlying statistics to be able to design a good experiment that will be able to answer your research question.

### *Making Decisions*

After performing any inferential statistical test, we get a probability value, or p-value. In general, biologists decide that there is a significant difference between means when the p-value is less than 0.05. This is also described as a 0.05 level of significance (occasionally scientists prefer to use a different level of significance). A 0.05 level of significance means that 95% of the time we will get data that look like this when they are actually different or, put another way, we will reject the null hypothesis incorrectly 5% of the time (Type I error).

For example, we might perform an experiment to determine if lizards run more quickly when temperatures are high than when they are low. We would proceed to hypothesize: the null hypothesis would be that there is no significant difference between lizard speeds at high and low body temperatures, and the experimental hypothesis would be that there is a significant difference between lizard speeds at high and low body temperatures. We would then set up and perform our experiment, making sure that we have a large sample size (50 lizards at each temperature, for example).

After the experiment, we would perform an inferential statistical test, preferably parametric. If our p-value was calculated to be less than 0.05, we would reject our null hypothesis and state that there is a significant difference in lizard speed at high versus low temperatures.

#### *About the 0.05 level of significance*

When doing statistical analyses, you must remember that setting the level of significance of the experiment is relatively subjective. Most scientists accept a 0.05 probability of incorrectly rejecting the null hypothesis. However, this may not be acceptable depending upon the situation. For example, what if you are testing medicines to be used on humans. If application of the drug could be a chance of death, then the testers may require a 0.001 level of significance before they will accept that a drug is effective and safe. On the other hand, some data can be extremely variable, such as data from large ecological experiments. There may be a lot of variation in the data due to natural variation among sites and individuals. In that case, it may be more appropriate to use a 0.1 level of significance for an experiment.

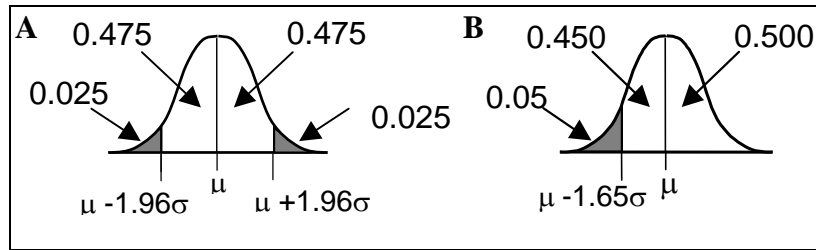
There is nothing cast in stone about this number. If your values approach the 0.05 level of significance it is probably a good idea to look carefully at the data to see if you think the results are **biologically** significant. (Although perhaps not strictly statistically significant at the 0.05 level).

#### *Two-tailed versus one-tailed tests: Choosing the right null hypothesis*

The terms one-tailed and two-tailed refer to the direction in which you expect the differences between means of groups to lie. In the above example, we were asking the question: Is there a difference in running speed of lizards in high and low temperatures? The question does not assume that they will run faster in low temperatures as compared to high temperatures or vice versa. However, sometimes we know something about the organisms we are studying and can make a guess as to which way the differences will be. Since lizards are ectothermic, it might be logical to assume that they will run slower at a low temperature than at a high temperature (unless it's too high). So we may reword our question to be: Do lizards run faster when they are at higher temperatures? This will lead to a different null hypothesis.

In the first case, our null hypothesis was that there would be no difference between running speed of lizards at high and low temperatures (two-tailed). In the one-tailed example, the new null hypothesis states that lizards at high temperature will run faster than those at low temperature. You must be very careful and be sure you have plenty of evidence to back up your decision to use a one-tailed *vs.* a two-tailed test. This is because if lizards at slower temperature run faster than lizards at high temperature, you will accept the null hypothesis in the one-tailed example just outlined, and reject it in the two-tailed example.

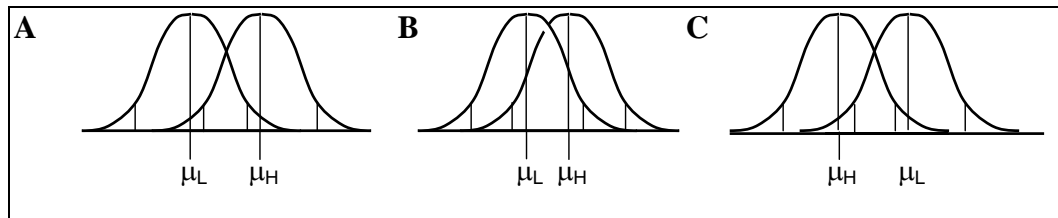
Why should we bother with one- and two-tailed tests? The reason is that we want to minimize the number of times we incorrectly reject the null hypothesis (Type I error). Look at the rejection areas under the normal curve for a two-tailed test (Figure 2). In the two-tailed example (A), the rejection areas (shaded) are distributed on both sides of the curve (0.025 on each side). In the one-tailed example (B), all of the rejection area is on one side of the curve only. This means that in a one-tailed test, there will be less chance to incorrectly reject the null hypothesis!



**Figure 2.** Rejection areas on a normal curve comparing a two-tailed test and one-tailed test at 0.05.

Suppose that  $\mu_L$  is the mean for lizards at low temperature, and  $\mu_H$  is the mean for lizards at high temperature and that their distributions are shown by the curves in Figure 3. Look at the first case (A). In this situation the means are significantly different because they are in the rejection curves. They would also be statistically significantly different if they lay completely outside of the distribution curves shown. On the other hand, look at case B. In this case, we will accept the null hypothesis because this shows that the means are not different enough to be considered statistically significantly different from each other. Thus, if we use a one-tailed test, the rejection area is larger and will allow us to be more accurate in assessing whether the populations are truly different from each other or not.

Now notice Case B. In that situation, the mean of lizards at low temperature is higher than the mean at high temperature. Since that is the case, in the one-tailed example when we were expecting the mean of lizards at high temperature to be greater than the mean at lower temperature, we will **accept** the null hypothesis, because the mean at high temperature was not greater than the mean speed at low temperature.



**Figure 3.** An example of data in which the null hypothesis will be rejected (A) because the means of each group are in the rejection area of the other curve as well as an example in which the null hypothesis will be accepted (B) because the means are too similar to be considered statistically significant.

In this experiment, there are three possibilities: lizard speeds will not be different from each other, lizards will run faster at high temperatures, lizards will run faster at low temperatures. Note that in the case of the one-tailed test, you will reject your null hypothesis **ONLY** if lizards run faster at high temperatures. It will not be rejected if they have the same running speed **OR** they run faster at low temperature (Figure 3C)! This is why you must be careful when choosing a one-tailed test to be sure you have good reason to do so. If you can, then you will be able to perform a much more accurate assessment.

## How to Choose the Right Test

In order to choose the right statistical test to run, you must understand what type of data you are dealing with and the question being asked. If you are dealing with categorical data, go to Decision Tree 1. If you are dealing with continuous or discrete data, go to Decision Tree 2 (you will have to know the number of treatment groups you have and you may have to know if data are paired or not paired). If you are not sure of what type of data you have, read the sections below.

### *Types of data*

There are three types of data: continuous, discrete, and categorical. **Continuous** data are those that are expressed in real numbers, that is, numbers that can be expressed as a decimal fraction. Data such as height, weight, etc. are continuous data. **Discrete** data are those that are expressed in integers, such as age, years in school, and number of toes. These data only make sense as integers. However, in general, we deal with them in much the same way as continuous variables. Even though they are integers, you can still describe the mean of the population with a real number. Hence we get the silly situation of saying things like “the average American household has 2.4 children”. Nonetheless, this is descriptive of a discrete data set (just remember that no one has 0.4 children...).

The other type of data is called **categorical**. This is when you have counts of individuals in certain categories. For example, you may have a population of 100 flowers and count the number with blue vs. white flowers. Your data set would be something like: 25 white and 75 blue. It makes no sense to calculate a mean of these numbers since they are already descriptive of the population you sampled. Geneticists looking at Mendelian traits often collect these types of data for a population or subsample of a population.

### *Paired vs. non-paired data*

These terms refer to whether or not your samples are dependent on each other. Let’s take the case of the lizard example we have been using. Suppose we are asking the question: What effect does temperature have on the running speed of lizards? We could measure the running speed of 50 lizards at 20°C and then we take the same lizards and measure their running speed at 40°C. Some of the variation in our data will be due to the individual lizards we sample. A slow lizard at 20°C is likely to be slower at 40°C as well. Since we are not interested in the variation of speeds among lizards, but only the overall effect of temperature on running speed, we can collect paired data (by keeping track of individual lizards) and use a special statistical test that can deal with these types of data.

We could answer the same question by collecting 50 lizards at random, measuring their running speed at 20°C and then collecting a second sample of 50 lizards (with none from the first sample) and measuring their running speed at 40°C. In this case, the data are non-paired. Most of the time, we use non-paired data, particularly for complex experiments (several temperatures and/or species of lizards), but often this is dictated by the amount of money and effort available to use for the research project. In the first case, we only had to catch 50 lizards, but in the other case, we would need to catch 100 lizards, so it would probably take twice as long to carry out the experiment!

An important point to remember is that the research question drives what types of data are collected and the analyses we run on those data. For example, if we are trying to breed fast running

lizards, then we would be very interested in the individual variation in running speeds and a paired test (which removes individual variation as “nuisance” variability) would not be appropriate. If the research question changes, so must the experiment and statistics used to analyze the experiment.

#### *A word about data transformations*

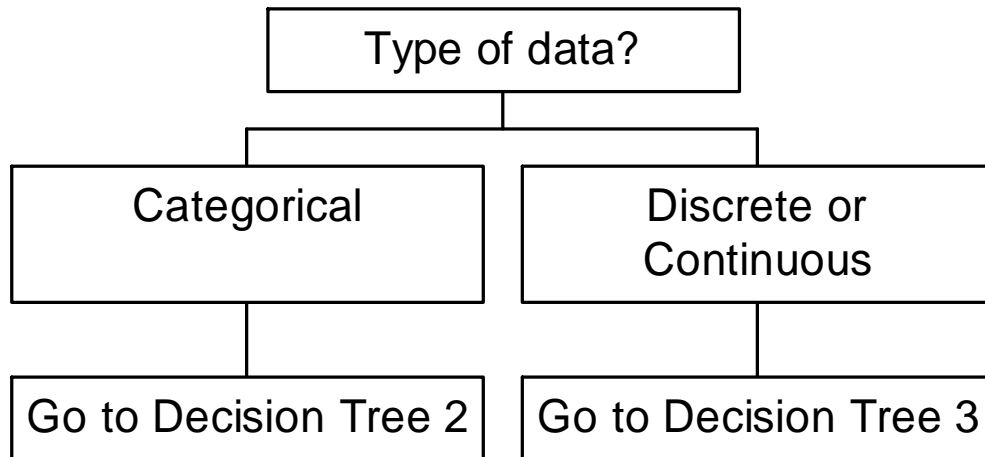
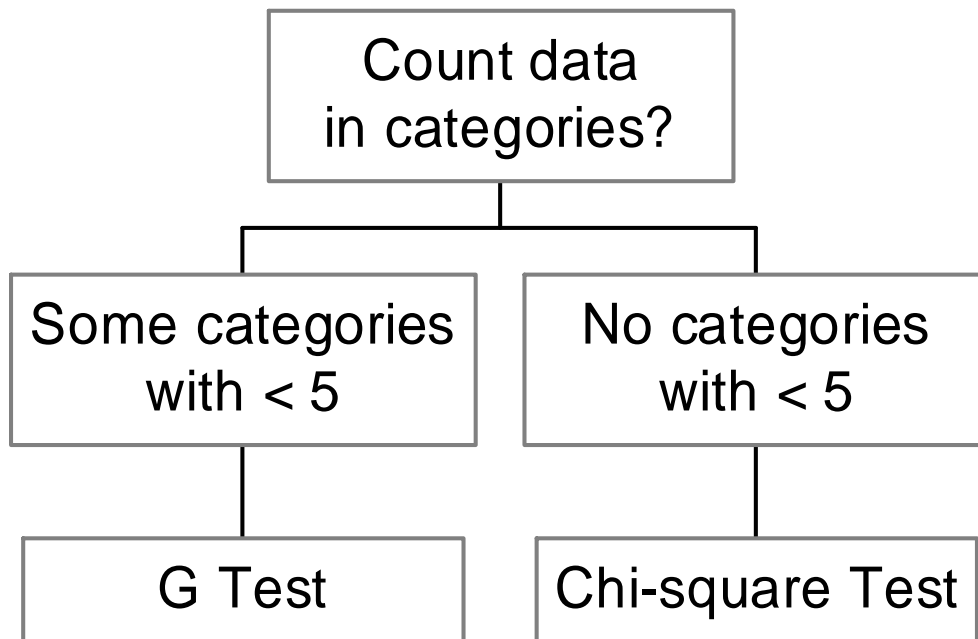
Some of you may think that transforming data is sort of “cheating”. Well, it really isn’t. All we are doing with a data transformation is putting the data on a different scale. So just as it doesn’t matter whether you measure heights in meters, centimeters, or feet, it doesn’t matter that we sometimes transform data to be able to run a parametric analysis on it. Until more powerful non-parametric analyses are available, this will be common practice. On the other hand, you should always remember what transformations you do. Obviously, you must transform all the data, not just a few data points. Also, think about what it may mean that your data are normal on a different scale than the one you used to take measurements.

#### *Common data transformations*

For height and weight data, log base ten or natural log are often used.

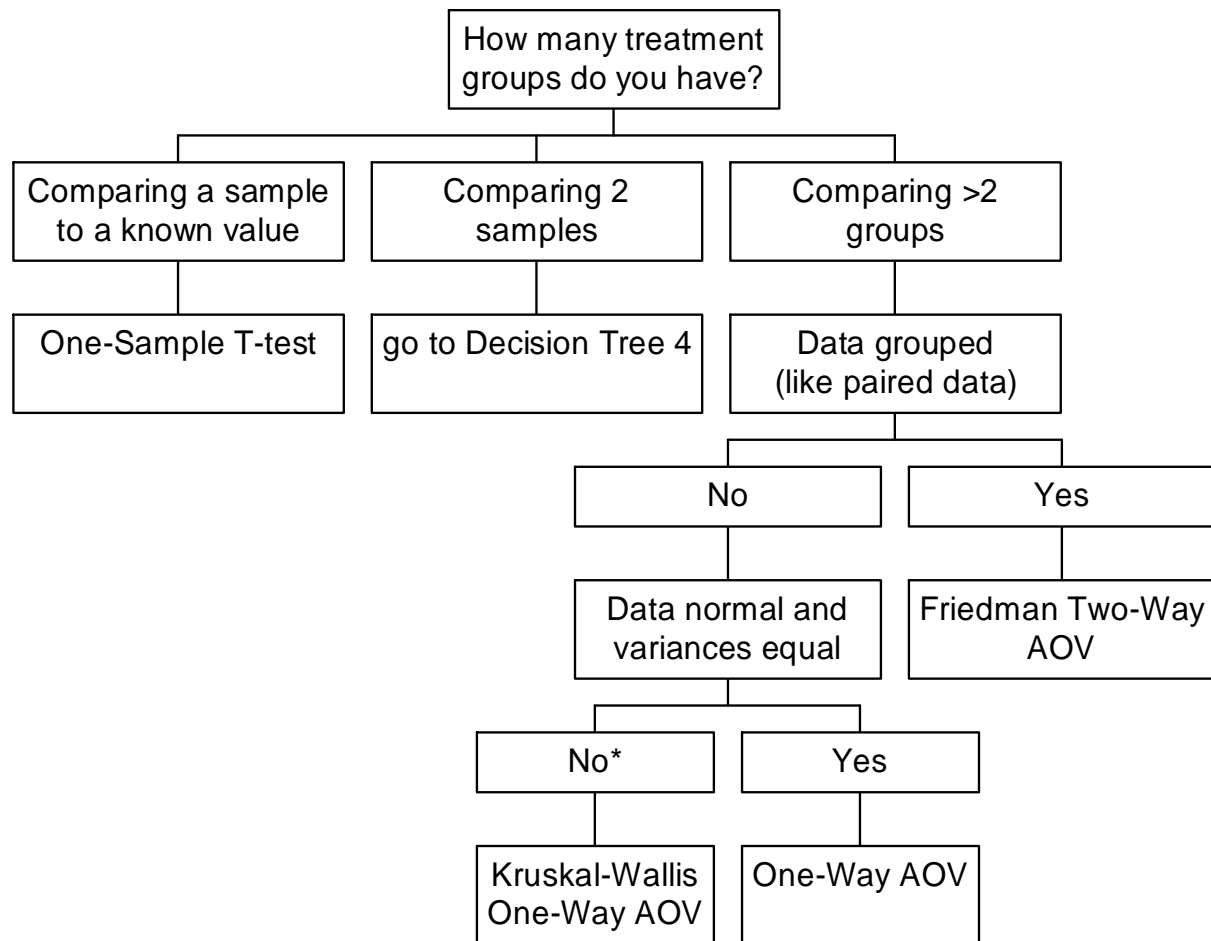
For count data, the square root of the data is often normal.

For ratios or proportions, the square root of the arcsin can be helpful.

*DECISION TREE 1 → START HERE!**DECISION TREE 2*

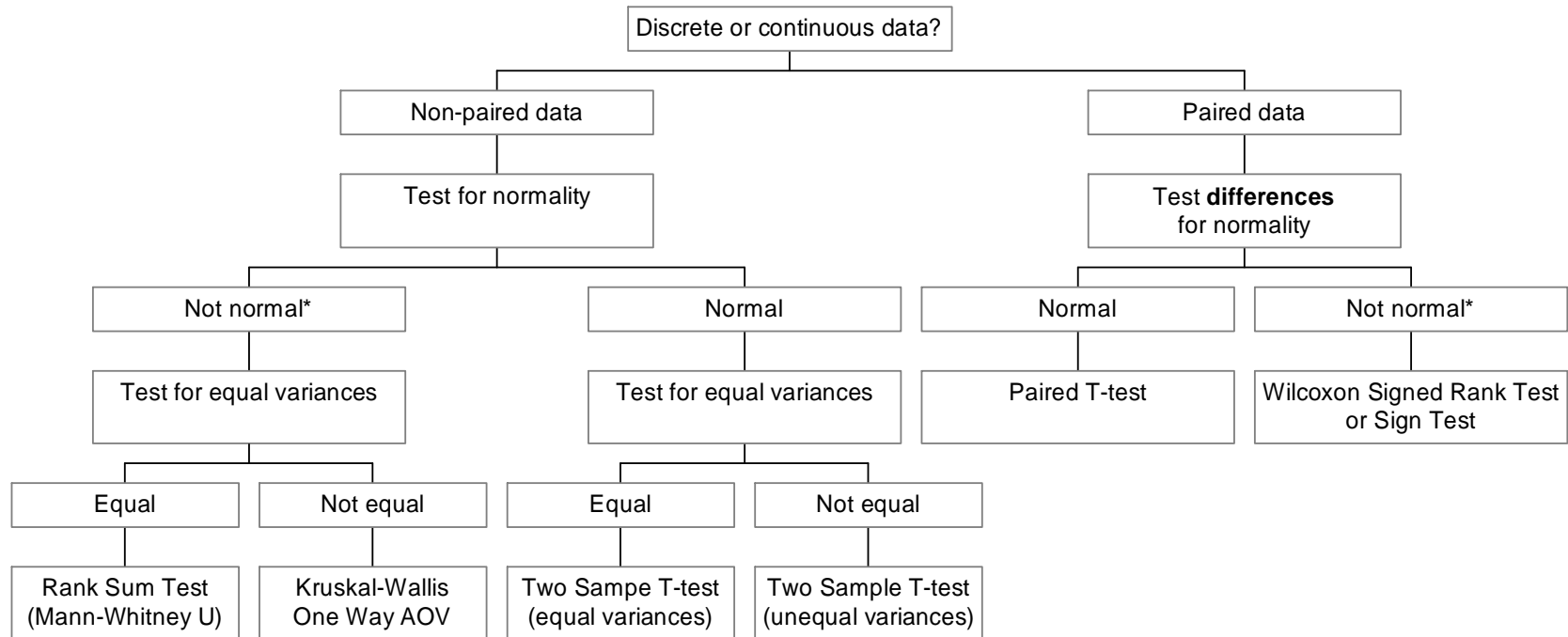
Directions on performing the Chi-square and G statistic test by hand (which are very easy) are given at the end of this appendix. Note that neither case can have a category equal to zero and they cannot be run on Statistix.

## DECISION TREE 3



\* If data are not normal, you can transform the entire dataset and retest to see if you can use the more powerful parametric test, see section on transformations for help with your data. Transforming data is usually done first to see if you can use the more powerful parametric test. If transformations do not help you data, then go ahead and use the non-parametric test available to you.

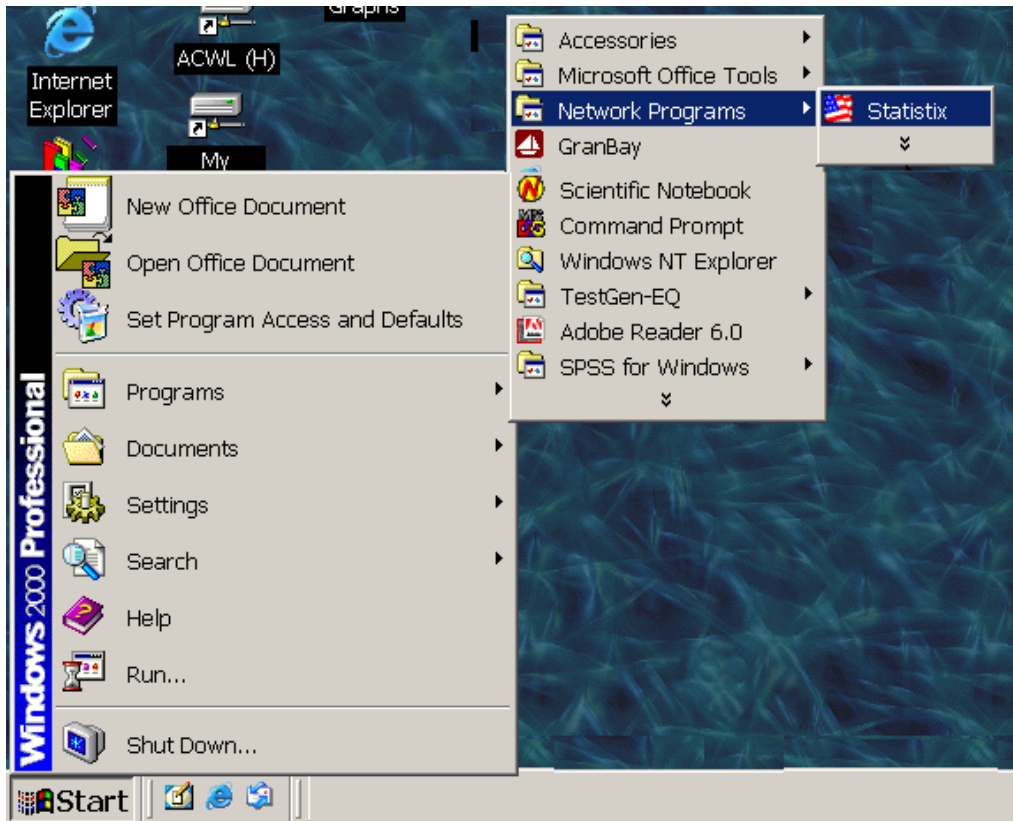
### DECISION TREE 4



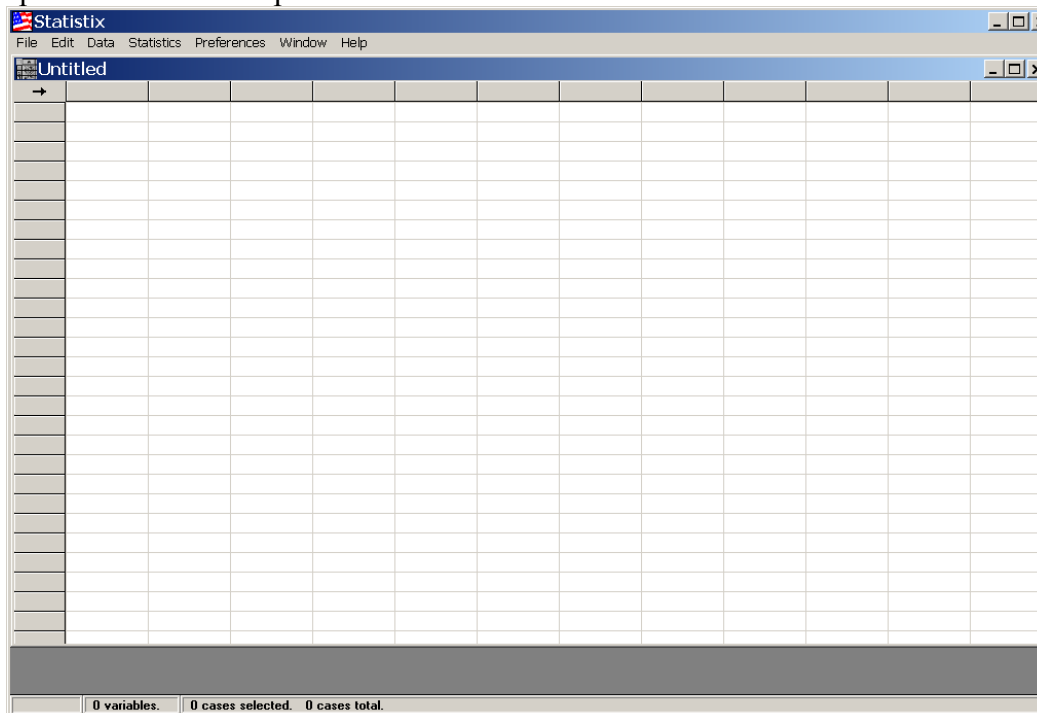
**\* If data are not normal, you can transform the entire dataset and retest to see if you can use the more powerful parametric test, see section on transformations for help with your data. Transforming data is usually done first to see if you can use the more powerful parametric test. If transformations do not help you data, then go ahead and use the non-parametric test available to you.**

To get to any of the above tests, choose the Statistics menu and then select “One, Two, Multi-Sample Tests”. Instructions for interpreting results from Statistix from all these tests are given in the next sections.

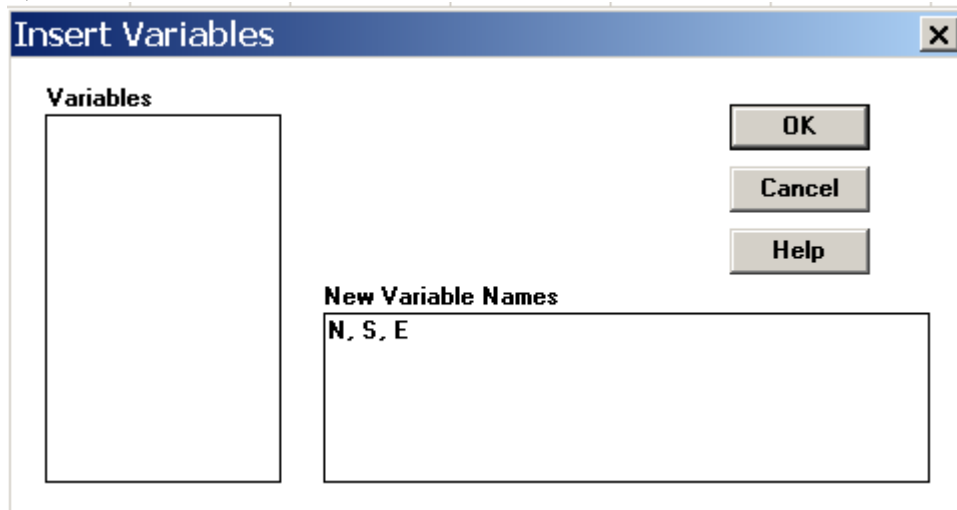
## How to Use Statistix For Windows



From the Start Menu, choose Programs, then Network Programs, and then Statistix. A blank spreadsheet should open like this:



You can insert variables by going to the menu “Data” and choosing “Insert” and Variables”. Type in variable names (no spaces) that are descriptive of your data. Then click on "OKAY". (You can enter several at a time and separate them with commas). In this example, I will enter three variables called N, S, and E.



Once you have named your variables, enter the data into each column under the appropriate variable. My data for N, S, and E look like this:

The image shows the Statistix software interface. The title bar reads "Statistix" with a small American flag icon. The menu bar includes "File", "Edit", "Data", "Statistics", "Preferences", "Window", and "Help". The main window title is "Untitled". Below the title bar is a data table with the following structure:

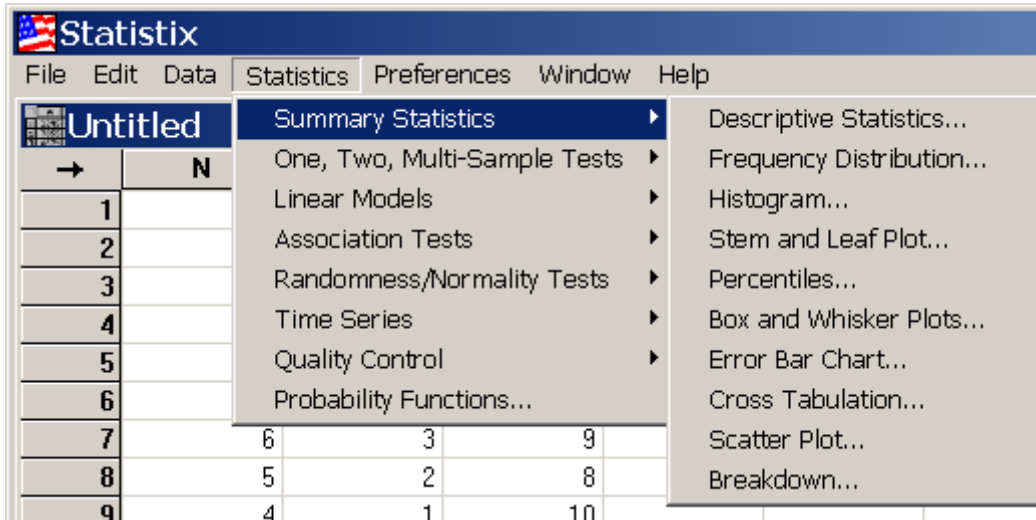
→	N	S	E		
1	4	2	5		
2	3	1	6		
3	7	4	7		
4	8	3	6		
5	9	3	7		
6	8	5	8		
7	6	3	9		
8	5	2	8		
9	4	1	10		
10	8	4	11		
11	9	6	13		
12	6	2	7		

The cell containing the number 8 in the 6th row and 4th column is highlighted with a blue border.

I am now ready to **look** at the data carefully using graphs and descriptive statistics and then begin my analysis.

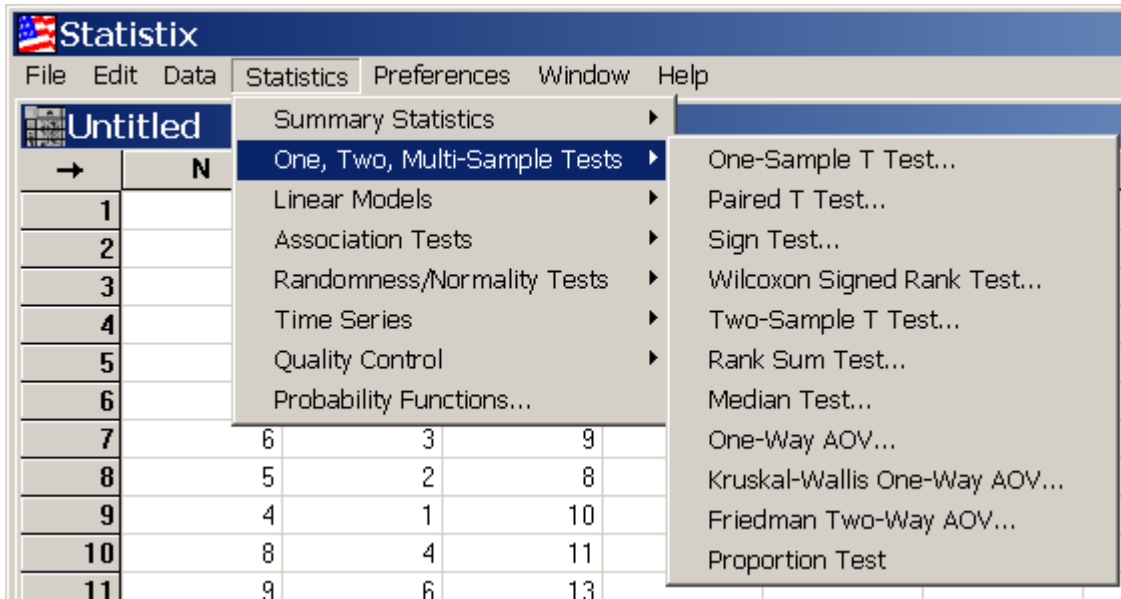
## How to Run Descriptive Statistics and Graphing in Statistix

Note that to run descriptive statistics and get graphs such as box and whisker plots, histograms, etc. you must go to the “Statistics” menu and select “Summary Statistics” as seen below:



## How to Get to Statistical Tests in Statistix.

Note that all the statistical tests referred to in Decision trees 3 and 4 can be found under the “Statistics” menu under “One, Two, Multi-Sample Tests” as seen below:



More sophisticated statistics are also available, but will not be covered in this handbook.

## How to Run Descriptive Statistics in Statistix

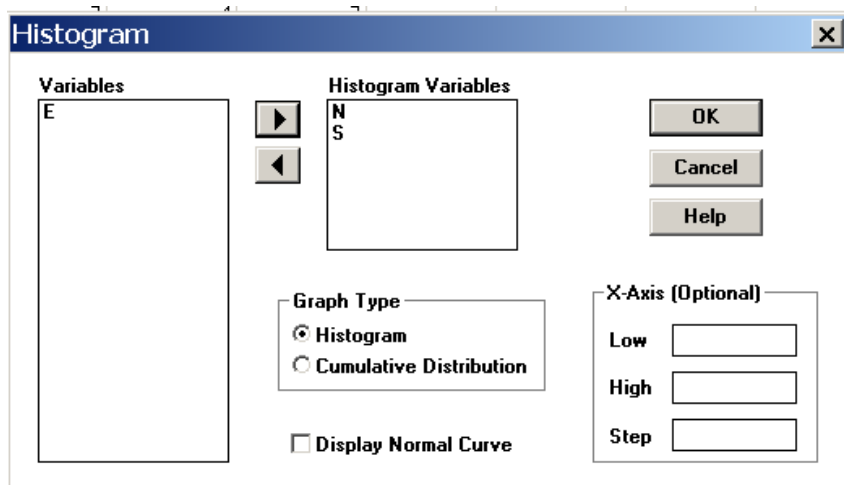
In order to run descriptive stats, enter your data and go to “Summary Statistics” menu and choose “Descriptive Statistics”. In the box that comes up, enter all the variables for which you want these statistics calculated. You will also have to choose which items you want to see. I suggest choosing at a minimum the options in the output box that follows. Note that N = sample size, SD = standard deviation, SE Mean = standard error of the mean, Min/Max = minimum and maximum values.

The following are results that we attain with the N and S data sets (data are at the end of this handbook):

DESCRIPTIVE STATISTICS			
	N	S	
N	12	12	
MEAN	6.4167	3.0000	
SD	2.0652	1.5374	
VARIANCE	4.2652	2.3636	
SE MEAN	0.5962	0.4438	
MINIMUM	3.0000	1.0000	
MAXIMUM	9.0000	6.0000	

*Annotation:* Note that there are an equal number of samples (12) in the “N” and “S” data set. The mean of the N group is larger as is the standard deviation. This is very often the case, so if you have two groups with very different means, variances will rarely be equal. The standard deviation is the square root of the variance and is in the units of the mean. So, if we measured growth in cm, the mean and SD (as well as SE Mean) is also in cm.

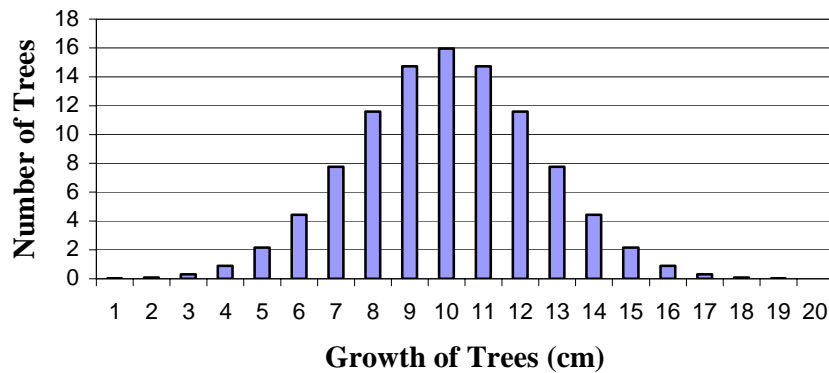
## How to Make a Histograms in Statistix



In order to make a histogram select “Histogram Variables” (they will all be in the same histogram) and then select “OK”. Note that you have some options to make the bins (“X-axis (optional)”) any size (step) you want as well as control the highest and lowest value.

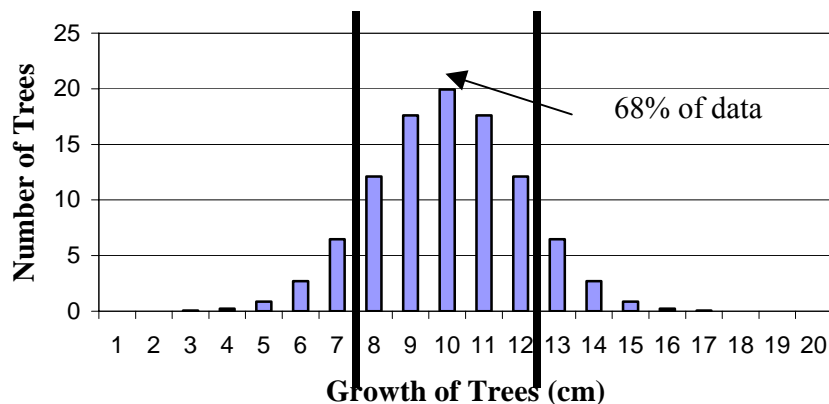
### What is a Histogram?

A histogram is a graph that shows all your data points placed in “bins” (on the x-axis) with the frequency of occurrence on the y-axis. In this case, we have a dataset of 100 trees. For each tree, we measured the yearly growth of that tree in centimeters. Each bin on the x-axis represents values that fall from the previous value to the one shown. For example, there were no trees that grew between 0 and 1 cm and there were almost 12 trees that grew more than 7 cm up to 8 cm. This example is showing you what a normal curve would look like. Notice that the mean is the highest point (10) and that it is perfectly symmetrical. It also has some other important properties related to the dispersion around the mean.



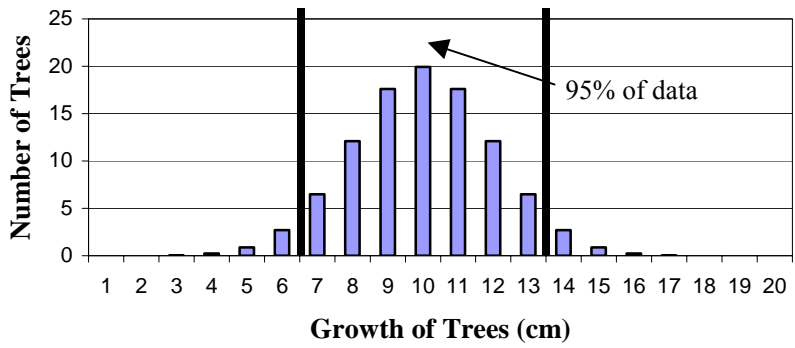
**Figure 4.** Histogram of the growth of 100 trees. Data show a normal distribution.

The meaning of standard deviation is relatively simple to understand on a normal curve. For these data, the standard deviation is 2. Since the data are normal, we know that within 1 standard deviation from the mean (that is  $+2$  and  $-2$  from the mean or from 8 to 12) 68% of our data points can be found there. You can add up the columns to check this out (Figure 5).



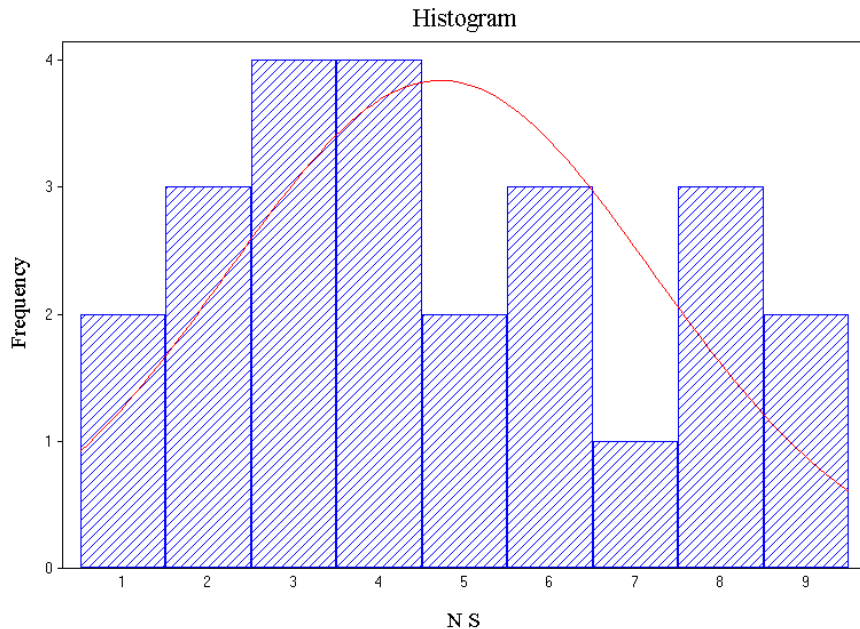
**Figure 5.** Histogram of 100 trees with one standard deviation from the mean highlighted. Note that this area contains approximately 68% of all data points.

Within 2 units of standard deviation, that is  $+4$  and  $-4$  from the mean, or from 6-14, we can find about 95% of all our data. Notice that only the very small tails are outside of this area (Figure 6).



**Figure 6.** Histogram of 100 trees with two standard deviations from the mean highlighted. Note that this area contains approximately 95% of all data points.

Most often, you will not get nice looking histograms like the ones shown above. This can happen for many reasons. The most likely reason you will encounter is that you do not have enough data points to really see the pattern of the distribution. Or, it could be that your data simply are not normal. Note that the data for S and N together do not closely follow the normal curve shown by the smooth line (Figure 7). This is for the above two reasons. The S and N data sets are different from each other, each having a different mean, so these data are bimodally distributed and will NOT be normal. However, a visual estimation of normality can be achieved using the Wilk-Shapiro/Rank It plot (see Testing Data for Normality in Statistics).



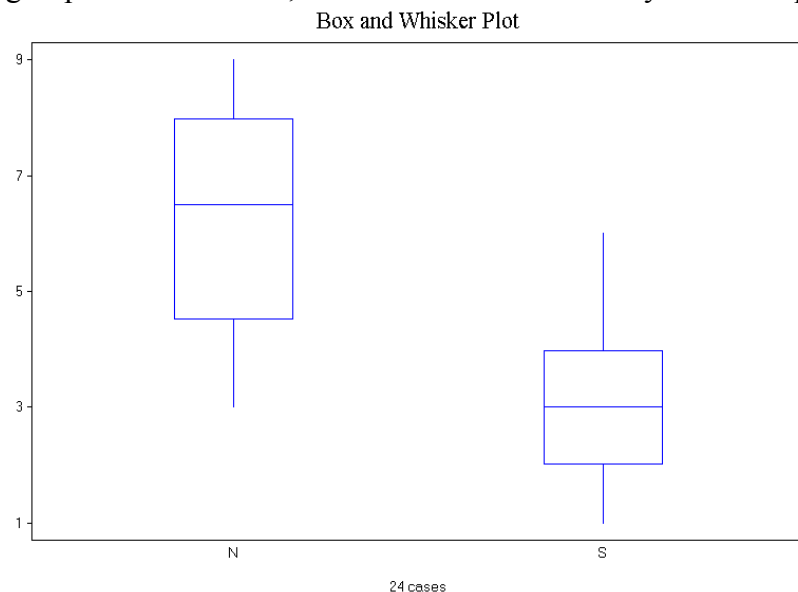
**Figure 7.** Statistix output for the N and S dataset shown together on one histogram. The line represents what a normal curve would look like for these data.

## How to Make a Box and Whisker Plot in Statistix

Another useful way to look at data is to use box and whisker plots. Unfortunately, these are not completely standardized, so you have to consult your specific statistical package help to see what exactly they display. To get a box and whisker plot in Statistix, go to the “Statistics” menu, choose “Summary Statistics” and then “Box and Whisker Plot”. If you choose the “Table” option, you can select the variables to plot under the “Table” box. In this case, I have chosen to see the N and S datasets (Figure 8).

The box itself contains the middle half of the data (that means half of the data points are found within the range of the box. The line through the box denotes the median. The median is not the same as the mean. It is the middle value of the data. That is, if you were to put all the data in numerical order from smallest to highest, this value would be the value precisely in the middle of these values. In a normal distribution, the mean and median are the same value. Note that some non-parametric statistical tests use the median instead of the mean since the median value is not affected as much by outliers (data points that fall far from the mean) as means.

The “whiskers” are the lines that extend up and down from the boxes. These encompass most if not all of the rest of the data. If there are data points that are greater than 1.5 times the size of the box, these are shown in the box and whisker plot as outliers. There are two types of outliers that Statistix will show. If values fall outside 1.5 times the width of the box, these are possible outliers denoted by an asterisk “\*”. Values that fall greater than 3 times the length of the box are considered probable outliers and are shown by an “o”. You should check your data for possible and probable outliers. These can greatly influence your means and therefore any statistical test that uses them. If you have outliers in one group and not another, the variances are also likely to be unequal.

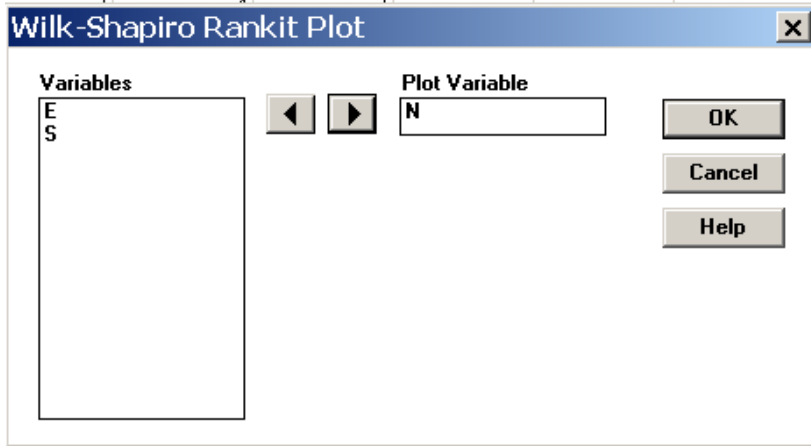


**Figure 8.** Statistix output for box and whisker plots using the N and S dataset. Note that although the two groups of data overlap in range, the medians are very different. Notice also that the upper whisker of the S group is longer than the lower whisker. This indicates that data are clumped toward the lesser values and that the distribution is NOT symmetrical (the normal distribution, remember, IS perfectly symmetrical) and there are no outliers.

There are many reasons for having outliers in data sets. The most common problem is a mistake in entering data (moved the decimal place a few points by mistake). This is a good point in the process to find these types of errors. It is possible, however, that your outliers represent real, good data points. In that case, you may want to consider keeping them in the data set and dealing with it by trying transformations (see sections “A Word about Data Transformations” and “Common Data Transformations”) or using a non-parametric statistical test (that is, one that does not assume data are normally distributed and have equal variances).

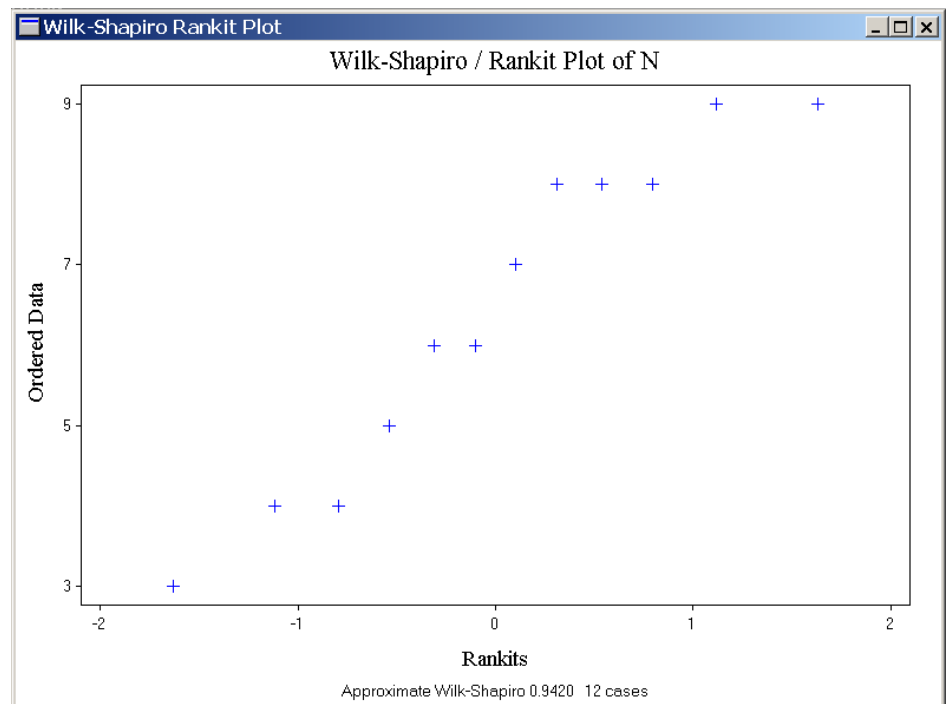
## Testing Data for Normality in Statistix

To see if your data can be considered normal or following a normal distribution, go to the menu Statistics| Randomness/Normality Tests | Wilk-Shapiro/Rank-it Plot... A menu will ask you to enter the appropriate variable.



Note that you can only enter one plot variable at a time. Select the variable you want to check (in this case, I chose "N") and then use the right arrow to move it into the "Plot Variable" box. Now you are ready to run the test, so choose "OK".

A scatter plot (like the one on the right) will be produced that shows your data correlated to what a normal distribution would look like. If your data are normal, the scatter of points should approximately be in a straight line (with minor deviations) and the approximate Wilk-Shapiro value should be close to 1. (The Wilk-Shapiro value is a correlation between your data and data that are normally distributed with the same mean and variance). To get an exact p value for this statistic, use a table of correlation coefficients and the degrees of freedom (number of points - 1).

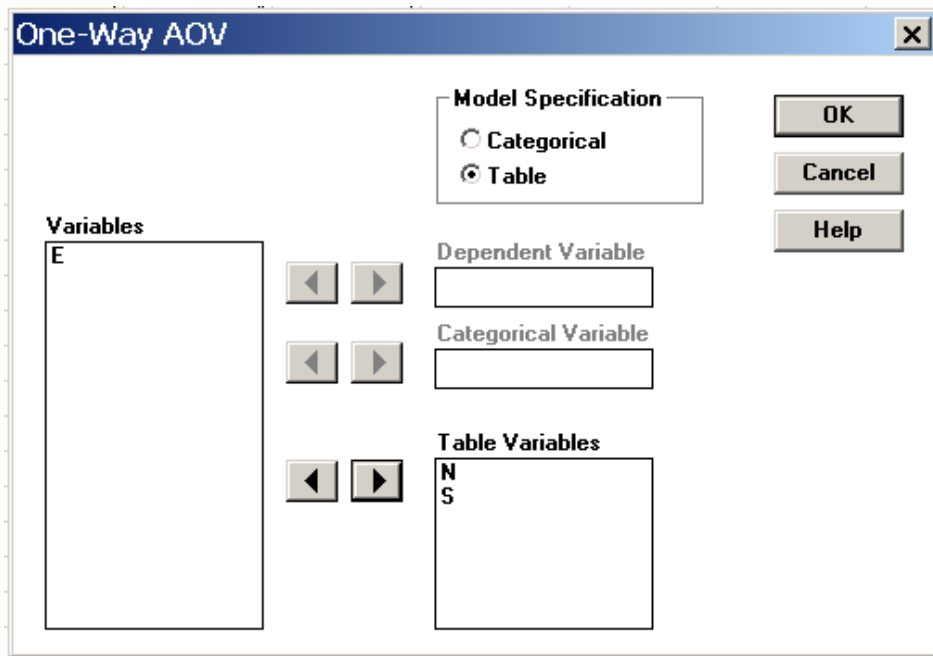


### Potential problems:

Please note that you must have at least seven values in order to use the Wilk-Shapiro value to test for normality. And, since a line is determined by two points, you will need at least three to be able to use the scatter plot. If you do not have enough data points to accurately assess the normality of your sample, you may proceed under the assumption that the data are normal (unless you have reason to believe that they came from a non-normal distribution), but proceed with caution!

## Testing for Equal Variances in Statistix

In Statistix, there is no separate test for equal variances. However, the One-Way AOV automatically calculates two tests of homoscedasticity you can use. The two tests are Bartlett's test of equal variances and Cochran's Q. Also, when you run a two-sample t-test, the output includes a test for equality of variances. Use both of these tests in order to test for equal variances (you don't have to write down the statistics from the other tests you are not interested in – only the results of the equal variance tests). Note that AOV stands for "Analysis Of Variance". It is often abbreviated ANOVA.



Note that if you have entered the variables as separate columns, you must select "Table" under "Model Specification". Then, just use the right arrow to put all the variables into the model. In this example, I am comparing two groups N and S and I will use this One-Way AOV to determine if my samples have equal variances or not.

For the test of equal variances provided in Statistix with the Two-Sample t-Test, see the section entitled "How to Interpret Results of a Two-Sample t-Test".

Note that all tests of normality or equality of variances, the null hypothesis is that the data are normal or variances are equal. Thus, you only have to worry about not meeting assumptions if the associated p value is less than 0.05 (thus rejecting the null hypothesis).

*Results from a One-Way AOV*

ONE-WAY AOV FOR: N S

SOURCE	DF	SS	MS	F	P
BETWEEN	1	70.0417	70.0417	21.13	0.0001
WITHIN	22	72.9167	3.31439		
TOTAL	23	142.958			

	CHI-SQ	DF	P
BARTLETT'S TEST OF EQUAL VARIANCES	0.90	1	0.3418

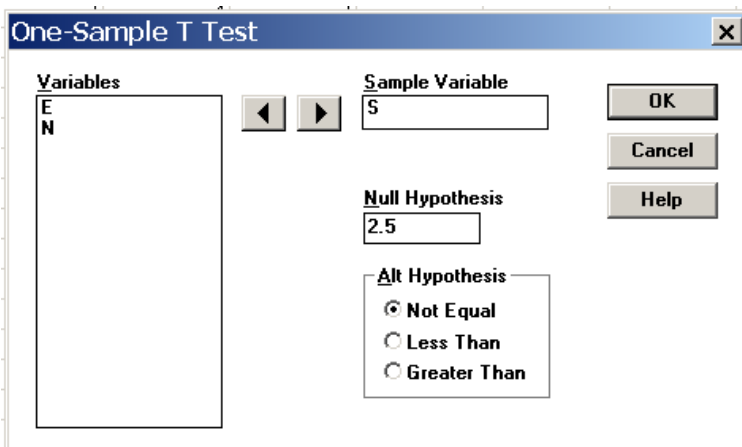
COCHRAN'S Q 0.6434  
LARGEST VAR / SMALLEST VAR 1.8045

COMPONENT OF VARIANCE FOR BETWEEN GROUPS 5.56061  
EFFECTIVE CELL SIZE 12.0

VARIABLE	MEAN	SAMPLE SIZE	GROUP STD DEV
N	6.4167	12	2.0652
S	3.0000	12	1.5374
TOTAL	4.7083	24	1.8205

CASES INCLUDED 24 MISSING CASES 24

*Annotation:* Bartlett's test of equal variances uses the chi-square statistic, be sure to write down the chi-square value, degrees of freedom, and the associated p value (so in this case,  $\chi^2 = 0.9$ ,  $df = 1$ ,  $p = 0.34$ ). Cochran's Q takes the largest variance and divides by the smallest variance to calculate a statistic. The associated p-value in this case is 0.6434 and the ratio that was calculated is 1.8045, so that the larger mean has almost twice the variance than the smaller mean. However, both of these tests in this case, show no deviations from equal variance since both p values are greater than 0.05.

**How to Interpret Results of a One-Sample t-Test**

Note that the One-Sample t-Test is used to compare a sample population to a known value. For example, suppose we read somewhere that all trees tend to grow 2.5 cm in diameter per year. We could see if our results for south-facing slopes fit this statement. You must choose your sample variable ("S" in this case) and enter the known value we are comparing to. The default value is 0.0. In this example, we have to change it to 2.5. Note that you have other options (less than or greater than).

ONE-SAMPLE T TEST FOR S

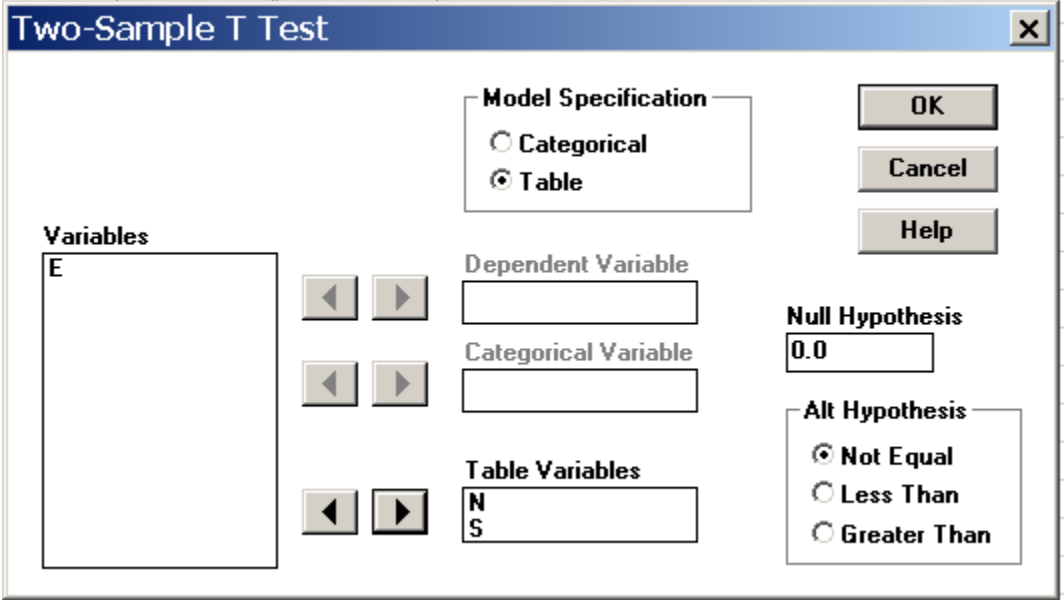
NULL HYPOTHESIS: MU = 2.5	
ALTERNATIVE HYP: MU <> 2.5	
MEAN	3.0000
STD ERROR	0.4438
MEAN - H0	0.5000
LO 95% CI	-0.4768
UP 95% CI	1.4768
T	1.13
DF	11
P	0.2839
CASES INCLUDED 12 MISSING CASES 0	

*Annotation:* Note that this test automatically runs a two-tailed test (see alternative hypothesis). The results provide us with some descriptive statistics (mean, standard error [std error], lower 95% confidence interval, and upper 95% confidence interval around the mean [LO 95% CI, UP 95% CI]). You should make a note of the T value, degrees of freedom, the calculated p value, and the sample size (“CASES INCLUDED”). Note that in this example,  $p > 0.05$ , so we accept the null hypothesis.

**How to Interpret Results of a Two-Sample t-Test**

*Two tailed example*

Notice how to enter your data if you have your data in columns in the Statistix spreadsheet.



TWO-SAMPLE T TESTS FOR N VS S				
VARIABLE	MEAN	SAMPLE SIZE	S.D.	S.E.
N	6.4167	12	2.0652	0.5962
S	3.0000	12	1.5374	0.4438
DIFFERENCE	3.4167			
NULL HYPOTHESIS: DIFFERENCE = 0				
ALTERNATIVE HYP: DIFFERENCE > 0				
ASSUMPTION	T	DF	P	95% CI FOR DIFFERENCE
EQUAL VARIANCES	4.60	22	0.0001	(1.8753, 4.9580)
UNEQUAL VARIANCES	4.60	20.3	0.0001	(1.8679, 4.9654)
TESTS FOR EQUALITY OF VARIANCES	F	NUM DF	DEN DF	P
	1.80	11	11	0.1710
CASES INCLUDED 24 MISSING CASES 24				

*Annotation:* Notice that the program automatically calculates sample size, means, standard deviations (S.D.) and standard errors (S.E.) for each variable. This procedure also automatically tests for equality of variances. The null hypothesis for the test for equality of variances is that the variances ARE equal. You can report the test for equality of variances as well ( $F = 1.8$ ,  $df = 11$ ,  $p = 0.17$ ). Since the variances are equal, you must report the results from the EQUAL VARIANCES test which has  $T = 4.6$ ,  $df = 22$ , and  $p = 0.0001$ . Since  $p < 0.05$ , in this case, N is significantly greater than S. If you are using the unequal variance test, notice that the degrees of freedom include a fraction. This is how the test adjusts for unequal variances. Also note that if the value under P is “0.0000”, then you should report that  $P < 0.0001$  since the value is never actually zero, but it may be a number less than the significant digits this program can compute.

### One-tailed example

The screenshot shows the 'Two-Sample T Test' dialog box. On the left, a list of variables contains 'E'. In the center, 'E' is entered in the 'Dependent Variable' field, and 'N' and 'S' are entered in the 'Table Variables' field. On the right, the 'Model Specification' section has 'Table' selected. The 'Null Hypothesis' is set to '0.0', and the 'Alt Hypothesis' is set to 'Less Than'. Buttons for 'OK', 'Cancel', and 'Help' are visible on the right side.

TWO-SAMPLE T TESTS FOR N VS S				
VARIABLE	MEAN	SAMPLE SIZE	S.D.	S.E.
N	6.4167	12	2.0652	0.5962
S	3.0000	12	1.5374	0.4438
DIFFERENCE	3.4167			
NULL HYPOTHESIS: DIFFERENCE = 0				
ALTERNATIVE HYP: DIFFERENCE < 0				
ASSUMPTION	T	DF	P	95% CI FOR DIFFERENCE
EQUAL VARIANCES	4.60	22	0.9999	(1.8753, 4.9580)
UNEQUAL VARIANCES	4.60	20.3	0.9999	(1.8679, 4.9654)
TESTS FOR EQUALITY OF VARIANCES	F	NUM DF	DEN DF	P
	1.80	11	11	0.1710
CASES INCLUDED 24 MISSING CASES 0				

*Annotation:* Note that when doing a one-tailed test, you have to choose “Less than” or “Greater than” options. These options use the values you gave IN THE ORDER YOU ENTERED THEM. So, in this case, I entered N first and then S (N is above S in the “Tables” box) and chose the option “Less than”. Note that again we have equal variances since the p value for the test for equality of variances is **not** < 0.05.

### How to Interpret the Results of a Rank Sum (Mann-Whitney U) Test

Enter data similar to way you enter data for a Two-sample T-test (see that section for details).

RANK SUM TWO-SAMPLE (MANN-WHITNEY) TEST FOR N VS S				
VARIABLE	RANK SUM	SAMPLE SIZE	U STAT	MEAN RANK
N	208.00	12	130.00	17.3
S	92.000	12	14.000	7.7
TOTAL	300.00	24		
NORMAL APPROXIMATION WITH CONTINUITY CORRECTION				3.320
TWO-TAILED P-VALUE FOR NORMAL APPROXIMATION				0.0009
TOTAL NUMBER OF VALUES THAT WERE TIED				23
MAXIMUM DIFFERENCE ALLOWED BETWEEN TIES				0.00001
CASES INCLUDED 24 MISSING CASES 0				

*Annotation:* The statistic is called a U value and in this case is 3.320 with an associated p value of 0.0009. This test compares the order of data (ranks) instead of the actual numbers of data themselves (notice the higher p value here as compared to the results of the t-test). Notice that there are no degrees of freedom for this test since it uses all ranks to calculate rank sums on which it bases the statistic (no assumptions of normality or equal variances).

## How to Interpret the Results of the Kruskal-Wallis One-Way AOV

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV					
VARIABLE	MEAN RANK	SAMPLE SIZE			
N	17.3	12			
S	7.7	12			
TOTAL	12.5	24			
KRUSKAL-WALLIS STATISTIC					11.3866
P-VALUE, USING CHI-SQUARED APPROXIMATION					0.0007
PARAMETRIC AOV APPLIED TO RANKS					
SOURCE	DF	SS	MS	F	P
BETWEEN	1	560.667	560.667	21.57	0.0001
WITHIN	22	571.833	25.9924		
TOTAL	23	1132.50			
TOTAL NUMBER OF VALUES THAT WERE TIED					23
MAX. DIFF. ALLOWED BETWEEN TIES					0.00001
CASES INCLUDED 24		MISSING CASES 24			

*Annotation:* For this test, you can report either the Kruskal-Wallis statistic (H) and the associated p value (in this case,  $H = 11.3866$  with  $p = 0.0007$ ) or the parametric analysis of variance that has been performed on the ranks (in this case,  $F = 21.57$  with  $p = 0.0001$ ). If the results are not in agreement, choose the Kruskal-Wallis test as the more conservative of the two, since you have violations of equal variances (which would still be required of ranks for the parametric AOV). Sample size is given along with means of ranks. **WATCH OUT!!** report actual means and standard deviations of your data in your report, NOT the means of the ranks. Note that this test also allows for multiple comparisons (if you have more than two groups). See the section on the One-Way AOV for how to do this.

## How to Interpret the Results of a Paired t-Test

Note that in this test, as in the Two-sample T test, you can choose to have a two-tailed (Alt Hypothesis = Not Equal) or a one-tailed test (Alt Hypothesis = Less Than or Greater Than). As before, the ORDER in which you put the variables will determine which value will be used as the lesser or greater value (this will be spelled out in the output).

PAIRED T TEST FOR N - S

NULL HYPOTHESIS: DIFFERENCE = 0  
ALTERNATIVE HYP: DIFFERENCE <> 0

MEAN	-3.4167
STD ERROR	0.3362
LO 95% CI	-4.1566
UP 95% CI	-2.6768
T	-10.16
DF	11
P	0.0000

CASES INCLUDED 12      MISSING CASES 0

*Annotation:* Null and alternative hypotheses are explicitly stated. If you are doing a one-tailed test, be sure the correct hypothesis is being tested. In this case the S values are subtracted from N values (N-S). Note that the mean, standard error and lower and upper 95% confidence intervals are given for the **differences** between pairs. **WATCH OUT!!** You want to report the means and standard deviations of the groups, NOT their differences. Also remember that the order in which you input the data is critical in this test since the data are paired. The sample size is the value given for “CASES INCLUDED” or one more than the degrees of freedom (DF). In the above example, there is a negative value for T. Since the T distribution is symmetrical, you can report this as either a negative or positive value (the associated P value will be the same). The T value is negative because the difference between N and S gave negative values (notice that the mean is negative).

### How to Interpret the Results of a Wilcoxon Signed-Rank Test

Enter data similar to way you enter data for a Two-sample T-test (see that section for details).

WILCOXON SIGNED RANK TEST FOR N - S

SUM OF NEGATIVE RANKS	0.0000
SUM OF POSITIVE RANKS	78.000

EXACT PROBABILITY OF A RESULT AS OR MORE EXTREME THAN THE OBSERVED RANKS (1 TAILED P-VALUE)	0.0002
--	--------

NORMAL APPROXIMATION WITH CONTINUITY CORRECTION	3.020
TWO TAILED P-VALUE FOR NORMAL APPROXIMATION	0.0025

TOTAL NUMBER OF VALUES THAT WERE TIED	10
NUMBER OF ZERO DIFFERENCES DROPPED	0
MAX. DIFF. ALLOWED BETWEEN TIES	0.00001

CASES INCLUDED 12      MISSING CASES 12

*Annotation:* This test uses the ranks of the data rather than the numbers themselves. This frees the analysis of distribution and variance problems. Note that there is no value for degrees of freedom since all data are used and converted to ranks. The Wilcoxon Signed-Rank calculates a Z statistic (in this case equal to 3.02 with an associated p value of 0.0025). The sample size is given in “CASES INCLUDED”, in this case 12. So, for this test, the information you need is:  $Z = 3.02$ ,  $p < 0.005$ ,  $n = 12$ ).

### How to Interpret the Results of a Sign Test

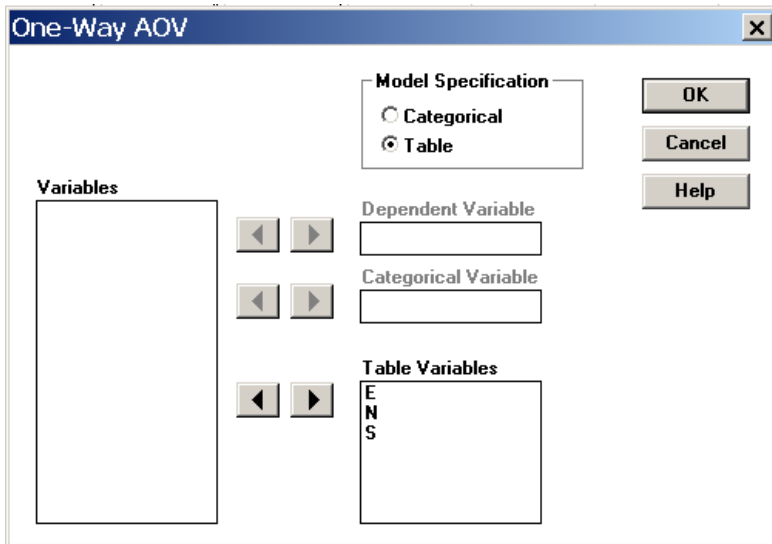
SIGN TEST FOR N - S

NUMBER OF NEGATIVE DIFFERENCES	0
NUMBER OF POSITIVE DIFFERENCES	12

NUMBER OF ZERO DIFFERENCES (IGNORED)	0
PROBABILITY OF A RESULT AS OR MORE EXTREME THAN OBSERVED	0.0002
A VALUE IS COUNTED AS A ZERO IF ITS ABSOLUTE VALUE IS LESS THAN 0.00001	
CASES INCLUDED 12	MISSING CASES 12

*Annotation:* The sign test simply compares the paired values and keeps track of the number of positive and negative differences. It is not as powerful as the Wilcoxon Signed Rank Test, but it is simple and non-parametric. Notice that there are no degrees of freedom associated with this test since all data are compared. There is no statistic value (it is calculated as a probability only), but you must report sample size (“CASES INCLUDED”) and the calculated p value (in this case, 0.0002).

### How to Interpret the Results of a One-Way Analysis of Variance



Note that you can enter more than two table variables as shown on the left.

Note that this test compares means for greater than two groups. For this example, we have added a third group called “E” (east-facing slopes). The test compares the amount of variance explained between groups and within groups. Simply put, if the variation within groups is greater than that between groups, the differences will not be significantly different. Notice that you get two tests of equal variances “for free” when you run this test (see section called “Testing for

Equal Variances in Statistix” to interpret these results). You also get some descriptive statistics for each of the groups (means, sample size, and standard deviations of each group). Note that the sample sizes do not have to be equal in order to run this test, although it is a more powerful test if the sample sizes are equal.

You should also note that there are many complicated and powerful variations on the One-Way Analysis of Variance which are covered in Statistix using the General AOV/ACOV under “Linear Models”. You can add a covariate, interaction terms, and create complex and intriguing models. However, these will not be covered in this handbook.

ONE-WAY AOV FOR: E N S

SOURCE	DF	SS	MS	F	P
BETWEEN	2	161.167	80.5833	20.17	0.0000
WITHIN	33	131.833	3.99495		
TOTAL	35	293.000			

	CHI-SQ	DF	P
BARTLETT'S TEST OF EQUAL VARIANCES	1.76	2	0.4154
COCHRAN'S Q		0.4469	
LARGEST VAR / SMALLEST VAR		2.2660	
COMPONENT OF VARIANCE FOR BETWEEN GROUPS			6.38237
EFFECTIVE CELL SIZE			12.0

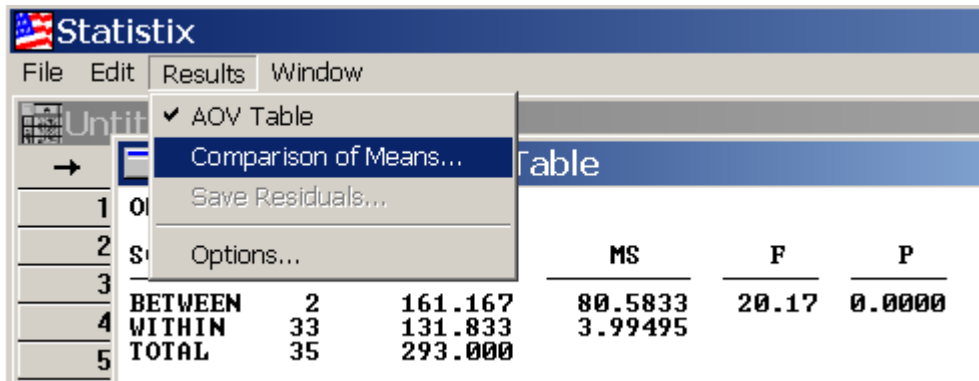
VARIABLE	MEAN	SAMPLE SIZE	GROUP STD DEV
E	8.0833	12	2.3143
N	6.4167	12	2.0652
S	3.0000	12	1.5374
TOTAL	5.8333	36	1.9987

CASES INCLUDED 36      MISSING CASES 0

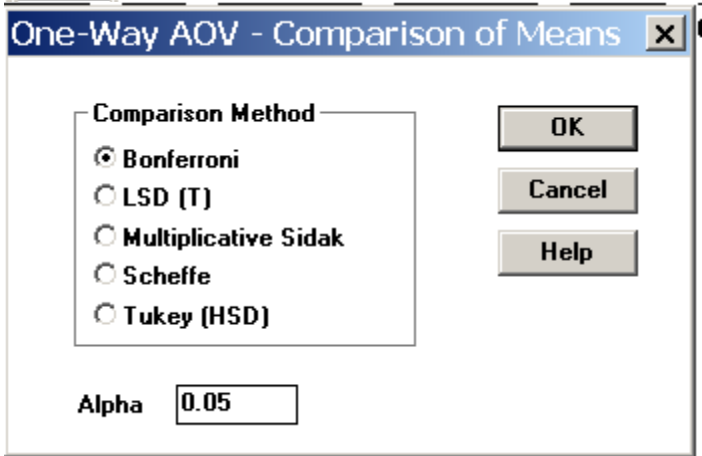
*Annotation:* The statistic the One-Way AOV uses is the F statistic (in this case,  $F = 20.17$  and  $p < 0.0001$ , note that  $p$  is NEVER zero). Sample size is the mean sample size of each group (in this case 12). If the calculated  $p$  value is  $< 0.05$ , then all you can say is that there is a difference between at least one pair of groups. Logically, what you can say for sure is that the group with the smallest mean is significantly smaller than the group with the largest mean.

The next step: If you get a significant  $p$  value for your variable, you will need to run a multiple

comparison test. You must do this WHILE YOUR AOV RESULTS ARE ON THE SCREEN!!



When your AOV results are visible, you can go to the "Results" menu and select "Comparison of Means".



You have several options for comparison method. They are listed from the most stringent to most lax in their ability to differentiate groups. You may wonder why we need so many choices. Well, some methods are more effective when your data are not normal (Scheffe) than others. Typically, you should use the Bonferroni method.

*One-Way AOV Comparison of Means*

BONFERRONI COMPARISON OF MEANS			
VARIABLE	MEAN	HOMOGENEOUS GROUPS	
E	8.0833	I	
N	6.4167	I	
S	3.0000	.. I	
THERE ARE 2 GROUPS IN WHICH THE MEANS ARE NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.			
CRITICAL T VALUE		2.522	REJECTION LEVEL 0.050
CRITICAL VALUE FOR COMPARISON		2.0581	
STANDARD ERROR FOR COMPARISON		0.8160	

*Annotation:* In this example, E and N are not significantly different from each other, although they are both significantly different than S. You can see that using the notation under “Homogeneous groups” (groups that have an “I” at the same level are not statistically different from each other) as well as by using the “Critical Value for Comparison.” The values need to be at least 2.06 greater than each other to be statistically significantly different and E and N are not that different from each other (difference is 1.6). However, the difference between N and S is 2.6, and E and S is about 5, so these are significantly different from each other.

**How to Calculate the Chi Square Statistic**

The chi square test is used on categorical count data. If we were to flip a coin 100 times, we would expect the coin to end up heads 50% and tails 50% of the time. In reality, however, maybe we got 45 heads and 55 tails. Is this really different from a 50/50 ratio? To determine if there is a significant difference between the observed and expected ratios, we perform the chi square test:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O is the observed value and E is the expected value.

	heads	tails
<b>observed (O)</b>	45	55
<b>expected (E)</b>	50	50
<b>(O-E)<sup>2</sup></b>	25	25
<b>(O-E)<sup>2</sup>/E</b>	0.5	0.5

The sum of (O-E)<sup>2</sup>/E gives us our Chi square (X<sup>2</sup>) value; in this case,  $\chi^2 = 1$ .

Examine a chi square table at 0.05 probability with one degree of freedom (two classes-- heads and tails-- minus one). The critical value according to the table is about 3.8. Since our calculated  $\chi^2$  is less than the critical value, we do not have a significant difference from what we expected according to the null hypothesis.

This test is commonly used in genetics. In genetics, we often have a set hypothesis about the genetic relationship of traits. In that case, we do not expect our results to differ from the null hypothesis. When we accept the null hypothesis (which defines the genetic relationships we expect) we support the relationships of genes and traits we are studying.

### How to Calculate the G Statistic

Let us take the same example used in calculating the chi-square above. Suppose you flipped a coin 100 times and it came up heads 45 times. Is this significantly different than the expected 50%? The formula we use for calculating the G statistic is as follows:

$$G = 2 \sum O \ln(O/E)$$

Where  $P_E$  is the expected proportion and  $P_O$  is the observed proportion.

So, we can fill in the table below...

	Heads	Tails
<b>Observed frequency (O)</b>	45	55
<b>Expected frequency (E)</b>	50	50
<b>O/E</b>	0.9	1.1
<b>ln(O/E)</b>	-0.1053	0.0953
<b>O ln(O/E)</b>	-4.7385	5.2415

Sum results to get: 0.503

And multiply by 2 to get  $G = 1.06$

Then, use a chi-square table to find the critical value. Our calculated value is again below the 3.8 critical value, so the null hypothesis is rejected.

There is currently much debate over using the G statistic over Chi-square. The leading biostatistics textbook (Sokal & Rohlf, 1995) does not even include the Chi-square test as discussed here. It is, however, a very old test that is still widely used. Note that neither test can be used if an expected category is zero (due to a mathematical problem in the calculation – you can't divide by zero).

### What Information You Need From a Statistical Analysis

Once you have described and analyzed your data properly, you often need to incorporate your results into a report or essay. These are the pieces of information you will need to make notes on so that you can report your results properly:

- The sample size (n)
- Degrees of freedom (df) (some tests do not have this)
- The value of the statistic (Chi-square, t, F, etc.) (only one test does not have this)

- The p value (or probability value) that is associated with the value of the statistic (this value is always calculated)
- Descriptive data (mean and some measure of dispersion, such as variance, standard deviation, or standard error – the standard deviation or standard error are most commonly used).

### *Sample size (n) and Degrees of freedom (df)*

The sample size is the number of independent samples that were taken and used in the statistical analysis. In the lizard example where we measured running speeds of lizards at 20°C and 40°C, the sample size is 50 even though we took a total of 100 measurements because there are only 50 independent samples for statistical purposes (the paired test actually uses the **differences** between the two measurements).

The degrees of freedom of a test is closely linked to the sample size (n) used. The degrees of freedom of a test is often one or two less than the sample size. This has mostly to do with the fact that you are taking a sample from a larger pool of possible samples. Thus the exact sample (if you were to repeat the experiment) changes every time, so there is some uncertainty associated with it.

### **How to Present Results of a Statistical Analysis in the Text of a Report**

Suppose you have counted a total of 115 flies and categorized them into red, white, and brown eyes and run a chi-square analysis on the data. Your chi-square value was 1.4 and the associated p value was .5. This is an example of how you might present your results in a report:

“Eye color of flies did not deviate significantly from the expected 1:2:1 ratio ( $n = 115$ ,  $\chi^2 = 1.4$ ,  $p > 0.05$ ).”

You can also explicitly state the associated p value if you choose:

“Eye color of flies did not deviate significantly from the expected 1:2:1 ratio ( $n = 115$ ,  $\chi^2 = 1.4$ ,  $p = 0.65$ ).”

Notice that we have used the word “**significantly**” in this sentence. The **ONLY** time you can use this word, is if you have done a statistical analysis on the data and the null hypothesis has been tested.

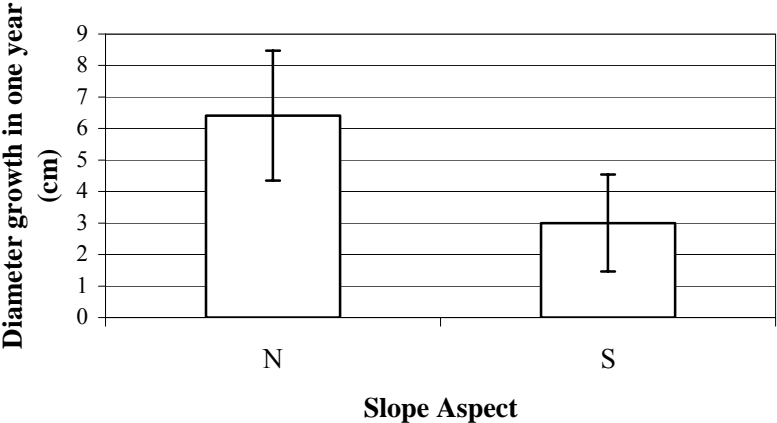
As a second example, suppose you have measured the heights of 50 oak trees 25 of which are growing on north facing slopes and 25 are on south facing slopes. You run a two-sample t-test on the data and get a t value of -4.6 and an associated p value of 0.0001. You can present your results as follows:

“Trees on north-facing slopes were significantly taller than those on south-facing slopes ( $n = 25$ ,  $t = 4.6$ ,  $df = 24$ ,  $p < 0.001$ ).”

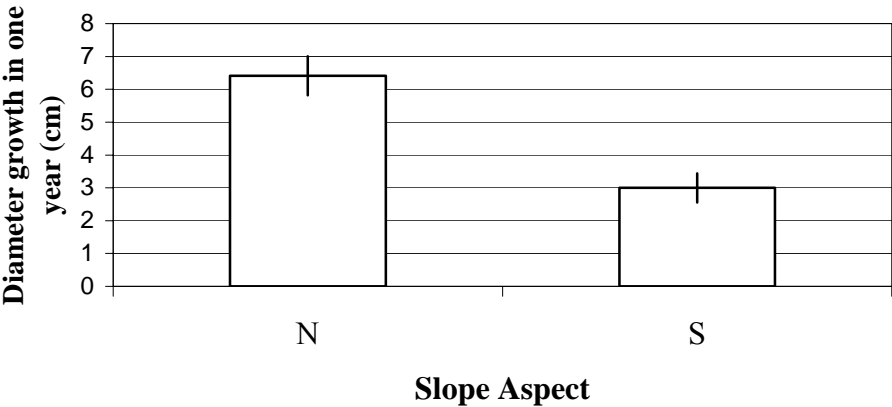
Notice that we have made a statement about the data and have used our statistical analysis to back it up. Also notice that we have noted that p is less than 0.001 (since it was actually equal to 0.0001). When we reject the null hypothesis, it is most common to say the p value was less than 0.05, 0.01, 0.001, etc. since these are considered critical values. Also notice that even though you may sometimes get a negative t value, it is usually reported as a positive number (the distribution is symmetrical).

**How to Present Results of a Statistical Analysis by using Graphs**

Suppose we want to include a figure in our report showing the different means of the N- and S-facing groups. See a guide on proper labeling and information needed for graphs available to you, such as Pechenik (2001). No matter what kind of graphing you do, however, you should always include some measure of central tendency with your means. The following two graphs compare using standard deviations and standard errors.



**Figure 9.** Means of tree growth from north and south-facing slopes. Error bars denote standard deviation.



**Figure 10.** Means of tree growth from north and south-facing slopes. Error bars denote standard errors.

If you compare Figures 9 & 10, you will see why standard errors are generally preferred by academic journals. You know from working with this data set (for example, see “Results of a Two-Sample t-Test”), that these two groups ARE statistically significantly different from each other. In general, if the standard error bars do not overlap, groups are significantly different. Standard error is adjusted for the number of samples. On the other hand, standard deviation, while in the units of the mean (and therefore appropriately placed on the same graph) are always larger and in this case do overlap between groups (Figure 9). No matter which measure of central tendency you use, you should definitely use one and be consistent in your paper. (Note that it is NOT appropriate to use variance since it is a squared measure of central tendency and therefore should not be on the same scale as the means.)

Note that these graphs were created in Excel. To get error bars into Excel graphs, you have to create the graph and save it as an object in your worksheet. Then, right click on the bars for which you want to add error bars. Choose “Format Data Series” and click on the “Y Error Bars” tab. You can choose if you want error bars only going up or down or in both directions. Be sure you have the standard errors in your worksheet. Click on the “Custom” button and click on the small worksheet icon to select the numbers off of your worksheet (or write in the cell names, such as B4:C4 in the example below to get standard errors). If you want both + and – bars, you will have to enter the same numbers in twice.

*To create the above graphs, the data in Excel looked like this:*

	<b>A</b>	<b>B</b>	<b>C</b>
<b>1</b>		N	S
<b>2</b>	Mean	6.41	3
<b>3</b>	SD	2.0652	1.5374
<b>4</b>	SE	0.5962	0.4438

## Data Used in Examples of the Handbook

*Data used in all the Statistix examples given in this appendix*

N data set = 4, 3, 7, 8, 9, 8, 6, 5, 4, 8, 9, 6

S data set = 2, 1, 4, 3, 3, 5, 3, 2, 1, 4, 6, 2

E data set = 5, 6, 7, 6, 7, 8, 9, 8, 10, 11, 13, 7

(Note that order of data is important when doing paired tests.)

*Data in Statistix looked like this:*

	<b>N</b>	<b>S</b>	<b>E</b>
<b>1</b>	4	2	5
<b>2</b>	3	1	6
<b>3</b>	7	4	7
<b>4</b>	8	3	6
<b>5</b>	9	3	7
<b>6</b>	8	5	8
<b>7</b>	6	3	9
<b>8</b>	5	2	8
<b>9</b>	4	1	10
<b>10</b>	8	4	11
<b>11</b>	9	6	13
<b>12</b>	6	2	7

Note: The “Table” option was always selected and then the variables were selected.

## References

Gould, J.L. and G.F. Gould. 2002. Biostats Basics: A Student Handbook. W.H. Freeman & Co, NY.

Pechenik, J.A. 2001. A Short Guide to Writing about Biology. Longman, NY.

Sokal, R.R. and F.J. Rohlf. 1995. Biometry: The Principles and Practice of Statistics in Biological Research. 3<sup>rd</sup> Edition. W.H. Freeman & Co., NY.