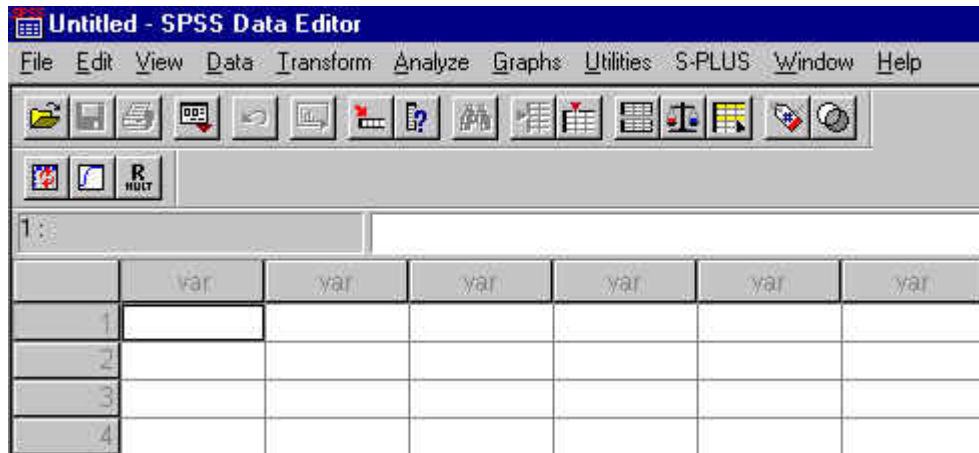


BIOLOGY 458

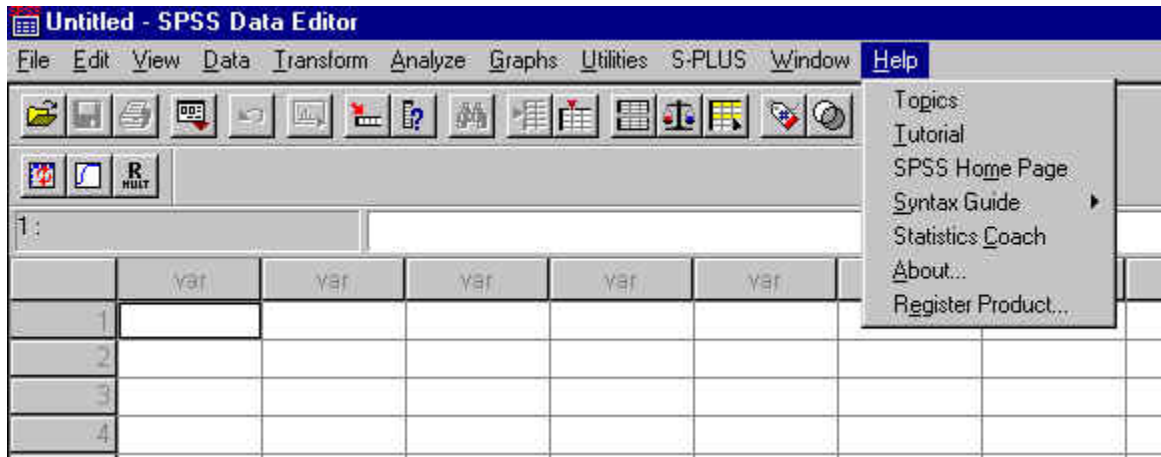
LAB 1 - AN INTRODUCTION TO SPSS

INTRODUCTION

SPSS (statistical package for the social sciences) is one of many statistical software packages available today. It is widely used by scientists and social scientists since it is a comprehensive package of numerous procedures for data analysis and interpretation. Mainframe, PC, and MAC versions of **SPSS** are available. Throughout this course, we will be learning how to use the PC or MAC versions of this program. **SPSS** is a menu driven "point and click" software package. When you first enter **SPSS**, the screen you will see is called the **Data Editor Window**.



A "toolbar" with pull-down menus for various commands is found at the top of the **Data Editor Window**. On the right-hand end of the toolbar in the **Data Editor Window** is the **Help Menu**. Like most other software packages **SPSS** offers extensive explanatory information. In addition to the encyclopedic list of help topics which may be accessed from the **Topics** command in the **Help Menu**, **SPSS** also offers a **Tutorial**, and a **Statistics Coach** which are both accessed from the **Help Menu**. The **Tutorial** is an interactive introduction to numerous aspects of using **SPSS**. It may be entered at a number of different points, so it is useful anytime you are attempting to use **SPSS** in a novel way. The **Statistics Coach** is also an interactive guide to data analysis and interpretation. **SPSS** asks you a series of questions and your answers lead the **SPSS Statistics Coach** to recommend what procedures to apply to answer the question you pose for the data you have.



Point To Remember: While **SPSS** offers an amazing array of options for data transformation, analysis, graphing, and output, only a small number of commands need to be mastered to find **SPSS** useful immediately. I personally do not know all the bells and whistles of **SPSS**, particularly the Windows based interface since I have used mainframe and DOS versions of **SPSS** for many years. I learn more as I need to use **SPSS** in new ways. Feel free to explore this program as much as you like!

Editorial Note: I have attempted to put the names of the various *Menus* in **SPSS** in ***bold and italic*** and the specific **Commands** available in a menu in **bold and underline**.

GETTING STARTED

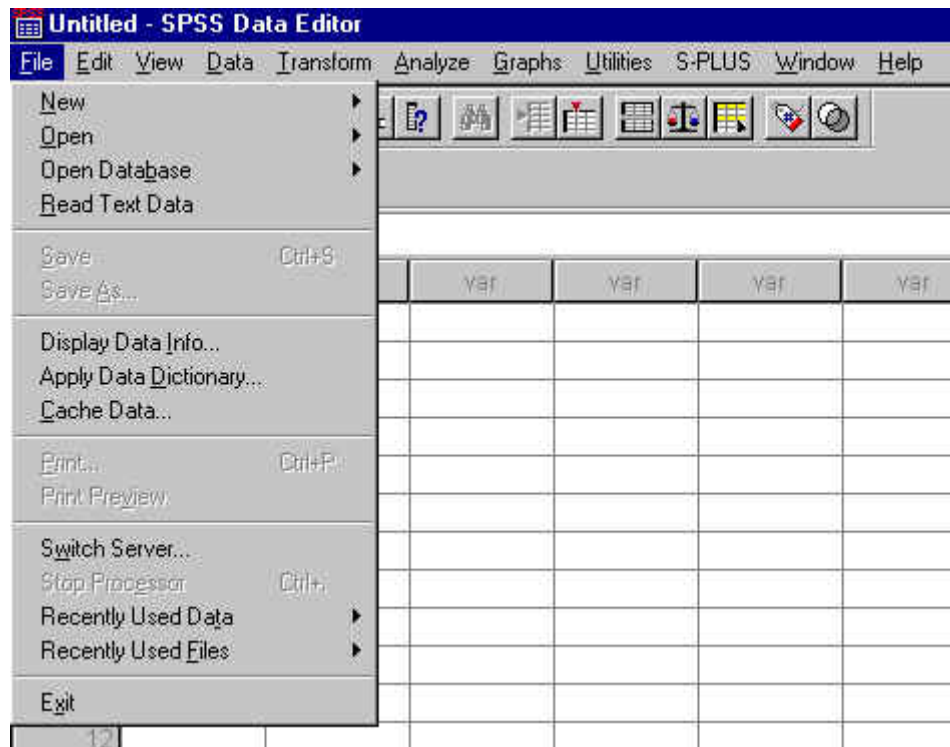
I have found the ***File***, ***Transform***, ***Statistics***, ***Graphs*** and ***Help Menus*** to be those that I use most frequently in **SPSS**. The five main steps in an **SPSS** session are

- [Enter the Data](#)
 - [Transforming Data Values](#)
 - [Analyzing Data](#)
 - [Graphing Data](#)
 - [Examining, Printing, and Saving Results](#)
-

Data Entry

Opening or Reading Data Files

Normally, the first item of business in any **SPSS** session is to open a data file, read in a data file, or to type in your data directly into the *Table* in the **Data Editor Window**. Data files may be "opened" using the **OPEN** command if they are **SPSS** files with the extension (.sav) or using the **OPEN DATABASE** command if they are *spreadsheet* files. They are read in using the **READ TEXT DATA** command if they are *ASCII*, *Text* or *DOS* files. The commands **Open**, **OPEN DATABASE**, and **READ TEXT DATA** are found in the **File Menu**. I have had less success opening **Excel** spreadsheet files using the instructions provided in the **Help Menu**, than I had transferring data into **SPSS** from **Excel** spreadsheets by highlighting the data I wanted to transfer, copying it to the clipboard, and pasting it into **SPSS**. The data can then be saved as an **SPSS** file for future analysis.



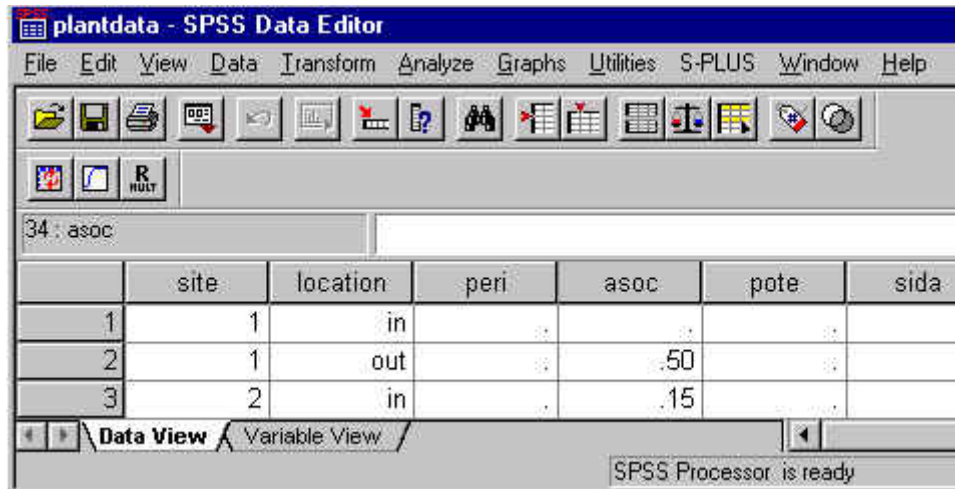
Typing Data into the SPSS Data Editor Window

You can enter your data directly into the *Table* in the **SPSS Data Editor Window**. Each line in the *Table* represents a different experimental subject and each column a different variable. You can use integer codes to represent particular treatment levels. Variable names can be entered into the *Table* by double-clicking the left mouse button on the variable name box at the top of a column and typing in a name in the appropriate sub-window.

Defining Missing Values

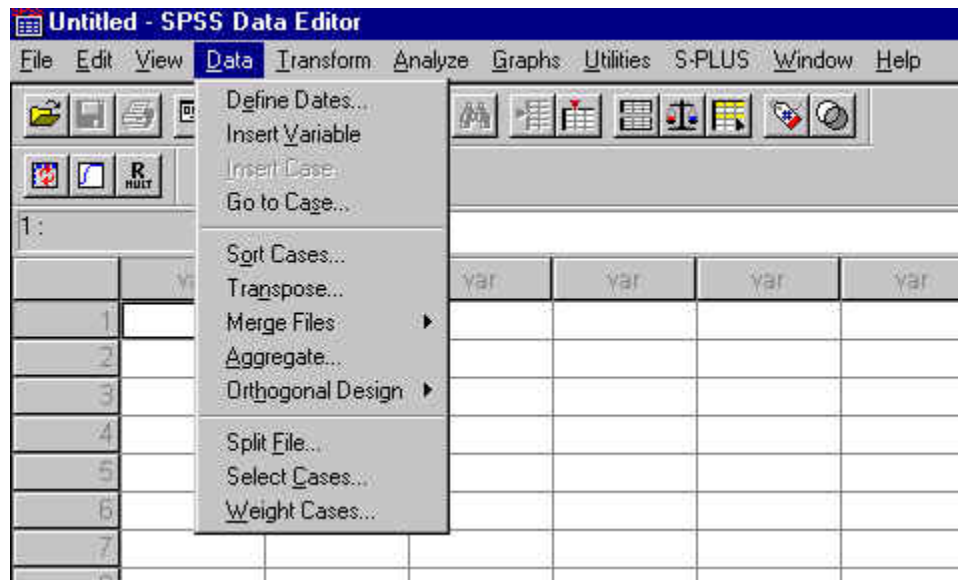
Often we are unable to collect some data on particular experimental subjects. When this occurs, we usually enter some value into the data file to hold the place of this "missing

observation." The value could be a set of blank spaces or a numeric value that is easy to distinguish from the actual observations. For example, one might enter the value "-1" for data on species diversity. Since species diversity must always be a positive real number, missing observations assigned the value "-1" are easy to distinguish from actual observations. In **SPSS** in the **DATA EDITOR WINDOW**, click on the command **VARIABLE VIEW** tab at the bottom of the window to see the options for defining the attributes of each of your variables. Note that one option is to define the value you use to indicate a missing value.



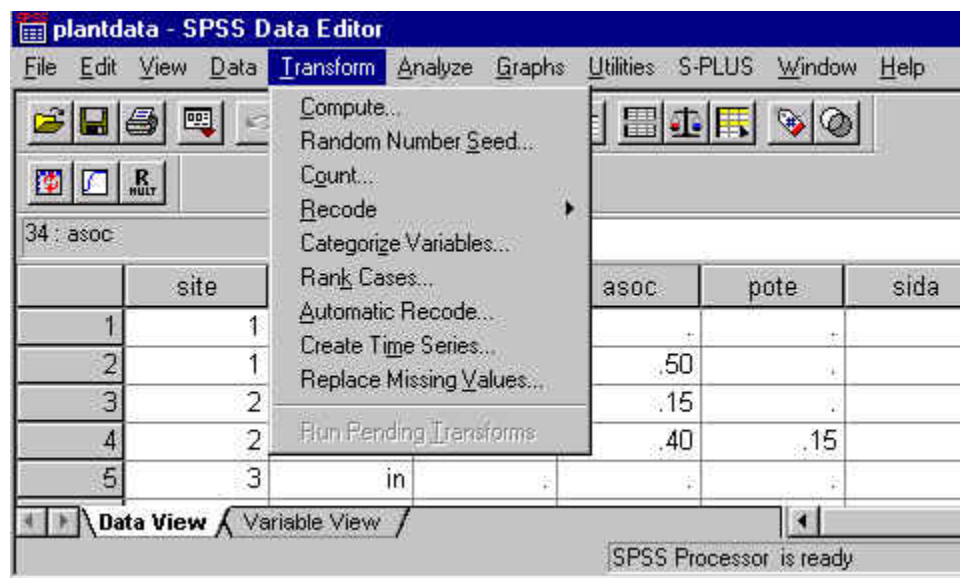
Selecting Subsets of the Data for Analysis

Sometimes we wish to only select a subset of the experimental subjects or to select a subset of the data from the entire range of values a particular variable may cover. For example, if we have data in which one column of the data file contains integer codes for sex with 1 indicating male and 2 indicating female, we may wish to perform an analysis or graph the data for females only. In the **Data Menu**, click on the command **Select Cases**.



Transforming Data Values

SPSS offers a number of options for transforming data values. However, I have found that the **Compute** and **Recode** commands under the **Transform Menu** are the ones I use most often.



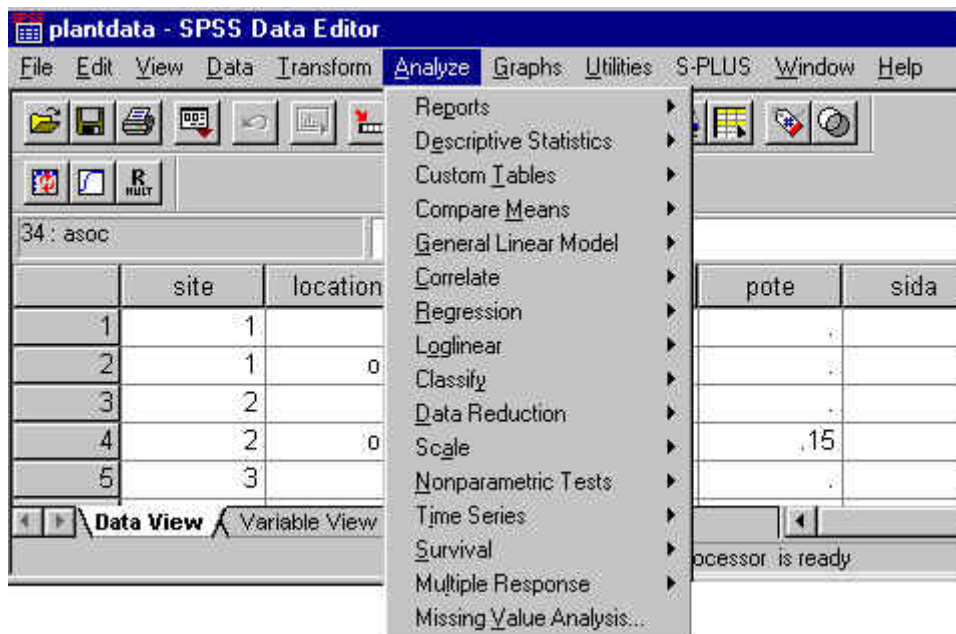
Compute - The **Compute** command allows you to perform a wide variety of numerical transformations to your data. Simply click on the **Compute** command, enter the name of the new variable you wish to create in the *Target Variable* sub-window, click on the name of an existing variable and the right arrow button to move that variable to the *Numeric Expression* sub-window, and then use the calculator like keypad or the *Functions* sub-window and button to construct your data transformation as an equation. Click on **OK** to have the transformation performed. **SPSS** will transform the data

according to the equation you entered, and place the new variable in the next empty column in the data file or *Table*.

Recode - The **Recode** command is useful as a means of redefining data values for the purpose of aggregating data from different categories or for replacing data values coded in ways that **SPSS** does not like. For example, if we had an additional variable in our data file that categorized animals into *age* categories, with 1 indicating infants, 2 indicating juveniles, 3 indicating sub-adults, 4 indicating sexually active adults, and 5 indicating non-sexually active adults, we might wish to perform an analysis that lumps all adult and sub-adult categories into just adults and non-adults. The **Recode** command would allow us to redefine values of the variable *age* to achieve this result. Simply click on the **Transform Menu** and the command **Recode**. Note that one can recode the existing data, that is, place the recoded values in the age column of the data file in place of the old codes, or create a new variable in the next empty column of the data file or *Table*, with the recoded values as a newly created variable. I tend to do the latter, since I often screw-up the recoding and need to repeat the process several times before I achieve my desired recoding. Again the exact process is a mouse and button clicking helluva good time.

Analyzing Data

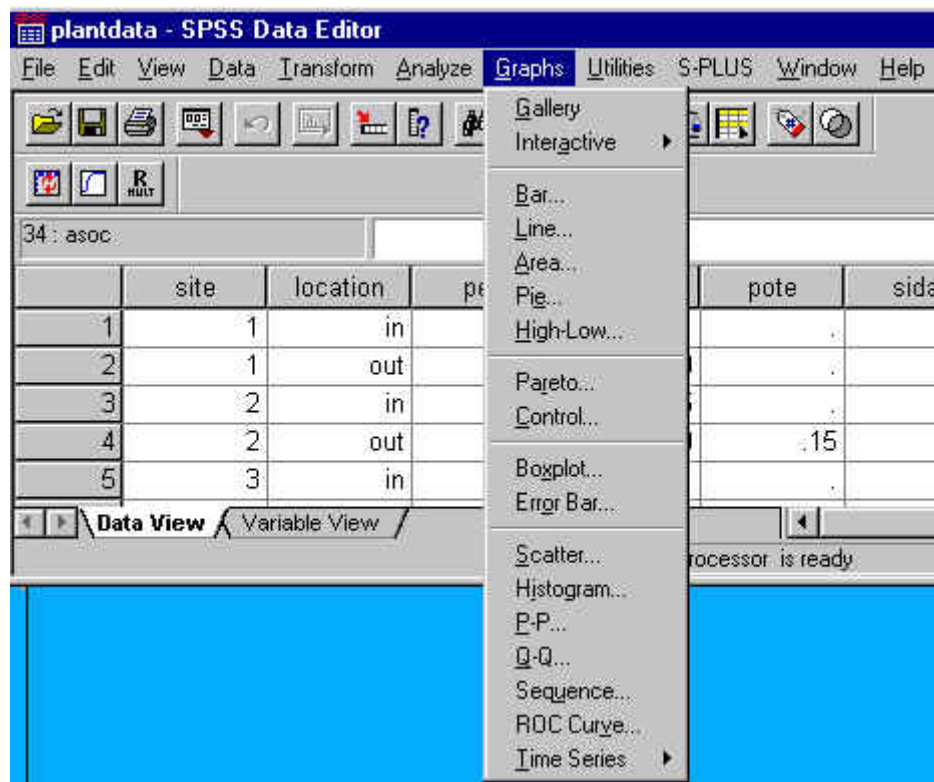
In **SPSS 10.0** the **Analyze Menu** is where one finds all the procedures in **SPSS** for the analysis of data. The commands **Descriptives** and **Custom Tables** are used to compute various summary statistics, such as means, variances, standard deviations, standard errors, skewness, kurtosis, and others for particular variables for all your data or for subsets of your data. Under **Descriptives**, the option **Explore** is particularly useful. The arrows on the right side of the **Analyze** menu indicate that more options are available by following the arrow with the mouse and cursor.



The commands **Compare Means**, **General Linear Model**, and **Nonparametric Tests** are used for most analyses in which the responses of subjects to a variety of experimental treatments or other categorical conditions are contrasted. The commands **Correlation** and **Regression** are used when examining the relationships between two or more continuous variables observed on each subject. These are the primary commands concerned with data analysis that we will use during this course.

Graphing Data

SPSS provides a variety of ways to graph your data for visual presentations mostly from the **Graphs Menu**, and considerable ability to edit and customize your graphs to suit your specific needs. Many of the procedures accessed from the **Analyze Menu** also have the option of generating graphs. You can save graphs to a disk file or print them out.



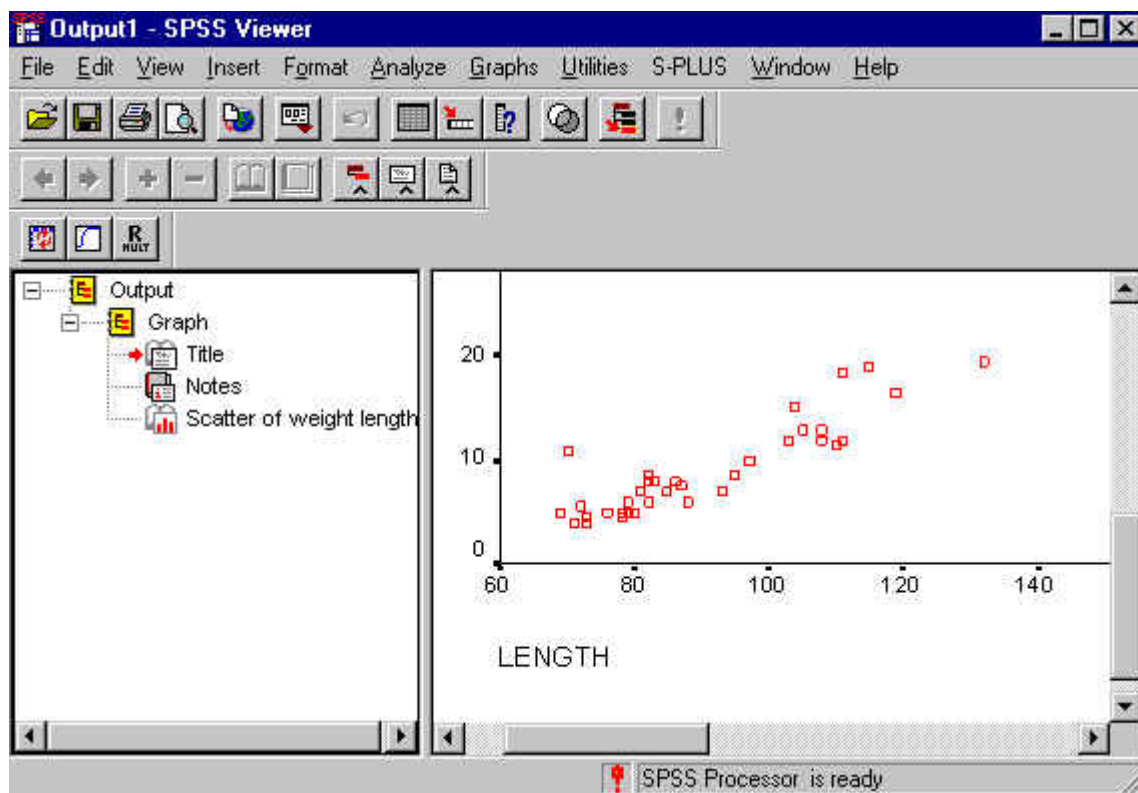
Examining, Printing, and Saving Results

The results of the application of any of the statistical procedures or graphing procedures from the **Analyze** or **Graphs Menus** appear in the **SPSS Output Navigator Window**. The **Output Navigator Window** is divided into two parts.

The left side of this window is an *Output Tree* which guides you through the various parts of your results. It may include a log of all **SPSS** commands that were executed in the *SPSS Log*. The results of each procedure executed and subsets of those results are

also shown in the output tree as seen in the example below for the command **Graph - Scatter**. One may access various parts of the output by double clicking on the icons in this tree, and one may suppress display of parts of the output by double clicking on an icon. An open book icon means the output is displayed and a closed book means it is not. A scroll bar is provided to allow you to move up and down the *Output Tree*.

The right side of the **Output Navigator Window** shows you the actual results of the data analysis or graphing procedure you have executed. You may use the scroll bar at the right side of this window to move up and down in the output. In addition, you can edit items in the output window such as captions on graphs and axis labels, or enter additional text or notes. These kinds of procedures are accomplished by clicking directly on the objects in the output window, by using commands in the **Edit Menu**, or by clicking on particular icons in the toolbar.



After generating output you may save it to a disk file or send it to a printer. These tasks are accomplished from the **Output Navigator Window** using commands on the **File Menu**.

A Point to Remember: All of this will become clearer as you use trial and error to get accustomed to SPSS and its user interface.

GETTING DOWN TO BUSINESS

As a brief introduction to using **SPSS**, complete the following exercise during the lab period. But, feel free to play with **SPSS** to get accustomed to its capabilities and its user interface. Feel free to use your own data to continue your exploration of SPSS, either during the lab period or in the public access labs at other times.

LAB EXERCISE

The following lab exercise is intended to allow you to become familiar with the **SPSS Menu/Help** system and to create and analyze some real data. You will learn how the commands function using the on-line **Menu/Help** system.

A few words about the data set:

The data for this week's lab come straight from some male-male competition work that a former student collected for a species of crab spider, *Misumna formosipes* (Family: Thomisidae). It is in the file [malrank](#), of which you will need to obtain a copy. The file contains data on specific spider males including: their code number (Column 1), carapace width in mm (Column 2), tibia length mm (Column 3), body length mm (Column 4), and weight in mg (Column 5). Another important feature of the data file is that weights for some of the spiders were not recorded. In order for **SPSS** to read the data format, a value is required. In these cases I often use 999.000 to indicate missing values. Keep this in mind for any analyses that use the variable "weight."

Again, the first thing to do in a program is to tell **SPSS** where to find the data, what form it is in, and names for the variables. In this case you have several options. You can have SPSS **Read Text Data**, but you will need to indicate the file name, its location, and how the data are delimited.

Next, obtain the mean and standard deviation for all the variables (except spider code obviously). There are several ways to do this, but the easiest commands are accessed through the **Analyze Menu - Descriptive Statistics (Descriptives or Explore)** and **Custom Tables**. You will need to define the missing values for the variable "weight", otherwise SPSS will treat the values of 999.00 as valid entries. You will need to exclude these values by using the **Variable View** tab in the **Data Editor Window** and defining the missing value to be "999.000".

Now make a plot (a scatter plot) of weight versus carapace width. Make sure to make weight the dependent variable (y-axis). This is found within the menu system under **Graph Menu**. Be sure to label the axes (vertical and horizontal), give the graph a title, and to remind **SPSS** that you have some missing data.

Next make a histogram of tibia lengths and superimpose a normal distribution curve over the histogram. Again look to the **Analyze Menu** for the **Explore** command, or the **Graph Menu** for the **Histogram** command.

Further Instructions for the Lab Exercise

The data file for lab 1 [malrank](#) is linked to the webpage. Simply click on the link and then save the file on your computer or mass storage unit (disk or USB drive). This makes the file local to your computer and you can then use the **Read Text Data** command and Wizard in **SPSS** to read the data into **SPSS**.

After completing the exercise, you can save your results by using the **Select All** command from the **Edit** Menu, and then the **Copy Objects** command from the **Edit** menu to copy your results to the clipboard. In the mean time, open a Word file and then **Paste** the results into the Word file. Name and save your Word file, and copy it to your mass storage unit, or e-mail it to yourself (or both). Now you can work on your lab write-up at home even if you don't have access to **SPSS**. If you have **SPSS** on your own computer, you can simply use the **Save As** command from the **File** Menu to save both your results from the **Output Viewer** window and the **SPSS** version of the data file from the **Data Editor** Window. Make sure that your computer at home or the office has at least as recent a version of **SPSS** as we use in lab, or the files may be incompatible. It is possible to save **SPSS** files in a format compatible with earlier versions of **SPSS**.

Lab Write-Ups

Lab write-ups should be submitted as Word or RTF files. Name the file "yourlastname-lab1" for example.

Lab write-ups should include answers to any questions posed by the lab exercise and report and interpret the results of any analysis performed. Lab write-ups that consist only of **SPSS** output without a narrative providing answer, results and interpretation will be treated as if nothing was submitted.

Not all of the tables and graphs you produce in a work session need to be turned in for any particular lab write-up. Only the tables and graphs you deem essential to convince me that you have mastered using **SPSS** to complete the exercise should be turned in.

No lab write-up need be turned in for lab 1. However, I have included below an example of what a write-up for lab 1 might look like. Take this as a lesson for what is expected of you for labs 2-10, all of which require a written lab write-up to be submitted.

Lab 1 Write-Up

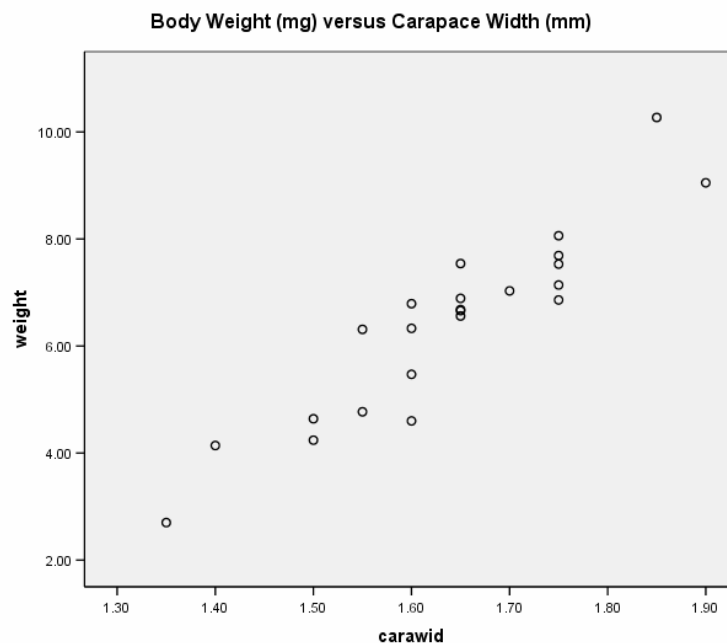
Using the data provided on body size and mass for male crab spiders (*Misumna formosipes*), I calculated the means and standard deviations of each metric variable using the “Descriptives” procedure in SPSS (see table below)

Of the linear measurements body length was most variable (sd = 0.307), but also had the highest mean value. Total body mass (weight) averaged 6.4 mgs, but was also quite variable among individual spiders. Only 23 of the 38 spiders had data on body mass.

Descriptive Statistics

	N	Mean	Std. Deviation
carawid	38	1.6553	.14695
tibalen	38	1.9645	.24243
bodylen	38	3.3421	.30724
weight	23	6.4326	1.70596
Valid N (listwise)	23		

I then proceeded to produce a scatter plot of body weight versus carapace width (see graph below). The graph suggests that weight increases linearly with body mass for these male spiders.



A histogram of tibia lengths suggests that tibia length in mm is approximately normally distributed, although there is some evidence that the distribution is slightly skewed to the right (the distribution is not symmetrical).

