Your Name: ______________________

Section Letter: ______________________

TA names: ______________________

INTRODUCTION TO STATISTICAL REASONING 36-201
Computer Lab Exercise # 1
Introduction to Minitab and Describing Data

Objectives:

1. To learn how to use the Macintosh and Minitab.
2. To read data into Minitab.
3. To produce and interpret histograms and boxplots

I. Introduction to the Macintosh and to Accessing Minitab

Terminology

The Mouse - is used to move objects on the screen, select text, and issue commands by selecting options from menus.

Arrow Pointer - the small arrow that appears on the screen. You move the arrow pointer by moving the mouse.

Clicking - is pressing the mouse button, then releasing it. Do not hold it down, just press and release.

Double-Clicking - is clicking twice quickly. Be sure the clicks are in rapid succession so they are interpreted as a single action rather than two separate actions. Hold the mouse stationary while double-clicking.

Dragging - means to press and hold the mouse button down while moving the arrow pointer across the screen.

Dialog Box - a box which contains several questions or options to which you can reply. Follow the directions. After choosing your option, *click* **OK**.

Icons - are pictures that represent programs, applications, documents, and folders. To open a folder or application, move the arrow pointer onto the folder or application icon and **double-click**.

The Window - is the rectangular box that appears when you select a program. All of your work will be done within a window.

Title Bar - includes the words and horizontal lines that appear along the top of the window.

Menu - a list of options that correspond to a given topic. To choose an option from a menu, move the arrow pointer to the desired menu, press the mouse button and hold it down; a list of words will then appear. Keep holding the mouse button down while you **drag** the arrow pointer along the list, until the choice you want to make is highlighted or has a black bar across the letters. When you release the mouse button, the highlighted option or command will be chosen and the corresponding action will take place.

Accessing Minitab

You will follow these steps each time you begin a lab.

1. *Double-click* the HD cluster icon.
2. *Double-click* the Applications icon.
3. *Double-click* the Minitab 10.5 folder.
4. *Double-click* the Minitab 10.5 icon.

II. Reading Data: Keyboard Entry

We will use data on REM latency, which is the time from sleep onset to the first REM (rapid eye movement) period. We are interested in describing the distribution of REM latency.

You will need to type in each of the data points. This is known as *keyboard entry.* First you will enter the REM latency data for the Normal subjects. You will then create a histogram. You will also explore different Minitab commands. The values of REM latency in minutes are

47, 90, 95, 51, 87, 60, 69, 51, 19, 54, 66, 138, 64, 66, 60, 84, 137, 60, 69, 63

1. In the Minitab window called "Untitled Worksheet", in the first row of the spreadsheet (the row pointed to by the arrow), click the first cell under C1. Type the word NREML, which stands for "normal REM latency". Now *click* on the arrow in the upper-left corner of the window so that it points down. Press **Return**.

2. If the the first cell in row which is numbered 1 is not highlighted, select it by clicking that cell. Type the first observation value, 47, and press **Return**. Type the next observation, 90, and press **Return**. Continue doing this until you have entered all 20 observations.

3. Notice that after awhile you can no longer see the first several observations that you entered. On the right hand side of your window in the *scrollbar*, find the arrow pointing up. Move the arrow pointer to the arrow pointing up and hold down the mouse button until you have moved to the top of the window.

4. You may resize the window so that all 20 observations will show on the screen simultaneously. You can do this by moving the arrow pointer to the square in the <u>bottom</u> right hand corner of the window (this square is called the *Size box.*) *Drag* the size box in a downward direction until all of the observations are visible.

5. You should always double check to be sure that you have entered all of the observations correctly. If you have made a typing error, *click* the cell containing the incorrect information, use the *delete* key to erase it, and retype the correct value.

III. Producing Graphical Displays

We next want to investigate the distribution of NREML using a *histogram* and a *boxplot*. (We may not have talked about boxplots yet but we will shortly. You may have read about them in the text.)

A. Histograms

1. *To produce a histogram*. From the **Graph** menu, choose **Histogram**. Under "Graph variables" type NREML in the cell in Row 1, under Column X. *Click* **OK**. The histogram will appear in the center of your screen. To make the histogram a little larger, move the arrow pointer to the bottom right corner of the graph window. *Drag* the size box downward to make the window larger. Try it.

Using your histogram you will now examine the features of the distribution of NREML. With your partner, answer the following questions.

♣ **Question #1**
i) What is the width of each class interval (i.e., each rectangle) of the histogram? ______
ii) Which interval contains the most observations? ______
iii) Approximately how many observations are in this interval? ______
iv) What are the actual NREML times that are contained in this interval? __________

Note: The interval that contains the greatest number of observations is called the **mode**.

Question #2 Look at the last interval in the histogram. This interval corresponds to which NREML times?

Question #3 What is the shape of this histogram? (Is this histogram approximately symmetric, positively skewed (skewed right), or negatively skewed (skewed left)?) What other features are noteworthy in this histogram?

You will now change the scale of the histogram, i.e., the width of the intervals, to see how this might change our impression of the features of the distribution of NREML. From the **Edit** menu, choose **Edit Last Command Dialog**. This is a very useful menu option that allows you to return to the last dialog box that you used. *Click* **Options**. Under "Definition of Intervals", *click* the button next to "Mid/cutpoint positions". Type *0:150/15* in the box next to "Mid/cutpoint positions". You are now telling Minitab to make a histogram which starts at 0 minutes and goes up to 150 minutes in intervals of width 15 minutes. *Click* **OK** and then *Click* **OK** again. You should now see a new histogram window.
Question #4 Look at the last interval in the histogram. What are the actual NREML times that are contained in this interval?

Question #5
i) Which interval contains the most observations? ______
ii) Approximately how many observations are in this interval? ______
iii) What are the actual NREML times that are contained in this interval? ______

Question #6 Approximately what is the shape of this histogram?

Question #7 What other features are present in this histogram?

To hide the histogram so that you can look at other plots on the screen, in the histogram window *click* the *Close box* in the upper left hand corner. This will hide the plot. You may need to do this two times. Now look at the **Window** menu. Both plots you created are listed at the bottom of the **Window** menu. Later we will retrieve the histograms to look at them again.

B. Boxplots

Another graphical display which is very useful for looking at the features of a distribution is the boxplot. We will now go through the steps for drawing a boxplot.

1. *To produce a boxplot.* From the **Graph** menu, choose **Boxplot**. Under "Graph variables: Y vs. X" type NREML in the first cell in Row 1, Column Y. *Click* **OK**.

The values of the REM latency times are displayed on the vertical axis. This is different from the histogram where the values of the distribution were on the horizontal axis. We will talk more about how to interpret boxplots in class. For now, notice that there is a rectangular box with a horizontal line inside of it, a vertical line going up from the box and one extending down from the box. Also note there is an asterisk above the box (actually there are two asterisks almost on top of each other).

To see these features better, find the square in the upper right hand corner of the boxplot window (this is the *Zoom box*). *Click* the *Zoom box.* The boxplot window fills up your screen. *Click* the Zoom box again and the boxplot window returns to its original size. Try this. Now *click* again so that the boxplot fills your screen.

Question #8 Find the horizontal line inside the rectangular box. This line represents the **median** of the distribution of NREML. The median is the value (number) that divides the distribution in half, i.e., half the observations are above the median and half are below the median. Looking at the vertical scale, approximately what is the value of the median for the distribution of NREML?

Question #9 Looking at the vertical scale, approximately what are the values of the asterisks sitting above the rectangular box in the boxplot? Unusual observations, sometimes called outliers, are represented in a boxplot by a $*$.

♣ **Question #10** What feature of the histogram of these data that you looked at earlier corresponds to the identification of the outliers in the boxplot in question #9?

Exercise

We now would like to compare the REM latency times for the normal subjects to the times for the depressed patients. The data points in minutes are

2, 5, 6, 10, 15, 17, 19, 21, 25, 39, 47, 48, 48, 48, 52, 65, 81, 81, 87

You must now enter the REM latency times for the depressed patients. To do this, you must create a new variable.

1. First to clean up the screen, you should close all of the open windows, except for the Worksheet window and the Session window . Move the arrow pointer to the *Close box* in the upper left hand corner of each other window and *click* once.
2. In the "Data" window, under C2 *click* the second cell in the top row of the grid. Type the word DREML under the C2. Press **Return**.
3. If the the second cell in row number 1 is not highlighted, select it by clicking that cell.
4. Now, follow steps 2 through 5 from Section II above with the DREML column to type in the new data. ♣

Part A

Produce a histogram and boxplot for the distribution of REM latency times for the depressed subjects. Follow the steps from the previous sections substituting DREML for NREML. You will need to change the scale of the histogram that you set for looking at the histogram of NREML. *Click* **Options** and next to "Mid/cutpoints positions" delete "0:150/15". Then *click* **OK**.

Study the features of this distribution. Which class interval in the histogram is the mode? Describe the shape of the histogram. Using the boxplot, what is the median of the distribution? Are there any outliers?

Part B (If you have time; otherwise go to the next section "To Finish ...")

We now would like to compare the distribution of REM latency times in normal subjects to the distribution of REM latency times in depressed patients. You will do this by creating the two boxplots next to each other on the same scale. First, you must create two new columns of data. The first will contain both sets of REM data in a single column. The second will contain subscripts that correspond to the two REM groups.

1. Close all plot windows, but keep the Worksheet and Session windows open.
2. From the **Manip** menu, choose **Stack**. *Click* the first box under "Stack the following blocks:". Type NREML. Press **Tab**. The insertion point (the blinking |) should be in the second box. Type *DREML*.

3. *Click* the box under "Store results in:" and type REML.

4. *Click* the check box to the left of "Store subscripts in". *Click* the box to the right of "Store subscripts in" and type *Group.*

5. *Click* **OK**.

Note in the "Data" window that the data in C3 is called REML and the data in C4 is called GROUP. It is important that you name any new variables (columns) you create while using Minitab. It makes working with the data much easier. You can now use these variable names when making plots and using other Minitab functions.

You should now have two new variables, REML and GROUP. The GROUP variable can take one of two values: 1 if the subject is in the normal patient group, or 2 if the subject is in the depressed patient group. Now you can plot both boxplots next to each other on the same scale.

1. From the **Graph** menu, choose **Boxplot**. Under "Graph variables: Y vs. X" *click* the cell in Row 1, Column Y, and type *REML* but *do not press* **Return.** .

2. Instead, *click* the cell in Row 1 under Column X and then type *Group.*

3. *Click* **OK**.

Note that the NREML group data corresponds to the boxplot with a '1' under it, the DREML group data corresponds to the boxplot with a '2' under it. You should always add this information to the boxplot, so that it is clear which group is which. It is also helpful to add a title to the boxplot so that you know what information it contains.

1. From the **Edit** menu, choose **Edit Last Command Dialog**. You should now see the "Boxplot" dialog box.

2. Move the arrow pointer to the downward pointing arrow next to "Annotation". Press the mouse button and hold. You should see a pop-up menu appear. Continue holding down the mouse button and choose **Title**.

3. In the window that appears called "Title" in the first row type *REM Latency for Depressed and Normal Controls.* Click **OK**.

4. Using the pop-menu next to Annotations again, this time choose **Footnote**.

5. In the window that appears called "Footnote" in the first row type *1 = Normal.*

6. To get to the second row press the **Down** arrow key on the keyboard. In the second row type *2 = Depressed. Click* **OK**.

7. *Click* **OK** once more.

Question #11 Comment on the differences between the medians for each distribution. Which group, as a whole, has shorter REM latency times?

If you have time left, you and your partner should play around with the different features of Minitab.

TO FINISH YOUR LAB SESSION

At the end of **every** lab session you must quit from Minitab and clean up your window. Use the following steps.

1. From the **File** menu, choose **Quit**.
2. A dialog box will appear asking you whether you wish to save your work. For this session *click* **Don't Save**.It will then ask you if you want to save changes to the Session window. *Click* **No**.
3. **Be sure to log yourself out of the machine before you leave. From the Key menu choose Logout.**

Be sure to turn in your cover sheet to your lab TA.