36-220 Lab #1
Introduction to MINITAB and Describing Data

Week of September 5, 2005

Please write your name below, tear off this front page and give it to a teaching assistant as you leave the lab. It will be a record of your participation in the lab. Please remember to include your section (A, B or C). Keep the rest of your lab write-up as a reference for doing homework and studying for exams.

Name:

Section:

- The symbol ♣ at the beginning of a question means that, after you answer that question, you should raise your hand and have either the TA or lab assistant review your answer. Once they have reviewed your work they will place a check in the appropriate space in the table below. The purpose of this check is to be sure you have answered the question correctly.
- Try to complete as much of the lab exercise as possible. We understand that students work at different paces and have tried to structure the exercise so that it can be completed in the allotted time. If you work systematically through the handout and still don't complete every question don't worry. The important thing is that you understand what you are doing. Nonetheless, you are encouraged to complete the lab on your own.

Check-Problem ♣	Instructor's Initials
Question 6	
Question 8	
Question 11	

36-220 Lab #1
Introduction to MINITAB and Describing Data

This lab has three goals: to help you start learning to use MINITAB; to learn how to read data into MINITAB; and to learn to make and read graphical displays and numerical summaries.

1 Accessing Minitab

MINITAB consists of a collection of commands that permit you to analyze data fairly easily by use of the computer. To use MINITAB to analyze data, you must first learn how to get into MINITAB so that you can issue commands in it.

To get into MINITAB you do the following:

1. Log onto the machine with the following information:
   Username: your andrew login
   Password: your andrew password
2. Click the **Start** button in the bottom left corner of the screen.
3. Highlight **All Programs** by placing the arrow on it.
4. Highlight **Math & Stats** by placing the arrow on it.
5. Highlight and then click on **Minitab 14**.

We can now begin to use MINITAB. In the window that appears, you will see two smaller windows– the upper **Session** window and the **Worksheet** window below it. Note that each of the two windows can be moved and resized for better viewing.

2 Getting a File off the Web

For today's lab, and many later labs, we will be using data available off the course website.

Open a Netscape or Internet Explorer browser session. Go to the course webiste, `http://www.stat.cmu.edu/~cshalizi/36-220`, and go to the bottom of the page for "Data Sets". Click on the data set "CRASH.DAT." Choose **Save this file to disk**, then click on Desktop along the left of the window, and finally press **Save**.

3 Importing Data

To import the data set "CRASH.DAT" into MINITAB,

1. Go to the **File** menu and select **Other Files** > **Import Special Text**.
2. There are 6 variables in this data set. In the window that says "Store data in column(s):", type "C1-C6". Click **OK**.
3. In the window that appears, choose **Desktop** under the **Look in:** menu and click on CRASH.DAT. Click **OPEN**.
4. Data will appear in the worksheet.

In a study designed to test the safety of various types of vehicles, vehicles were crashed into a wall at 35 mph with a crash-test dummy strapped in the driver's seat. We are going to analyze the data.
C1 - HEAD - measure of head injuries
C2 - CHEST - chest deceleration, which is a measure of chest injury
C3 - AIRBAG - 1 indicates an airbag, 0 indicates only seatbelts were used
C4 - DOORS - number of doors (equals 6 for vans and sport utility vehicles)
C5 - YEAR - year of vehicle
C6 - WEIGHT - weight of vehicle

4 Naming Columns

It is very useful when you are analyzing data to name each column with a short informative name. For example, to name the column C1 *HEAD,* click on the cell immediately below "C1", in the row *above* row number 1 and type *HEAD.* Similarly, label the remaining five variables in the appropriate columns.

5 Descriptive Statistics

Once we have data in MINITAB, it's often useful to have MINITAB tell us some key features of the data. One way to do this is through what MINITAB calls "Descriptive Statistics." In order to calculate the descriptive statistics, access the following through the pulldown menus at the top of the screen:

Stat → Basic Statistics → Display Descriptive Statistics

A dialog box will appear. Double click on "C2 Chest" in the left panel, and it will appear in the right panel under "Variables." Then, click the "OK" button. You should see information about the data appear in the session window.

Descriptive statistics gives several pieces of information about the data. At this time, the following are of interest: N (THE NUMBER OF DATA VALUES), MEAN, MEDIAN, STDEV, MINIMUM, MAXIMUM, Q1, and Q3.

6 Displaying Data

6.1 Stem-and-Leaf Plots

1. Go to the **Graph** menu and select **Stem-and-Leaf...**. Click on *CHEST* and then click on **Select**. *CHEST* is a measure of chest injuries sustained by the crash-test dummy. Larger values indicate more severe injuries. You should see the *CHEST* variable appear under "variables:" on the right side of the box, indicating it has been selected for the graph. Click on **OK**.

2. A stem-and-leaf plot of the data will appear in the Session Window. You may have to scroll up in the window or enlarge it to see the entire plot.

   Question #1: Describe the distribution of *CHEST* in terms of peaks, skewness and outliers.

   Question#2: What are the most common values of *CHEST*? (Give an interval)

3. The numbers to the left of the stems give what is called the "Depth". For example, look at the stem for values in the interval 45-49. The depth value for this stem is 74. This means that there are 74 observations with values of *CHEST* that are less than or equal to 49. The next stem for values in the interval 50-54 has the entry (31). The parentheses indicate that this stem contains the median, and that there are 31 observations on this stem. Finally, look at the stem for values in the interval 70-74. The depth value is 6. This means that there are 6 observations with values of *CHEST* that are greater than or equal to 70.

   Question #3: Using the depth information, determine how many observations have a value of *CHEST* that is $\geq$ 60.

6.2 Histograms

1. Select **Histogram...** from the **Graph** menu. A window will appear with four options; choose the **Simple** type of histogram. In the next window, select the *CHEST* variable by double-clicking on it (this is a short cut for highlighting the variable and clicking on **Select**). To ensure that this is a "density" histogram, click on the button *Scale* and then choose the *Y-Scale Type* tab. Be sure the button *Density* is selected from the list. Click on **OK** to exit this window, and then on **OK** to create the graph.

   Question #4: In this histogram of the *CHEST* variable, the program automatically selected the number of intervals. How many intervals are there?

2. Experiment with changing the number of intervals. To do this, double-click on any bar of the histogram. Choose the **Binning** tab along the top, press the button next to **Number of intervals**, and set it to "10." Click **OK** again to close the window.

   Question #5: How has the histogram changed?

3. Change the number of intervals again. This time use 25 intervals. Compare the 3 histograms.

   ♣ Question #6: Which of the 3 histograms would you use in a presentation or a report for your supervisor? Why?

   Whenever you construct a stem-and-leaf plot or a histogram, you have to decide how many stems or how many intervals to use. Your decision will greatly affect how informative the plot will be. In some cases, a bad decision can obscure important characteristics of the data.

4. Select **Window > Close All Graphs** to get rid of the histograms. Don't save any of them.

6.3 Boxplots

1. Suppose we wanted to see if airbags prevented injuries. Since *AIRBAG* is a **Categorical** or **Qualitative** variable and *CHEST* is a **Quantitative** variable, a good form of analysis would be a side-by-side boxplot. To

make a boxplot, select **Boxplot** from the **Graph** menu. Choose the type **One Y, With Groups**. Select *CHEST* as the **Graph variable** and **AIRBAG** as the **Categorical variable for grouping**. Click on **OK**.

Question #7: Compare the distribution of chest deceleration for airbag (*AIRBAG*= 1) and non-airbag (*AIRBAG*= 0) cars. Do airbags appear to prevent injury?

How do the spreads of the two distributions compare?

Is either distribution skewed? If so, describe the skewness.

If there is an outlier, this will appear in the boxplot as a "*" above or below the boxplot. Are there any outliers?

2. Now let's compare the chest injuries of vehicles with different numbers of doors. Create side-by-side boxplots, as above, using these two variables.

   ♣ **Question #8: Does there appear to be a relationship between the number of doors and chest deceleration? What about the plot suggests a relationship?**

   Question #9: Which type of vehicle (in terms of number of doors) tends to have the least severe injuries?

   Question #10: Why do you think that chest deceleration might be related to number of doors on the vehicle?

3. To test one reasonable theory, make a boxplot of *WEIGHT* by *DOORS*.

♣ **Question #11: What do the results of this boxplot suggest about the results of the previous boxplot comparing chest deceleration by number of doors?**

7 Saving Your Output

1. When doing homework assignments for this class, you will often be using MINITAB. In this section, you will learn the different ways to save your output so that you can turn it in with your homework assignment.

2. Graphs such as histograms, boxplots and scatterplots (to be discussed later in the semester) all appear as high-resolution graphs in a separate window. It is very simple to make a printout of these graphs– just make sure the window with the graph is the active window, then select **File > Print graph**. Whenever you turn in a graph with your homework, make sure that the graph, the x-axis and the y-axis are labeled, and that it is clear which homework exercise the graph was made for. Your graphs can also be saved as BitMap files (**File > Save Graph As**, then select **Windows BMP**) and included in your favorite Word processor.

3. Make a stem-and-leaf plot of the variable *HEAD*, which is a measure of head injuries.

4. Stem-and-leaf plots appear in the session window. To save output from the session window, you will need to make a log file. Select **File > Save Session Window As**. Click on **Save as type**. Select the *Rich Text Format*. Click on **File Name**. Enter a name for your log file. At this point, if you really wanted to save the file, you would click on **Save**. Since you don't actually need to save this particular file, click on **Cancel**. Note that *Rich Text Format* files can be opened in Microsoft Word.

Note: It is never acceptable to turn in MINITAB output alone. It should always be well labeled and come with a written explanation that interprets the output.

8 Saving Your Work

During lab, you will not need to save your work. However, there may be times while doing your homework that you will find it useful to save your file. If you wanted to save your file, you would do so using the pulldown menus as follows

File → Save Current Worksheet As

In the **Save in** drop down dialog box, select "Desktop." Type "crash" in the **File name** text box. Choose **Save**.

9 Closing Worksheets

At times, you'll basically want a clean slate to work from in MINITAB. You can close your current worksheet in MINITAB by clicking on the × button in the upper right hand corner of the *Worksheet*. This will eliminate the current worksheet; a new one will automatically appear.

10 More Getting Data off the Web

It's also possible to cut and paste data off the web into MINITAB. Throughout the semester, there may be times where you'll need to get data for your homework.

Open up a web browser and go to the web page `http://lib.stat.cmu.edu/DASL/Datafiles/Ageandheight.html`.

Here, you'll find data on the average age (in months) and average height (in centimeters) of a group of children in an Egyptian village. The measurements were taken repeatedly over several months.

Highlight the data, starting with the word "age" downwards and to the right, until the end of the dataset (the number 83.5 in the second column). Then "copy" the data (**Edit → Copy**). Return to MINITAB. Close the current worksheet. A new, clean worksheet should appear.

Now, "paste" the data into MINITAB: Right-Click on the cell immediately below "C1" (*above* row 1) and select "paste cells." Alternatively, after selecting that cell, you could use the pull down menus and select **Edit → Paste Cells**.

Before logging out of your computer, remove the file CRASH.DAT by placing it in the Recycle Bin.