BASICS

1) To open MINITAB, you just need to click on the appropriate icon.
2) Once opened, you will have two environments in the software. The top part is called the “session window”, the lower part is a spreadsheet (the grid) with columns (C1, C2, ...) and rows (1, 2, ...). This part is called the “Worksheet”. This is the place in which we enter the data.
3) Essentially that’s all we need to know to start with MINITAB!

DATA ENTRY

1) Place your cursor on the cell 1-C1 which is the cell in the first row and the first column of the worksheet. Notice that the frame of the working cell is highlighted. Type the first number in your data-set and hit ”Enter” key in your keyboard. This will take you to cell 2-C1 or the second row in the first column. In MINITAB, each column represents a set of data (as oppose to each row). When the data is large you might want to enter it row-wise (however, each column is still a different data-set.) To do this, click on the upper-left-corner square in your worksheet. The vertical arrow becomes horizontal and the data entry can be done row-by-row.

2) To edit a cell simply click on it. If you want to add a new data-point in the middle of a column, highlight the cell positioned after the desired row and right click on your mouse. By clicking on the ”insert rows” option, you can add a new data point. You can delete a cell in the same fashion except you use ”delete cells” command in the last stage. Alternatively, all the editing options can be performed using ”edit” from the menu bar. A fine way of getting comfortable with the program is to create data and try editing it.

3) The best way to keep track of data-sets (or columns) is to name them. To name each column, you can click on the empty cell within the same column but above the first row. Later, when you apply various statistical methods, the software will recognize that column with the name you chose for it.

4) To obtain MINITAB data files from other sources, choose FILE from the menu bar, then click on ”OpenWorkSheet ...”. Finally, notice that you can switch between the two sessions by choosing ”Window” option from the menu bar. At the very bottom of the ”Window” menu, you can view the open worksheets and the window session.

5) We will discuss the importance and applications of window session later, after we introduce the statistical techniques. Briefly, the window session in the place at which the results of the statistical analysis will appear.

Creating Graphs

Histograms

From the ”Graph” menu, choose ”Histogram ...”. The names of the columns that you have already created will appear on the box located at the left side of the ”Histogram” application. Simply click on the column of interest and hit ”select”. The selected column will appear under
the "X" mark. By hitting "OK", you will create your histogram. There are other options in this menu that we will learn about as we make progress into the quarter.

**Pie Charts**

The same procedure can be followed. From the "Graph" menu choose "Pie Chart ...". Click on the empty box: "Chart Data in", then the list of all columns will appear. Click on the column of interest and hit "select". Next, hit "OK" to obtain the pie chart.

**Boxplots**

Boxplots are extremely useful in assessing the shape and variability of a sample statistic. To create a boxplot, you need to go to Graph, click on Boxplot, select the variable (column) of interest and put it under Y box. Hit "OK". We discuss the various points and important features of the boxplots in class. Alternatively, you might be interested in drawing side-by-side boxplots. To do this and in the Boxplot option, place the variables (columns) of interest under Y (first variable in the first row, second variable in the second row, etc.), and in the Frame option click on Multiple Graphs .... Choose Overlay graphs on the same page. Hit "OK" twice. You should be able to observe the side-by-side box plots.

**Simple Calculations**

We can apply various simple operations on a given column. For example, let’s assume that we want to multiply each component of the first column (C1) by 6 and place the results in the fifth column (C5). Choose "Calc" from the menu bar and choose "Calculator". In the empty box in front of "Store result variable", type: C5. Click on the empty area under "Expression". There, type 6*C1. C1 can either be manually typed or can be selected using "Select" option. Hit "OK". You should have the result in the fifth-column.

Also, in "Calculator" and under the "Functions" menu, there is a whole list of functions you can use. For example, if you wanted to get the square root of column C1 and place it in C5, all you need to do is to type "Square root (C1)" in the expression box. But "Square root" is a provided function that is, you only need to double click on it from the "Functions" menu.
SA T SCORES AGAIN ...

Problem 1. In your MINITAB software, and using File, go to Open Worksheet ... and open the file Ga.MTW. This is the datafile we briefly introduced in the previous lab. It turns out that the three columns are the SAT verbal, Math, and also the GPAs of 100 students in Northeastern university. Let’s explore the data more.

(a) draw a histogram for the verbal scores.

(b) draw a histogram for the Math scores.

(c) draw a histogram for GPAs.

(d) draw a boxplot for the verbal scores.

(e) draw a boxplot for the math scores.

(f) draw a boxplot for GPAs.

(g) Go to Stat then Basic Statistics then click on Display Descriptive Statistics choose all three columns and hit "OK". Summarize your findings. Most importantly, compare your findings with the plots you made in the previous sections.

(h) Go to Graph menu and choose plot. First, select the "verbal" variable, and place it under Y. Next, select the "GPA" variable and make sure it is placed under X. Now Hit "OK". You just created a scatter plot that suppose to show the relationship between the Verbal scores and the GPAs of those 100 students.

(i) Repeat the above for plotting "math" versus "GPA" followed by "math" versus "verbal”.

(j) Interpret the three graphs. That is, make a link between the overall performance of each student in the context of those three measurements.

(k) Go back to Stat → Basic Statistics menu. Click on Correlation. In the Correlation box and under the Variables: select the variables "verbal" and "GPA". Record the correlation between the two variables. Repeat this procedure and obtain the correlation between the variables "math" and "GPA" followed by "Math" and "verbal". Compare your findings with the scatter plots you made previously.

(l) Go to Stat → Basic Statistics → Normality Test .... On the Variable box select "verbal. Also, click on Kolmogorov-Smirnov test. Hit "OK". You should have a "Normal Probability Plot" associated with "Verbal”. By the way, this is the same "Q-Q plot" we talked about in the previous lecture. Comment on the normality of this random variable. Also, keep track of $D^+$, $D^-$, $D$, and KS p-value. I will talk about these measurements towards the end of this lab.

(m) Repeat the same process for "Math”.

(n) Repeat the same process for "GPA”.

(o) Go to Stat → Basic Statistics and choose 1-sample t .... Under Variables, select "verbal” and in the empty box in front of Test mean: type 600. Hit OK. Record your results. Comment on the "p-value” and the "confidence interval”.


(p) Repeat the same process for "verbal". This time, before hitting "OK" and in your 1-sample t menu, click on Options. Change the confidence level from 95 to 99 and hit "OK" twice. Record your findings. Go back and repeat the same testing procedure now at 90% level. Compare this two-sided test with respect to those three different confidence levels.

(q) We are not done yet! In the 1-sample t menu, and again for “verbal”, click on Options. This time let the confidence level to remain at 95% but for the Alternative, choose less than. Record your findings. Repeat the same process at 95% level but change the alternative to greater than. Compare the results.

(r) We are not done yet!! Now, we are going to use MINITAB to compare two populations. Namely, "Verbal" versus "Math". Choose Stat → Basic Statistics → two-sample t. Click on Samples in different columns. Place "Math" in the first box. Place "Verbal" in the second box. Click on Assume equal variances. Hit "OK". Summarize your findings. Repeat the same process but leave the Assume equal variances box empty. Any changes? Why?

(s) In the two-sample t, leave the Assume equal variances box empty. Now click on Options. In the box called Test mean: type 50. Hit "OK" twice. Compare your findings with the previous section! How would you rationalize such strikingly different results?

Problem 2.

(a) Close the previous worksheet and open a new one.

(b) Go to Calc → Random Data → Normal .... Generate 10 rows of data and Store in column: C1. Let mean and Standard deviation remain at 1 and 0 respectively. Hit "OK". Now, create a histogram of those 10 randomly generated numbers. Do not close the graph.

(c) Repeat the above but this time generate 100 rows of Normal(0,1) data (overwrite the newly generated data on C1). Draw a histogram for those 100 randomly generated numbers.

(d) Repeat the same procedure and generate 1000 rows of normal data. Histogram those numbers.

(e) Repeat the same for 10000 rows of data. What did we just learn from all of this? By the way, what you just did is called simulation.

(f) Go to Stat → Basic Statistics → Display Descriptive Statistics. Summarize your findings.

Now let’s try some COOL things:

(g) Go to Calc → Sort and sort column C1 and Store sorted column into C2. Also, in the empty box in front of Sort by column select "C1". You should have the sorted version of C1 placed in C2.
(h) Now, go to Manip → Code → Numeric to Numeric. In the box Code data from columns select "C2". In the box Into columns select "C3". In the first row of Original values type -1.96:1.96 and under New: in the same row type 1. In the second row under Original values type 1.96:5 and under New: type 0. In the third row under Original values type -5:-1.96 and under New: type 0. Go to Calc → Column Statistics and select Sum. Hit "OK".

(i) Go to Calc → Probability Distributions → Normal ... → Cumulative Probabilty. Select Input Constant and in the empty box in front of it type: 1.96. Hit "OK". Repeat the same procedure with constant value of -1.96. Compare your findings with the results you obtained in the previous sections.

Problem 3. More on QQ-plots To discover QQ-plots further, consider the following three scenarios:

(a) Open a new worksheet.

(b) Generate 1000 random data using Normal distribution (see previous problem). Put the results in "C1".

(c) Generate 1000 random data using a \( \chi^2 \) distribution with 2 degrees of freedom. To do this, go to: Calc → Random Data → Chi Square .... Store the results in column "C2". Choose 2 for degrees of freedom.

(d) Generate 1000 random data using a \( \beta \)-distribution! To do this, go to: Calc → Random Data → Beta .... For the first shape parameter choose 15, and for the second shape parameter choose 3. Put the results in "C3".

(e) Histogram "C1", "C2", "C3". Compare their Normal Probability Plot!