Math 321
Introduction to MINITAB
Lab #1

Login with your ID and pass word. If you do not have this, you need to go the 7th floor of the Math Science building and get it. You will have to wait at least 45 minutes until you can use your password after getting it.

Click on **Minitab 14.**
Spend a few moments finding your way around the MINITAB windows. Our main interest will be in the DATA window, which serves as the worksheet for entering data, and the MENU bar that allows the selection of commands to perform various tasks. Results are presented in the SESSION window.

Example 1: Entering data
The values below represent the grade point averages of 30 randomly selected U. of C. Students:

       2.0 3.1 1.9 2.5 1.9 2.3 2.6 3.1 2.5 2.1
       2.9 3.0 2.7 2.5 2.4 2.7 2.5 2.4 3.0 3.4
       2.6 2.8 2.5 2.7 2.9 2.7 2.8 2.2 2.7 2.1

i) Enter this data into the first column, c1, of the worksheet in the DATA window.
Enter GPA in the blank cell under the c1. This assigns a name for the "variable" in that column.
Before proceeding, be sure that the highlighted cell is not one of your data values – click anywhere else on the worksheet.

ii) From the MENU BAR select GRAPH > HISTOGRAM... four pictures appear. Since we are just dealing with one data set, click on SIMPLE then click OK.
Click the mouse on c1 GPA in the rectangle on the left of the box: then press SELECT. This enters the name of your variable as one to be graphed. {Or type c1 in the GRAPH VARIABLES cell.}
Click OK.

A Histogram should appear. You can go back and try another type of histogram with a curve superimposed. Try changing the interval width etc…

Go exploring with different graphs…

iii) Find the mean, range, variance and standard deviation for the above data set. You can use your calculator or MINITAB to do this.

STAT>Basic Statistics> Display Descriptive Stats. Click the Variable Box , double click on C1. Click on OK or hit enter

Example 2: Using MINITAB files
Students in an introductory statistics course participated in a simple experiment. Each student recorded his or her height, weight, gender, smoking preference, usual activity level, and resting pulse. Then they each flipped a coin and those whose coin came up heads ran in place for one minute. The entire class then recorded their pulse rates again. The data is saved in a worksheet called PULSE

To retrieve the worksheet: Choose FILE>OPEN WORKSHEET ... a dialog box appears. Scroll through the files listed below it until you find the pulse file (make sure you don't click on Pulse1) and click on it.
Press OK. The results of this experiment appear in the worksheet in the DATA window.

Variable	Description	Variable	Description
C1 'PULSE1'	Resting pulse rate	C5 'SEX'	1=male 2=female
C2 'PULSE2'	Second pulse rate	C6 'HEIGHT'	Height in inches
C3 'RAN'	1= ran in place; 2=did not run in place	C7 'WEIGHT'	Weight in pounds
C4 'SMOKES'	1=smokes regularly 2=does not smoke regularly	C8 'ACTIVITY'	Usual level of physical activity; 1=slight; 2=moderate;3= quite often

i) Compare the resting pulse rates of the students according to whether they smoke:
GRAPH>Histograms
Select PULSE1 as the variable to be plotted {highlight and SELECT} or type in "graph variable box"
Click the "categorical variable box": double click on c4 SMOKES {or click and SELECT}
Click OK and the histograms are given one on top of the other..

ii) Compare the descriptive statistics of the resting pulse rate according to activity level in a similar fashion.

To compare the actual values: STAT>Basic Statistics> Display Descriptive Stats. Click the Variable Box , double click on PULSE1. Click the By Variable box. Double click on Smoke. Click on OK or hit enter

iv) Use MINITAB to help you answer the following question:
1. Compare the resting pulse rate of males and females by discussing the relative position of their central measures; the spread and shape of the respective distributions.
2. What is the range of resting pulse rate for those classified as regular smokers?
3. What is the difference in the mean 'second pulse' rate for those who ran in place and the mean of those who did not?

Input some other files into MINTAB or input your own data set. Try out all the features.

Additional Questions:

1. Visitors to Yellowstone National Park consider an eruption of the Old Faithful geyser to be a major attraction that should not be missed. The given table summarizes a sample of times (in minutes) between eruptions and the relative frequency

Time	Relative Frequency
40-49	0.04
50-59	0.22
60-69	0.115
70-79	0.03
80-89	0.535
90-99	0.055
100-109	

(a) Complete the table by filling in the missing relative frequency. (.005)
(b) If the sample consisted of 200, find the frequency distribution.
(c) Draw the appropriate relative frequency histogram for this data. Describe the distribution of the data. (Make sure there are no gaps)
(d) What percentage of the times between eruptions was at least 80 minutes? (59.5%)
(a) What number of times did visitors have to wait between 60 and under80 minutes between eruptions? (29)

2. The campus branch of the Royal Bank randomly selected a sample of student accounts and checked the corresponding balance (rounded to the nearest dollar amount). The resulting data follows :

458, 623, 449, 721, 441, 498, 609, 557, 587, 435, 567, 760, 505, 876, 535, 684

a) Find the following: mean, and standard deviation (581.6,126.7)
b) Using the Chebyshev's inequality, within what range would you expect to find at least 75% of the account balances? (328.1, 835.0)
c) What percentage of the data should fall within 1 standard deviation of the mean if the data is normally distributed? (~68%) What is the actual percentage from the data? (62.5%) Why aren't they exactly the same?

3. A sample of 100 values has a mean of 41 and a standard deviation of 6. The shape of the distribution is unknown. How many values would you expect to be contained in the interval 20 to 62? (91.84 or between 91 and 92)

4. For a list of 50 sample values, $\sum y = 50$ and $\sum y^2 = 646$. In what interval would you expect to find at least 8/9 of the data if the distribution is not normal? (-9.4628, 11.4628)

5. It is known from experience the 15% of a restaurant's customers order fried chicken on a given night. A researcher collects data on 10 randomly selected nights and finds that 10% of the customers ordered fried chicken. Identify the population, parameter, sample, statistic and the variable in this situation.

6. A researcher wanted to find the average credit hours for students attending a certain university. The credit hours of a random sample of 46 students were recorded as
    11, 15, 13, 12, 7, 13, 9, 12, 15, 8, 16, 13, 11, 10, 12, 8, 12, 9, 9, 14, 6, 7, 14, 12, 8, 14, 12, 10, 11, 17, 10, 17, 17, 9, 18, 11, 12, 13, 13, 18, 19, 16, 14, 15, 6, 15,
    a. Calculate the mean and standard deviation of the sample. (12.2608,3.3693)
    b. Calculate the approximation to s and compare it to the actual s in (a).
    c. Produce a frequency distribution of the data. (Note: You can use MINTAB to sort the data for you.)
    d. Produce a relative frequency histogram of the data with 8 classes.
    e. Describe the shape of the distribution.
    f. If the data was normally distributed,
        (i)   what percentage of the data should lie within 2 standard deviations of the mean? (~95%)
        (ii)   what percentage of the actual data lies within 2 standard deviations of the mean? Is this answer the same or different as the answer in (i)? If different explain why. (97.83%)
        (iii) what percentage of the data should lie below 1 standard deviation of the mean? (~16%)
        (iv)   what percentage of the actual data lies below 1 standard deviation of the mean? Is this answer the same or different as the answer in (iii)? (15.22%)
    g. If the data is not normally distributed,
        (i)   what percentage of the data should lie within 2 standard deviations of the mean? (at least 75%)
        (ii)   what percentage of the data should lie within 1.2 standard deviations of the mean? (at least 30.56%)
        (iii) what percentage of the actual data lies within 1.2 standard deviation of the mean? (73.33%)

7. Show how $s^2 = \dfrac{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}{n-1}$ is derived from $s^2 = \dfrac{\sum(y-\bar{y})^2}{n-1}$.

8. A traffic study conducted at one point on the Trans Canada Highway shows that vehicle speeds (in km/h) have mean of 112.2 and a standard deviation of 6.6. Approximately what percentage of cars are traveling more than 118.8 km/h? **(~16%)** What assumption did you make concerning the distribution of speeds in order to answer this question?

9. Test scores have a mean of 10 with a standard deviation of 7. Why would these values lead you to believe that the test scores are not normally distributed?

10. Do all the questions in Chapter 1.