Name: _____ **MAT 121**
Summer 2018 **Lab 1**

Problem 1: The following are resting heart rates for a group of men and women.

Male	71.3	70.7	68.1	75.1	76.2	74.9	67.9	66.1	72.3	73.5
Female	80.5	85	76.6	83.8	76.7	78.3	77.4	83.7	88.1	82.5

(a) Label two columns in Minitab 'Male' and 'Female', respectively. Enter the heart rates into the corresponding columns.

(b) Note one can sort data in Minitab. Click Data and find Sort. Select the Male and Female columns and sort them from lowest resting heart rate to highest resting heart rate.

(c) Under Graph, select a (simple) Histogram and create a histogram of the Male and Female heart rates. Sketch these in the space below. Describe each histogram: is the distribution symmetric or left/right skewed?

(d) Now produce a histogram with *both* male and female heart rates plotted. Does there appear to be a difference between them? Explain. Be sure to provide a plot.

(e) Find the 5-number summary for male heart rates. Do this by hand, showing all the necessary computations:

Min	Q1	Median	Q3	Max

(f) Under Stat, Basic Statistics, choose 'Display Descriptive Statistics'. Use this to produce a 5-number summary for the female resting heart rate. Give the numbers below.

Min	Q1	Median	Q3	Max

(g) What is the IQR for males? For females? Are there any outliers for either group?

(h) By hand, find the mean male resting heart rate.

(i) Using the following table, find the standard deviation for the male resting heart rate.[1]

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
71.3		
70.7		
68.1		
75.1		
76.2		
74.9		
67.9		
66.1		
72.3		
73.5		
	Total:	

[1] Recall the sample standard variance is given by $s^2 = \frac{1}{n-1}\sum(x_i - \overline{x})^2$

(j) Use Minitab to find the mean and standard deviation for the female resting heart rate:

Mean:

Standard Deviation:

(k) Using the previous parts, does there seem to be a difference between male and female resting heart rate? Explain using your computations and/or graphs from the previous parts.

(l) Given your answer in the previous part, give some reasons for why there is/is not a difference between male and female resting heart rate. Explain why this might be the case.

(m) Assuming the sample of males and females were a good representatives of the population of all men and women, we shall use this to describe the *human* resting heart rate. We need to combine each data set into one data set. Choose 'Data' then 'Stack'. Select the male and female columns and combine them into a column labeled 'Human.'

(n) Use Minitab to produce a histogram of the resting human heart rate. Be sure to provide the image. Find also the 5-number summary as well as the mean, standard deviation, and variance for the resting human heart rate. Then below the data, provide a box plot of the data.

Min:

Q1:

Median:

Q3:

Max:

Mean:

Variance:

Standard Deviation:

Problem 2: Below are the yearly high stock prices for Microsoft from 1990 to the year 2017.

1.12	2.33	2.97	3.06
4.07	6.89	10.77	18.84
36.00	61.09	59.13	38.08
35.31	30.00	30.20	28.25
30.26	37.50	35.96	31.50
31.58	29.46	32.95	38.98
50.05	56.85	64.10	87.50

Enter these values into a column in Minitab labeled 'Microsoft'. Select 'Graph' and then select 'Time Series Plot'. Select the simple option, choose the Microsoft data, and then Time/Scale. Select the desired time scale and appropriate tmei interval. Starting at the year 1990, create a time series plot of this data. Be sure to provide the plot. Describe any trends you see.

Problem 3: Create a stem-and-leaf plot of the following data *by hand*. Use this stem-and-leaf plot to describe the data. Is the distribution symmetric or skewed left/right?

$$42, 43, 40, 61, 43, 56, 69, 41, 47, 46, 84, 62, 50, 64, 46, 87, 56, 58, 61, 60$$

Problem 4: Go to https://www.triolastats.com/es13-datasets and download the Minitab datasets. Open the dataset labeled 'Body Data'.

(a) How are the males and females entered into this spreadsheet?

(b) What type of variable is Gender? (1=Male) Explain.

(c) What other variables are included in this data set?

(d) We want to compare male and female heights. To do this, we will need to separate the data. Go to Data and Select Unstack Columns. Choosing height as the data to unstack (sorting subscripts in Gender), place the Male and Female heights into columns labeled 'M Height' and 'F Height'.

(e) For the data produced in the previous part, use Minitab to find the following:

Male	**Female**
Min:	Min:
Q1:	Q1:
Median:	Median:
Q3:	Q3:
Max:	Max:
Mean:	Mean:
Variance:	Variance:
Standard Deviation:	Standard Deviation:

Does there seem to be a difference between male and female height? Explain.

Problem 5: Go to https://www.triolastats.com/es13-datasets and download the Minitab datasets. Open the dataset labeled 'Bear Measurements'. Examine the data carefully.

(a) What were the variables in this study? What were the (most likely) units for each variable?

(b) For each variable in the study, identify if the variable was quantitative or categorical and what type of measurement it was.

(c) What was the average age of the bear examined? What was the range of ages examined?

(d) What was the median bear length?

(e) How many of the bears were male?

(f) We shall calculate the r value for the linear regression between age and weight. Do you expect there to be a relationship between a bear's age and its weight? Explain.

(g) Go to Stat, Basic Statistics, then Correlation. Enter the bear age and weight. What is the r value, i.e. the Pearson correlation coefficient?

(h) Is r positive or negative? Explain why the sign of r is positive or negative.

(i) Is the data linear? What 'percent' linear is the data?

(j) Given the previous three parts, explain why there seems to be a somewhat linear relationship between a bear's age and its weight. Be sure to explain both mathematically and using 'real world' terms.

Problem 6: Go to https://www.triolastats.com/es13-datasets and download the Minitab datasets. Open the dataset labeled 'Family Heights'. We want to investigate whether there is a relationship between the Father's height and the height of the son.

(a) First, we need to see if the data is linear. Go to Graph, select Scatterplot, then select simple. Choose father's height and first son's height. Examine the scatterplot. Does there appear to be a linear relationship between the two? Explain

(b) If one were to perform a linear regression, would one expect r to be positive or negative? Explain.

(c) If one were to perform a linear regression, would one expect r to be close to 1, close to 0, or close to -1? Explain.

(d) We shall now construct and plot a linear regression for the data. Select Stat, Regression, then Fitted Line Plot. Select for x the father's height and for y the first son's height. Select a linear regression.

(e) What is the linear equation produced?

(f) What is the r value? Is this what we might expect before seeing the plot? After? Explain.

(g) What 'percent' of this data is explained by the linear model?

(h) Do this work imply there is no relationship between father's height and first son's height? Explain

(i) Using the model, predict the height of the first son if the father's height was 72 in.

Problem 7: Explain what is wrong with the following: a data scientist examined hundreds of hours of video footage of NBA games, counting the number of times players from each team high fived other members of their team. Examining the resulting data, the scientist realized that the there was a strong relationship between the number of high fives given and whether the team won or not. They then suggested to Jim Boeheim that his players high five as often as possible. What is wrong with this conclusion and suggestion? What might explain what the scientist observed?

Problem 8: Choose one of the graphs from Homework 2 and use Minitab to create an improved version of the graphic chosen. Be sure to indicate which graph you are using and to provide the new plot.