

Last Name _____, First Name _____

Student # _____

Lab Section (IMPORTANT) _____

Due Tuesday October 11 IN CLASS.

Total mark 100

Part I. Lab questions.

- Data used in this lab are in the Excel file on CuLearn of the course. You will need to copy the data from Excel and paste them into a Minitab worksheet (Open such a worksheet by double-clicking on Minitab).

- Do not include ANY Minitab code to your assignment. Use spaces left to answer lab questions, and attach the printed graphs.

1. The age (in years) of 100 randomly selected tourists in a resort recorded in the column titles "Age" in the Excel file.

a. Construct a frequency histogram for these data such that the first class interval is 0.83 - 10.93. [2]

- *Enter the data in column C1*
- *Select Graph: Histogram. Enter C1 in the Graph variable window. Click OK to view the histogram*
- *Edit the horizontal axis scale. Double click on x-axis and under the Binning tab, select Interval Definition, choose Cutpoint and then enter the two endpoints 0.83 10.93 (with a blank space in between) for the first interval. This way Minitab will construct a histogram with the classes 0.83 - 10.93, 10.93-21.03, etc.*
- *Print your histogram and include it with your assignment*

b. Describe the shape of distribution of this data set. [2] _____

c. What proportion of observations are older than 10.93? [2] _____

d. The mean or the median, which one is greater? [1] _____ why? [1] _____

2. (Refer to “Age” data): Comparing Empirical rule to Tchebycheff theorem.
- a. Use *desc* command (Enable command editor “MTB >” by checking “Enable commands” from Editor in the bar menu) to find the mean [1] _____ and the standard deviation [1]_____.
- b. [2] The following is meant to check how many student heights fall between $\bar{x} \pm 2s$. You will use Minitab to construct a column C5 which will contain only values 1 or 0 according to whether the corresponding age (in column C1) falls in the interval $\bar{x} \pm 2s$, by typing in the following: *let c5=(c1>= $\bar{X} - 2 * S$ and c1<= $\bar{X} + 2 * S$).*
Note Before typing in you will replace \bar{X} and S by their respective values found in part **b.** above.
 Next you will check how many ages did fall in the interval $\bar{X} \pm 2S$ by typing in the following: *tally c5*
- c. What is the percentage of ages that fall between $\bar{X} \pm 2S$? [1] _____ . Is this value close to what the empirical rule suggests for the interval $\bar{X} \pm 2S$. [1] _____
 Does this value agree with Tchebysheff theorem? [1] _____- why? [1] _____
3. The average sales (in Canadian dollars) per customer transaction of 45 randomly selected convenience stores were recorded. The data are listed in the column titled “Sales in CAD per customer transaction-2014” in the excel file.
 Construct a stem-and-leaf chart for this set of data (print your graph)[2] to answer the following questions by:
- *Enter the data in column C2*
 - *Select Graph: Stem-and-Leaf. Enter C2 in the Graph variable window. Click OK to view the graph*
 - *Print your stem-and-leaf graph and include it with your assignment*
- a. How many stores had average sales per transaction less than 13 CAD? [2] _____
- b. How many stores had average sales per transaction of 16 CAD or more? [2] _____
- c. what is the median of the average sales per transaction? [1] _____
4. Refer to the “Sales in CAD per customer transaction-2014” data set above.

Construct a boxplot for this set of data(print it) [4] to answer the following questions by:

- *Select Graph: Boxplots. In the Boxplots window Choose Simple Under "One Y". Click OK. Enter C2 in the variable window. Click OK to view the graph*
- *Print your boxplot and include it with your assignment*

- a. Based on the boxplot, how would you describe the shape of the distribution of this data set [1] _____
- b. The interquartile range (IQR) is approximately equal to [1] _____
- c. Does the data set has any outliers. [1] _____

5. Columns E, F and G are Height, Weight and (body mass index) BMI for 20 patients.

- a. Construct a scatterplot with height marked along the horizontal axis and weight marked along the vertical axis.

You do not need to print out your scatterplot

Calculate the correlation coefficient: [1] _____. If appropriate, fit a least square regression line using height to predict weight (Response variable). What is the equation of regression line? [2] _____

- *Select Stat; Regression. Enter the response variable and the predictor variable. click OK.*

- b. Construct a scatterplot with height marked along the horizontal axis and BMI marked along the vertical axis. Calculate the correlation coefficient: [1] _____. If appropriate, fit a least square regression line using height to predict BMI (Response variable). What is the equation of regression line? [2] _____

- c. Construct a scatterplot with weight marked along the horizontal axis and BMI marked along the vertical axis. Calculate the correlation coefficient: [1] _____. If appropriate, fit a least square regression line using weight to predict BMI. What is the equation of regression line? [2] _____

- d. What is the predicted BMI if weight is 134? [1] _____. If weight is 200. [1] _____. Can you predict the weight if BMI is 25? [1] _____. Why? [1] _____.

Part II. Long-answer questions; Give the solutions for the following questions in details

1. Identify each of the following variables as categorical (i.e. qualitative), discrete or continuous.
 - a. [1] Number of times per year a person catches a cold _____
 - b. [1] Wind speed (Km/hour) in Chicago _____
 - c. [1] The color of a ball drawn from a box containing two red, and 3 white balls _____
 - d. [1] Monthly unemployment rate in Canada _____
 - e. [1] The month in which Ottawa's first major Winter storm will happen in 2017 _____

2. A data set consists of 10 values that are fairly close together. The largest value is replaced by another value but the new value is an outlier (very far away from the other ones).
 - a. [2] How is the mean affected?
 - b. [2] How is the median affected?

3. [4] The average monthly salary of full professors in universities in Canada is 10,800 CAD with the standard deviation 1500 CAD. If a full professor's monthly salary is 12,000 CAD, would consider this salary unusual? (Hint: Use the z-score to justify your answer)

4. The waiting time (in minutes) to speak to an agent of an insurance company of 20 randomly selected callers are recorded as below.

3.06, 2.98, 2.86, 2.81, 2.99, 2.71, 3.11, 2.98, 3.03, 2.95, 2.81, 2.94, 3.09, 2.85, 3.05, 2.96, 3.05, 2.84, 2.84, 2.97

 - a. [2] What is the average waiting time according to this data
 - b. [4] What is the standard deviation of the waiting time data
 - c. [4] Construct a stem-and-leaf graph for this data (by hand)
 - d. [2] Roughly, how would you describe the shape of the graph
 - e. [6] What are the values of the three quartiles (Q_1 , median, Q_3)
 - f. [2] What is the proportion of the measurements in $\bar{X} \pm 2S$
 - g. [2] (Refers to (f)) Is this proportion close to what the empirical rule suggests? Why?

5. The annual fuel cost in CAD for 24 popular vehicles in Canada are as listed below. (Source Statistic Canada Website).

3100, 3100, 900, 1000, 1750, 1750
2250, 1850, 1850, 2250, 3400, 2550
1600, 1700, 2250, 1750, 2000, 2150
2400, 1550, 1550, 1100, 1700, 1850

- [2] 75% of the annual fuel costs are greater than what value.
 - [2] 50% of the annual fuel costs are greater than what value.
 - [2] 75% of the annual fuel costs are smaller than what value.
 - [1] What is the range that contains approximately 50% of the middle values of the annual fuel costs. (Hint: Think of the interpretation of the IQR)
 - Compute the lower and upper fences [2] and construct a box-plot (by hand) for this data set [1], and identify any potential outliers [1].
6. The math and stats final grades (out of 100) for 6 second year students are recorded in the following table.

student	1	2	3	4	5	6
math grade	15	88	91	72	64	58
stats grade	5	92	94	72	62	54

- [4] Find the correlation coefficient between math and stats grads
- [5] Find the required regression line that enables you to predict the stats grades of the students based on their math grads.
- [2] What is your prediction for the stats grade of a student whose math grade is 75.