Stat 2507	Assignment (1)	Fall 2019

Last Name: _____ **First Name:** _____
Student ID: _____ **Lab Section:** _____

Section A due Monday, September 30 before 10:05 am
Section B due Tuesday, October 1 before 6:05 pm
Section C due Tuesday, October 1 before 10:05 am
Section D due Monday, September 30 before 1:05 pm

Important Notes:

- You MUST write your lab section in the space provided above.
- You MUST use Minitab for the first 3 questions. The other questions must be solved by hand.
- Pick up your marked assignment from the TA during your lab time.

Question:	1	2	3	4	5	6	7	**Total**
Marks:	11	15	11	6	8	21	18	90
Score:								

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 and paste them into a Minitab worksheet (Open such a worksheet by double-clicking on Minitab).
- Use spaces left to answer lab questions, and attach the printed graphs. Do not include ANY Minitab code to any question on your assignment unless it is said so.

1. (11 points) The weights (in kgs) of 50 ground beef packages are listed in the column titled "WEIGHT"
   (a) (2 points) Construct a stem-and-leaf chart for this set of data and use it to answer the following questions

Solution:

Stem-and-Leaf Display: WEIGHT

Stem-and-leaf of WEIGHT N = 50

2	11	68
4	12	12
12	12	55578899
15	13	114
(13)	13	5566677778999
22	14	00012234
14	14	5678999
7	15	11
5	15	667
2	16	12

Leaf Unit = 0.1

(b) (3 points) What are the smallest and the largest weights? **11.6** and **16.2**

(c) (2 points) What is the average weight of the ground beef packages? **13.858**

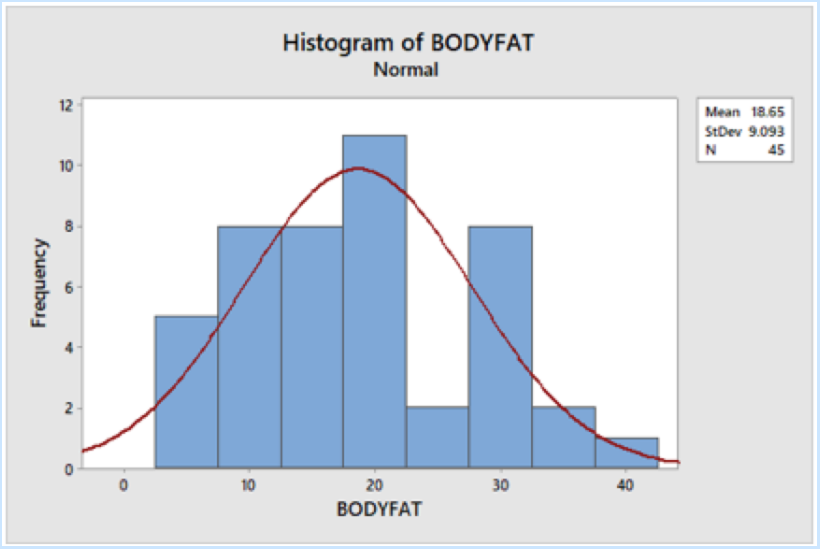
(d) (2 points) Find the first quartile Q_1 and the third quartile Q_3. **13.05** and **14.625**

(e) (2 points) Two Quarters of the weights are below what value? **Q_2=13.85**

2. (15 points) The Bodyfat of a sample of 45 patients at a general clinic are recorded in the column titled "BODYFAT" in the Excel file

(a) (3 points) Construct a frequency histogram for these data as following

Solution:



(b) (2 points) Is the variable "bodyfat of patients" qualitative or quantitative? **Quantitative**

(c) (2 points) Describe the shape of distribution of this data set. **bell-shaped or symmetrical**

(d) (2 points) Based on the shape of your histogram,how would you expect the mean to relate to the median? Explain **Based on the shape of histogram,we expected the mean to be close or equal to median because the shape is bell-shaped or symmetrical**

(e) (4 points) Obtain summary statistics for this data set as following

- *Click "View" in the bar menu and check "Command line/history".*
- *Use the command desc for the column C2*

What values does MINITAB give as the mean and median? The mean is **18.65** and median is **19**.
Do the values of the mean and median agree with your answer in part (e)? **Yes**
Do these values represent population parameters or sample statistics? Explain.

Solution: These values represent sample statistics because they are computed based on a sample of 45 patients at a general clinic.

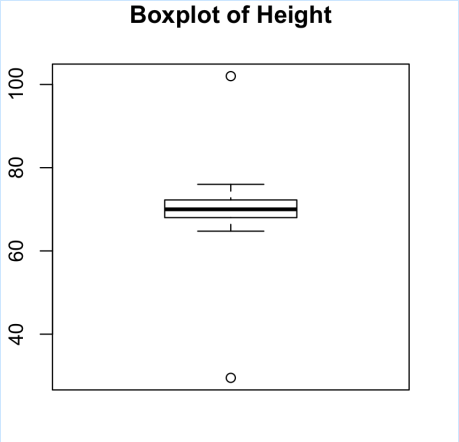
(f) (2 points) Would you use TChebychev's Theorem or the empirical rule for this data?

Solution: we can use both. TChebychev's Theorem and Empirical Rule. TChebychev's Theorem can be applied to any shape distribution and also we can use the Empirical Rule since the shape of data set is symmetrical

3. (11 points) The data in the column "HEIGHT" are the recorded height in cm of the sample of 45 patients

(a) (2 points) Construct a boxplot for this data set

Solution:



(b) (1 point) Describe the shape of the boxplot. **symmetrical**

(c) (1 point) Based on your boxplot, do you see any outlier? **yes** if Yes, list the value(s) you think can be outliers **The possible outliers are 29.5 and 102**

(d) (3 points) Find the first quartile Q_1 **67.88** the second quartile Q_2 **70.00** and the third quartile Q_3 **72.25**

(e) (2 points) By hand, Compute the lower fence **LF= Q_1 - 1.5 *IQR= 67.88 - 1.5*(4.37)= 61.325** and the upper fence **UF=Q_3 + 1.5 *IQR= 72.25+ 1.5*(4.37) = 78.805**

(f) (2 points) Compare your candidates for outliers in part (c) to the lower and the upper fence. **The value 29.50 is below LF and the value 102 above UF**.Do you think that theses value(s) in part (c) still are outlier? **yes both values as outliers**

Part II. Long-answer questions; Give the solutions for the following questions IN DETAILS.

4. (6 points) Identify each of the following variables as qualitative, (i.e. categorical) or qualitative. If a variable is quantitative then identify whether it is discrete or continuous
   (a) (1 point) Types of foods soccer players in Brazil like to eat. **Qualitative**.
   (b) (1 point) Years of education required to get a job opportunity in nuclear engineering. **Quantitative,discrete or Continuous**.
   (c) (1 point) The size of a team soccer T-shirt (small, medium, large, extra large). **Qualitative**
   (d) (1 point) The dollar amount on an accounts receivable invoice **Quantitative,discrete**.
   (e) (1 point) Temperature of a pizza oven at a particular time **Quantitative, Continuous**.
   (f) (1 point) The weight (in grams) of the sandwich meat prepared weekly in a local restaurant **Quantitative, Continuous**

5. (8 points) The number of children less than 18 years in a randomly selected sample of 27 families is given below

1	0	2	4	4	2	3	2	0	2	0	0	0	0
1	4	0	2	0	2	0	2	3	2	1	2	2	

   (a) (2 points) Construct a frequency distribution table using single-valued classes (i.e. 0, 1, $\cdots$)
   (b) (3 points) Compute the relative frequencies, cumulative relative frequencies, and percentages for all classes.

Solution:
part (a) and (b) The frequency distribution, relative frequencies and percentages are shown in the following Table

Number of children <18 years of age	**Frequency Frequency**	**Relative frequency**	**Cumulative relative frequency**	**Percentage (probability)**
0	9	0.334	0.334	33.3 %
1	3	0.111	0.445	11.1 %
2	10	0.370	0.815	37 %
3	2	0.074	0.889	7.4%
4	3	0.111	1.0	11.1%
Total	27			100 %

   (c) (1.5 points) Find the number of families with one or two children less than 18 years of age.

Solution: Number of families with one or two children under 18 years of age $= 3 + 10 = 13$.

   (d) (1.5 points) Find the number of families with two or three children less than 18 years of age.

Solution: Number of families with two or three children under 18 years of age $= 10 + 2 = 12$.

6. (21 points) The number of hours worked by 24 employees of a company is given below:

40	43	40	39	36	44	40	39	39	52	27	50
41	47	40	48	38	36	25	41	35	36	16	40

(a) (6 points) Calculate the mean, variance and standard derivation for the given data

Solution: The sample mean is

$$\bar{x} = \frac{\sum_{i=1}^{24} x_i}{24} = \frac{932}{24} = 38.83$$

The sample variance is

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{24} x_i^2 - \frac{(\sum_{i=1}^{24} x_i)^2}{n}\right] = \frac{1}{24-1}\left[37594 - \frac{(932)^2}{24}\right] = 60.93$$

The sample standard deviation is $\sqrt{s^2} = \sqrt{60.93} = 7.81$

(b) (6 points) Calculate the three quartiles (Q_1, Q_2, and Q_3) and the Interquartile range (IQR).

Solution: $Q_1 = 25^{th}$ percentile
Position of 25^{th} percentile $= 0.25(24+1) = 6.25$
$Q_1 = 25^{th}$ percentile $= 36 + 0.25(36–36) = 36$

$Q_2 = 50^{th}$ percentile
Position of 50^{th} percentile $= 0.50(24+1) = 12.5$
$Q_2 = 50^{th}$ percentile $= 40 + 0.5(40–40) = 40$

$Q_3 = 75^{th}$ percentile
Position of 75^{th} percentile $= 0.75(24+1) = 18.75$
$Q_3 = 75^{th}$ percentile $= 41 + 0.75(43–41) = 42.5$

(c) (4 points) Calculate the values of the lower fence and the upper fence for a boxplot.

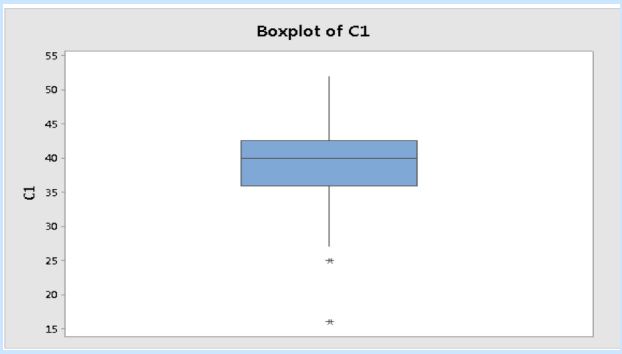
Solution: We define the lower and upper limits, the numbers that lie, respectively 1.5 IQRs below the first quartile and 1.5 IQRs above the third quartile. That is

Lower limit $= Q_1$–$1.5\ IQR$
Upper limit $= Q_3 + 1.5\ IQR$
The interquartile is $IQR = Q_3 - Q_1 = 42.5 - 36 = 6.5$

Lower fence $= Q_1$–$1.5\ IQR = 36 - 1.5(6.5) = 26.25$
Upper fence $= Q_3 + 1.5\ IQR = 42.5 + 1.5(6.5) = 52.25$

(d) (5 points) Construct a box-and-whisker plot. Comment on the shape of the distribution of the data. List any potential outliers, if any.

Solution: The shape of the distribution of the data is symmetrical. There are two outliers 16 and 25 shown by an asterisk, which are below lower fence 26.25



7. (18 points) The following is the number of passengers per flight in a sample of 34 flights from Ottawa, Ontario, to Hampton, Washington in 2018.

78	73	75	99	50	58	25	56	57	55	59	55	62	69	77	66	51
21	53	30	51	63	52	57	68	75	66	65	69	79	72	65	53	50

(a) (2 points) What is the population under consideration here? What is the used measurement unit?

Solution:
The **population** is all flights from from Ottawa, Ontario, to Hampton, Washington in 2018.
The **unit** is a particular flight from from Ottawa, Ontario, to Hampton, Washington in 2018.

(b) (2 points) What is the variable of interest here? Is it quantitative and continuous, or quantitative and discrete?

Solution:
The **variable** is the number of passengers per flight. It is quantitative and discrete.

(c) (2 points) Thirty percent of flights have passengers exceeding what value?

Solution:
The Position of 70^{th} percentile$= 0.70(n+1) = 0.70(34+1) = 24.5$
The 70^{th} percentile $= 68 + 0.5(69-68) = 68.5$

(d) (3 points) Draw a stem-and-leaf plot with stems 1, 2, 3 $\cdots$. What is the leaf unit?

Solution:
The stem-and-leaf plot follows. Note that the leaf unit, LU = 1.0. 2 | 1
2 | 5
3 | 0
3 |
4 |
5 | 0011233
5 | 55667899
6 | 23
6 | 5566899
7 | 23
7 | 55789
8 |
8 |
9 |
9 | 9

(e) (1 point) Describe the shape of the data set.

Solution:
The stem and leaf plot shows that the data is approximately bell-shaped and symmetric with long tails, or slightly left-skewed.

(f) (6 points) Find the percentage of measurements falling in the intervals $\bar{x} \pm ks \; for \; k = 1, 2, 3$.

Solution:
The following table gives the actual percentage of measurements falling in the intervals $\bar{x} \pm ks \; for \; k = 1, 2, 3$.

k	$\bar{x} \pm ks$	Interval	No in Interval	Percentage
1	60.412 ± 15.45	$(44.962, 75.862)$	27	79 %
2	$60.412 \pm 2(15.45)$	$(29.512, 91.312)$	31	91 %
3	$60.412 \pm 3(15.45)$	$(14.062, 106.762)$	34	100 %

(g) (2 points) How do you compare the percentages obtained in part (f) with those given by the Empirical Rule? Explain.

Solution:
The percentages in part (f) do not agree too closely with those given by the Empirical Rule, especially in the one standard deviation range. This is caused by the lack of mounding (indicated by the gap) in the distribution.

End of Assignment 1.