

Carleton University

School of Mathematics and Statistics

STAT 2507 – Assignment #1

Section A due Tuesday, October 2, 2012 before 10:05am

Section B due Tuesday, October 2, 2012 before 6:05pm

Section C due Monday, October 1, 2012 before 2:35pm

INSTRUCTIONS: For written questions, show all of your work. No credit will be given for answers without justification. Do not use MINITAB for a question unless it specifically says to do so. (Total Marks=50)

Part A: Lab Part

- I. Consider the following 50 measurements for questions 1,2,3: (**File -> Open Worksheet...**...select Asst1data1.txt. The data appear in column C1.)

3.1 4.5 2.9 2.7 3.8 5.1 2.5 3.6 4.3 5.6
4.9 3.5 2.1 4 6.2 1.8 3.6 6.1 5.7 4.9
2.8 3.7 3.5 4.4 2.5 5.6 5.1 4.7 3.7 4.2
3.6 4.1 4 3.7 2.9 2.2 4.8 3.9 4.6 3.1
2.5 4.9 3.7 4.2 2.8 3.4 1.6 3.9 4 3.9

1. Construct a histogram for these 50 measurements.(Click **Graph->Histogram...**,select **Simple**)
- Describe the shape of the histogram. Do you see any outliers?[1]
 - For the class with a midpoint of 5, specify (1) the class width, (2) the lower class boundary, and (3) the upper class boundary. [1.5]
 - What percentage of measurements are less than 5.25? [0.5]

Solutions:

a The histogram has a mound shaped distribution. No outlier.

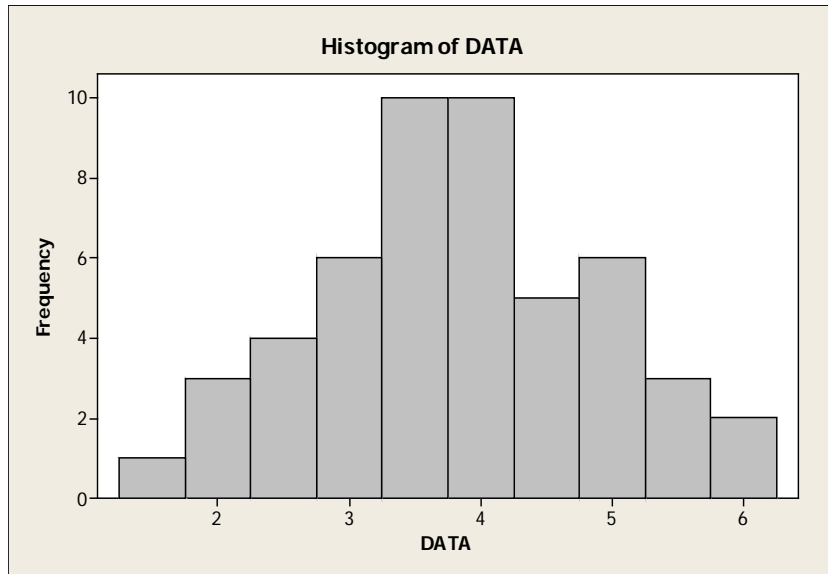
b (1) Class width is the difference between any two midpoints.

Therefore the class width $= (3-2)/2 = 0.5$

(2) The lower class boundary $= 5 - 0.5/2 = 4.75$

(3) The upper class boundary $= 5 + 0.5/2 = 5.25$

c 90% of measurements are less than 5.25



2. Construct a stem-and-leaf plot for these 50 measurements.(Click **Graph->Stem and leaf**)
- Describe the shape of the distribution. Do you see any outliers?[1]
 - Use the stem and leaf plot to find the smallest observation.[1]
 - Find the eighth and ninth largest observation.[1]

Solution:

- The stem and leaf display has a mound shaped distribution. No outlier.
- From the stem and leaf display, the smallest observation is 1.6.
- The eighth and ninth largest observations are both 4.9 .

Stem-and-Leaf Display: DATA

Stem-and-leaf of DATA N = 50
Leaf Unit = 0.10

```

2   1   68
4   2   12
12  2   55578899
15  3   114
(13) 3   5566677778999
22  4   00012234
14  4   5678999
7   5   11
5   5   667
2   6   12

```

3. Calculate the mean, median and standard deviation for these 50 measurements.(**Click->Basic Statistics->Display Describe Statistics...**)[1.5]

Solutions: Mean is 3.858 , Median is 3.850, Standard deviation is 1.086

- II. Four Eagles Service Station selected a random sample of customers. The amount of gasoline(in gallons) purchased by each customer is given in the following: (Consider the following 30 measurements for questions 4,5,6.). (To open the data: Click **File -> Open Worksheet...**...select Asst1data2.txt. The data appear in column C1.)

7.3 9.6 8.9 9 9.1 8 14 7.9 9.1 12.4
 9.5 7.7 10.1 13.1 8.3 6 9.1 8.2 11.3 6.1
 9.2 11.2 7.4 8.6 11 13 11.6 7.3 10 17

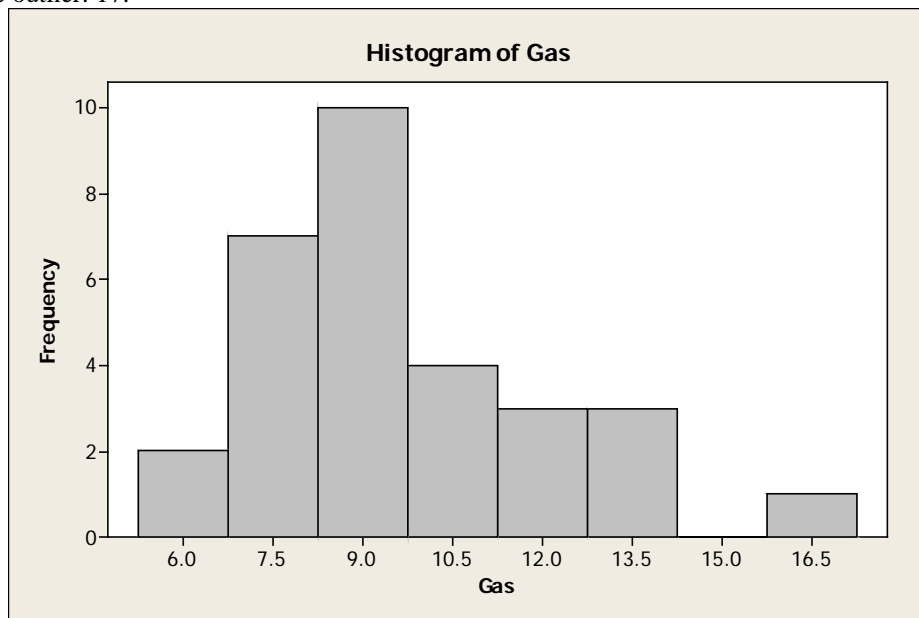
4. Calculate the mean, median and standard deviation for these 30 measurements.(**Click->Basic Statistics->Display Describe Statistics...**). Use these measures to decide whether or not the data are symmetric or skewed.[2]

Solutions: Mean is 9.7 , Median is 9.1, Standard deviation is 2.445 . The mean is larger than the median, indicate a distribution skewed to the right.

5. Construct a histogram for these 30 measurements.(Click **Graph->Histogram...**,select **Simple**)
- Describe the shape of the distribution. [0.5]
 - Are there any outliers in the set? If so, which numbers are outliers?[1]

Solution:

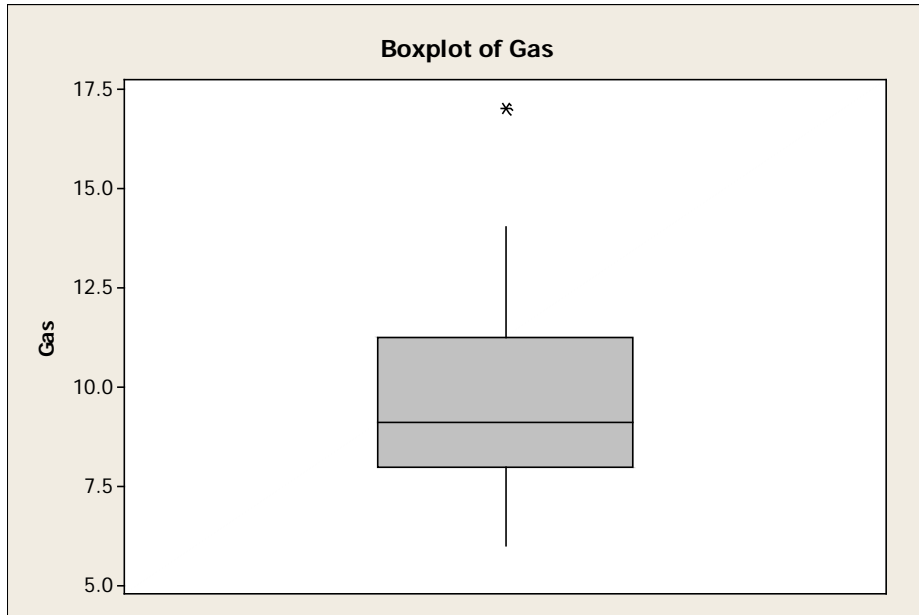
- The distribution is skewed to the right
- There is one outlier: 17.



6. Draw a box plot for these 30 measurements.(Click **Graph->Box plot...**)
- Describe the shape of the distribution. [0.5]
 - Are there any outliers in the set? If so, which numbers are outliers?[1]

Solution:

- a The distribution is skewed to the right
- b There is one outlier: 17.

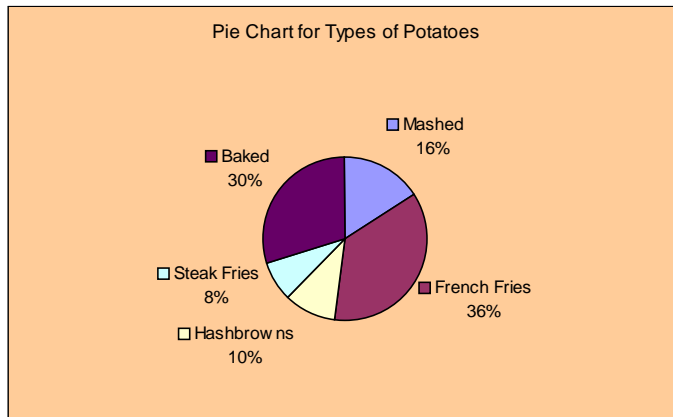


- III. A study was performed to examine preferences for various types of potatoes: Mashed, French Fries, Hash browns, Steak Fries, and Baked. The following data were recorded regarding the preferences of 150 people: (Consider the following 5 measurements for questions 7,8,9.). (To open the data: Click **File** -> **Open Worksheet**...select Asst1data3.txt. The data appear in column C1.)

| Type of Potato | Frequency |
|----------------|-----------|
| Mashed | 24 |
| French Fries | 54 |
| Hash browns | 15 |
| Steak Fries | 12 |
| Baked | 45 |

7. Construct a pie chart to describe the data.[1]

Solution:



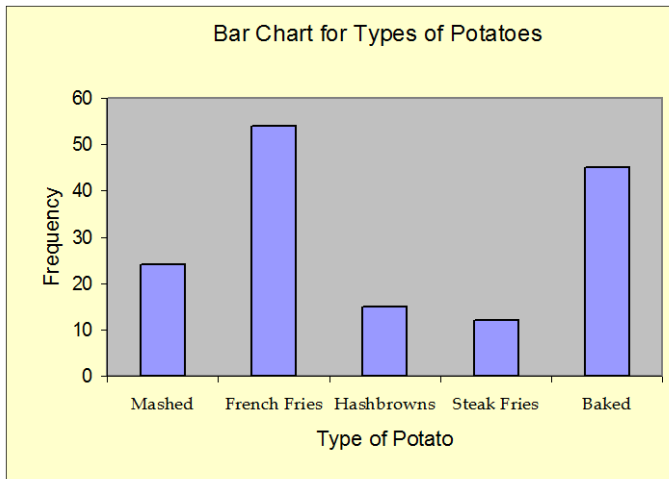
8. What proportion of people prefers French Fries or Steak Fries?[1]

Solution:

Proportion of people prefers French Fries or Steak Fries = $0.36 + 0.08 = 0.44$.

9. Construct a bar chart to describe the data.[1]

Solution:



Part B: Written Questions

1. [3.5] Identify each of the following variables as either quantitative discrete, quantitative continuous, or qualitative.
 - a. The brands of ice cream that you purchase
 - b. The number of heads obtained if a coin is tossed three times.
 - c. The amount of sugar consumed by Canadians in one year

- d. The species of fish in the zoo
- e. The lengths of time children wait for the school bus
- f. Your favourite professional football team
- g. Rating of the effectiveness of a new cold remedy (Not effective, effective).

Solutions : Qualitative, Discrete, Continuous, Qualitative, Continuous, Qualitative, Qualitative.

2. The numbers of rooms for 15 home recently sold were;

8, 8, 8, 5, 9, 8, 7, 6, 6, 7, 7, 7, 7, 9, 9

- a. [3]Compute 10th, 65th, and 90th percentiles.
- b. [4]Compute Q_1 , Q_2 , Q_3 and IQR.
- c. [3]Compute mean, median, mode.
- d. [4]Compute range, variance, standard deviation.

Solution:

- a. 6, 8, 9.
- b. 7, 7, 8, 1
- c. 7.4, 7, 7,
- d. 4, 1.4, 1.183

3. Consider the following data:

4.7 4.9 6.8 5.1 5.4 3.5 6.6 6.1 6.4 3.9
 4.1 6.4 5 3.6 7.1 2.9 6.4 2.7 3.6 4.7
 5.8 5.5 4.1 5.2 4.2 5.7 5.9 5.1 3.6 5.7

- a. [4]Draw a stem-and-leaf plot for the data with stem unit 1 and a leaf unit = 0.1.
- b. [3]Find the mean , the median and the standard deviation.
- c. [1]Describe the shape of the distribution (is it symmetric, skewed left, skewed right, bell-shaped,etc., are there obvious outliers).
- d. [6]Find the percentage of measurements falling in the intervals $\bar{x} \pm ks$ for $k = 1, 2, 3$.
- e. [2]How do the percentages obtained in part d compare with those given by the empirical Rule? Explain.
- f. [1] The largest numbers is 7.1. Is this number unusually big?

Solution:

a.

Stem-and-Leaf Display: C1

Stem-and-leaf of C1 N = 30

Leaf Unit = 0.10

```

2  2  79
2  3
7  3  56669
10 4  112
13 4  779
(5) 5  01124
12 5  57789
7  6  1444
3  6  68
1  7  1

```

b. Mean is 5.023 , median is 5.1, standard deviation is 1.199.

c. The stem and leaf plot shows that the data is roughly bell-shaped.

d. 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 | Number in Interval | Percentage |
|-----|-----------------------|----------------|--------------------|------------|
| 1 | 5.023 ± 1.199 | 3.824 to 6.222 | 21 | 70% |
| 2 | $5.023 \pm 2 * 1.199$ | 2.625 to 7.421 | 30 | 100% |
| 3 | $5.023 \pm 3 * 1.199$ | 1.426 to 8.620 | 30 | 100% |

e. The percentages in part **d** 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 centre of the distribution.

f. For $x=7.1$,

$$z_score = \frac{x - \bar{x}}{\sigma} = \frac{7.1 - 5.023}{1.199} = 1.7323 < 2$$

So, 7.1 is not unusually big.