

Assignment #1 SOLUTIONS

1)

The Executive Board of Facebook wanted to know which kinds of professional people were accessing their website regularly over a two-month period. To that end, a random sample from its database of active users was collected. The following data were recorded:

User	Number	Average Frequency of Access per Week
Business Persons	24	35
Academics in the Arts	54	21
Engineers	15	10
Academics in the Sciences	12	8
Lawyers	45	15
Medical Personnel	50	11

1a) Determine the relative frequencies of users.

(1 point)

ANS:

User	Number	Relative frequency
Business Persons	24	$24/200=0.12$
Academics in the Arts	54	$54/200=0.27$
Engineers	15	$15/200=0.075$
Academics in the Sciences	12	$12/200=0.06$
Lawyers	45	$45/200=0.225$
Medical Personnel	50	$50/200=0.25$

1b) Approximately what proportion of their users are academics?

(1 point)

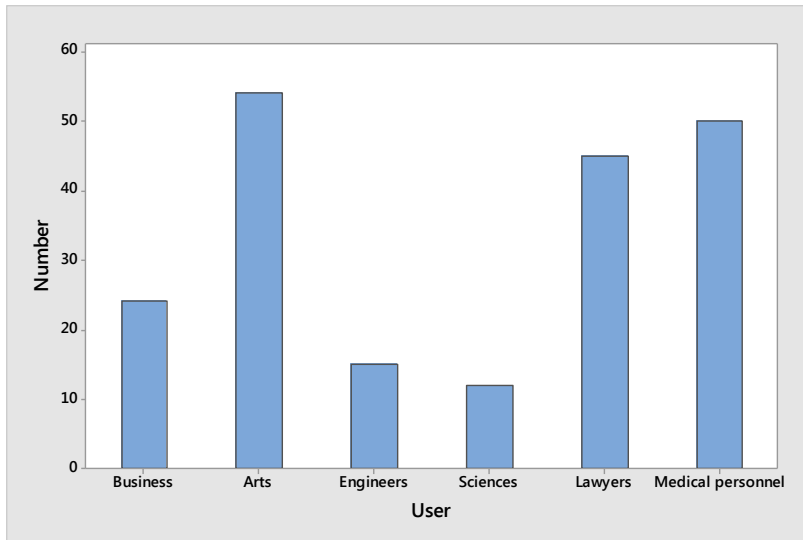
ANS:

Proportion of their users which are academics is $0.27 + 0.06 = 0.33$.

1c) Construct a relative frequency bar chart. Interpret the bar chart.

(2 points)

ANS: Note: Shape cannot really be described except in general such as which one is 'most frequent'.



2)

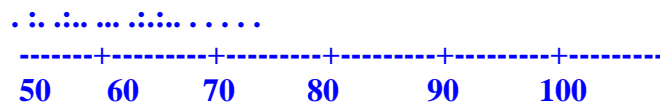
The MAST 250 instructor would like to present the midterm results to her class in a way that shows the overall spread of the data. The 25 test scores for the midterms are listed below.

45 78 62 98 50 61 91 89 57 64 77 69 82
73 70 73 79 80 68 72 62 65 78 50 95

2a) Construct and interpret the dotplot.

(2 points)

ANS: The points are fairly evenly distributed between 45 and 98, with most scores between 60 and 80.



2b) Calculate Z-scores for the largest and smallest values to determine whether there are any outliers. Report on your findings.

(1 point)

ANS: There are no outliers. If using $n-1$ for SD, Z-score for largest value (98) is 1.92, and Z-score for smallest value (45) is -1.92. If using n for SD Z-scores are 1.96 and -1.96, respectively.

3)

Montreal's Department of Sanitation would like to start charging by the weight of a customer's garbage rather than the number of cans. The weights (in kilograms) of 90 randomly selected cans of garbage are summarized in the chart below.

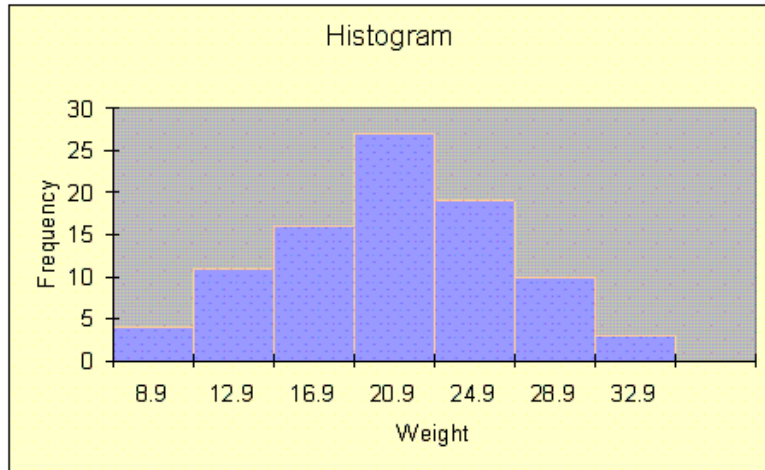
Class	Interval	Frequency
1	4.9 to < 8.9	4
2	8.9 to < 12.9	11

3	12.9 to <16.9	16
4	16.9 to <20.9	27
5	20.9 to <24.9	19
6	24.9 to <28.9	10
7	28.9 to <32.9	3

3a) Construct the frequency histogram.

(2 points)

ANS:



3b) What percentage of the cans weigh less than 20.9 kilograms?

(1 point)

ANS: $(58/90) * (100\%) = 64.44\%$

3c) What percentage of the cans weigh at least 24.9 kilograms?

(1 point)

ANS: $(13/90) (100\%) = 14.44\%$

4)

Eight doctors at the University Health Center were asked how many flu shots they had given to students this fall. The numbers of flu shots were 6, 3, 5, 24, 2, 6, 0, and 8.

4a) Find the sample mean and median number of flu shots given.

(1 point)

ANS:

$$\bar{x} = 6.75$$

$$m = 5.5$$

4b) Based on the values of the mean and median in the previous two questions, are the measurements symmetric or skewed? Justify your answer.

(1 point)

ANS:

Since the mean is larger than the median, we conclude that the measurements are skewed to the right.

5)

The following data represent the number of almond pieces for a sample of eight chocolate almond bars: 4, 6, 10, 1, 3, 1, 25, and 8.

5a) What is the average number of almond pieces per bar?

(1 point)

ANS: $\bar{x} = 7.25$

5b) Which, if any, of the observations appear to be outliers? Justify your answer.

(1 point)

ANS: The value 25 has a z-score of 2.26 making it a suspect outlier. (May also come to this conclusion based on calculating IQR)

5c) What is the standard deviation for the number of almond pieces per bar.

(1 point)

ANS:

$$s = \sqrt{\frac{852 - (58)^2 / 8}{7}} = 7.85$$

6)

Twenty-eight applicants interested in working at Statistics Canada took an examination designed to measure their aptitude for data analysis. A stem-and-leaf plot of the 28 scores appears below, in which the first column is the count per “branch,” the second column is the stem value, and the remaining digits are the leaves.

Count Stems Leaves

1	4	6
1	5	9
4	6	3688
6	7	026799
9	8	145667788
7	9	1234788

6a) Should the Empirical Rule be applied to this data set? Justify your answer.

(1 point)

ANS:

No. The data do not appear to be mound-shaped.

6b) Use the range approximation to determine an approximate value for the standard deviation. Is this a good approximation?

(1 point)

ANS: $s \approx R/4 = 13$. This approximation is very close to the actual value of $s = 12.85$.

6c) Construct a box plot for these data.

(2 points)

Position of first quartile = $0.25(29) = 7.25$, then $Q_1 = 70 + 0.25(2) = 70.5$

Position of third quartile = $0.75(29) = 21.75$, then $Q_3 = 88 + 0.75(3) = 90.25$

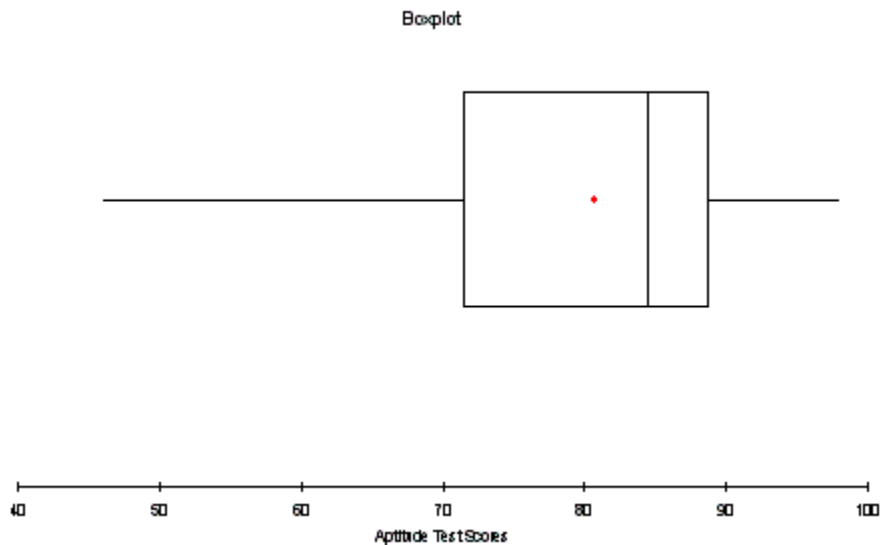
Whiskers should extend to 46 (smallest value) and 98 (largest value)

(May also be calculated but is not shown automatically by minitab and so is omitted from below):

Lower fence: $Q_1 - 1.5(IQR) = 70.5 - 1.5(19.75) = 40.875$, and

Upper fence: $Q_3 + 1.5(IQR) = 90.25 + 1.5(19.75) = 119.875$

ANS:



6d) Is there any indication of outliers?

(1 point)

ANS:

There do not appear to be any outliers present since there are no observations outside the fences.

7)

The following data represent the numbers of Facebook “pokes” a sample of people receive per year: 73, 74, 76, 77, 79, 79, 83, 84, 88, 84, 84, 85, 86, 86, 87, 87, 88, 91, 92, 92, 93, 97, 98, 98, 81, and 82.

7a) Create a stem-and-leaf plot for the distribution of pokes.

(2 points)

ANS:

<u>Stems</u>	<u>Leaves</u>
7	34
7	6799
8	123444
8	5667788
9	1223
9	788

7b) What percentage of measurements would you expect to be between 71.60 and 99.48?
(2 points)

Mean: 85.54, SD: 6.97

2SD below the mean: 71.60

1SD below the mean: 78.57

1SD above the mean: 92.51

2SD above the mean: 99.48

ANS: Since the distribution appears to be relatively mound-shaped, the Empirical Rule applies. The interval (71.60, 99.48) represents two standard deviations from the mean, so we would expect approximately 95% of the measurements to lie in this interval.