Tutorial Worksheet 1 - Reviews of Descriptive Statistics

Objective Questions

1. The _____ is the value you calculate when you want the arithmetic average.

   (a) Mean   (b) Median   (c) Mode   (d) All of the above

2. The process of arranging data into rows and columns is called

   (a) Classification   (b) Frequency distribution   (c) Tabulation   (d) Array

3. Find the median of the following data: 160, 180, 200, 280, 300, 320, 400

   (a) 140   (b) 300   (c) 180   (d) 280

4. The "average" type of grass used in UAE lawns is best described by

   (a) the mean   (b) the median   (c) the mode   (d) the standard deviation

5. The median is a better measure of central tendency than the mean if

   (a) the variable is discrete   (b) the distribution is skewed   (c) the variable is continuous   (d) the distribution is symmetric

6. A set of data points follow a simple linear relation $y = 3x + 2$, where x is any integer number. The mean of the values of y for all values of x in the range $[1 \ldots 100]$ (equally probable) is

   (a) 50   (b) 50.5   (c) 152   (d) 153.5

7. The GM of the following data will be calculated as $X = [50, 125, 70, 56, 49, 98]$

   (a) 70   (b) 74   (c) 100   (d) 101

8. What is the primary characteristic of a set of data for which the standard deviation is zero?

   (a) All values of the variable appear with equal frequency.   (b) All values of the variable have the same value.   (c) The mean of the values is also zero.   (d) None of the above is correct.

9. If the standard deviation of x, y, z is p then the standard deviation of $3x + 5, 3y + 5, 3z + 5$ is?

   (a) $3p + 5$ (b) $3p$ (c) $p + 5$ (d) $9p + 15$

Subjective Questions

Problem 1. The wickets taken by a bowler in 10 cricket matches are as follows: 2   6   4   5   0   2   1   3   2   3   Find the mode of the data.

Problem 2. If the mean of a frequency distribution is 100 and the coefficient of variation is 45%, then what is the value of variance.

Problem 3. For a given sample, the observation is as follows.

x	1	2	3	4	5	6
$f(x)$	25	50	10	30	40	20

x denotes a sample value and $f(x)$ denotes the frequency of occurrence of x. Find the five-point summary of the above data.

X	1	2	3	4	5	6	7	8	9	10
F(x)	1	3	5	7	9	2	4	6	1	0

Problem 4. Calculate the mean, median and mode of the following data: $5, 10, 10, 12, 13$. Are these three equal?

Problem 5. A frequency distribution of a set of 10 data is given below. Calculate the coefficient of variance of the data.

Problem 6. The accompanying data on the number of minutes used for cell phone calls in one month was generated to be consistent with summary statistics published in a report of a marketing study of San Diego residents (**Tele-Truth, March 2009**):

$$
\begin{array}{ccccccccc}
189 & 0 & 189 & 177 & 106 & 201 & 0 & 212 & 0 \\
306 & 0 & 0 & 59 & 224 & 0 & 189 & 142 & 3 \\
71 & 165 & 236 & 0 & 142 & 236 & 130 & &
\end{array}
$$

(a) Would you recommend the mean or the median as a measure of center for this data set? Give a brief explanation of your choice.

(b) Compute a trimmed mean by deleting the three smallest observations and the three largest observations in the data set and then averaging the remaining 19 observations. What is the trimming percentage for this trimmed mean?

(c) What trimming percentage would you need to use in order to delete all of the 0 minute values from the data set? Would you recommend a trimmed mean with this trimming percentage? Explain why or why not.

Problem 7. Suppose that 10 patients with meningitis received treatment with large doses of penicillin. Three days later, temperatures were recorded, and the treatment was considered successful if there had been a reduction in a patient's temperature. Denoting success by S and failure by F, the 10 observations are

$$\text{S \quad S \quad F \quad S \quad S \quad S \quad F \quad F \quad S \quad S}$$

(a) What is the value of the sample proportion of successes?

(b) Replace each S with a 1 and each F with a 0. Then calculate $\bar{x}$ for this numerically coded sample. How does $\bar{x}$ compare to $\hat{p}$?

(c) Suppose that it is decided to include 15 more patients in the study. How many of these would have to be S's to give $\hat{p} = .80$ for the entire sample of 25 patients?

Problem 8. The Insurance Institute for Highway Safety, published data on repair costs for cars involved in different types of accidents on June 11, 2009. In one study, seven different 2009 models of mini-cars and micro-cars were driven at 6 mph straight into a fixed barrier. The following table gives the cost of repairing damage to the bumper for each of the seven models:

Model	Smart Fortwo	Chevrolet Aveo	Mini Cooper	Toyota Yaris	Honda Fit	Hyundai Accent	Kia Rio
Repair Cost	$1,480	$1,071	$2,291	$1,688	$1,124	$3,476	$3,701

(a) Compute the values of the variance and standard deviation. The standard deviation is fairly large. What does this tell you about the repair costs?

(b) The Insurance Institute for Highway Safety also gave bumper repair costs in a study of six models of minivans (December 30, 2007). Write a few sentences describing how mini-cars, micro-cars, and minivans differ with respect to typical bumper repair cost and bumper repair cost variability.

Model	Honda Odyssey	Dodge Grand Caravan	Toyota Sienna	Chevrolet Uplander	Kia Sedona	Nissan Quest
Repair Cost	$1,538	$1,347	$840	$1,631	$1,176	$1,603

Problem 9. For the data on the number of minutes used for cell phone calls published by San Diego residents used in Problem 6.

(a) Compute the values of the quartiles and the interquartile range for this data set.

(b) Explain why the lower quartile is equal to the minimum value for this data set. Will this be the case for every data set? Explain.

Problem 10. Fiber content (in grams per serving) and sugar content (in grams per serving) for 18 high fiber cereals are shown below.

Fiber Content	7	10	10	7	8	7	12	12	8	13	10	8	12	7	14	7	8	8
Sugar Content	11	6	14	13	0	18	9	10	19	6	10	17	10	10	0	9	5	11

(a) Find the median, quartiles, and interquartile range for the fiber content data set.

(b) Find the median, quartiles, and interquartile range for the sugar content data set.

(c) Are there any outliers in the sugar content data set?

(d) Explain why the minimum value for the fiber content data set and the lower quartile for the fiber content data set are equal.

(e) Construct a comparative boxplot and use it to comment on the differences and similarities in the fiber and sugar distributions.