

DESCRIPTIVE STATISTICS

VOCABULARY & PROBLEMS

1. VOCABULARY

- (a) **Population:** The complete set of numerical information on a particular QUANTITY in which an INVESTIGATOR is interested. We assume a population consists of N values.
- (b) **Sample:** An observed subset of POPULATION VALUES. We assume a sample consists of n values.
- (c) **Mean(arithmetic mean or average):** The sum of all the values in the sample divided by the number of values in the sample/population.
- (d) **Median:** The value separating the higher half of a sample/population from the lower half. Found by arranging all the values from the lowest to highest and taking the middle one (or the mean of the middle two if there are an even number of values).
- (e) **Variance:** A measure of the amount of spread (variation) in a set of data; the larger the variance, the more scattered the observations on average.
- (f) **Frequency:** The number of times that a particular value occurs as an observation.
- (g) **Frequency distribution:** The information consisting of the possible values/groups and the corresponding frequencies is called the frequency distribution.
- (h) **Measures of central tendency:** Measures of the location of the middle or the center of a distribution. The definition of MIDDLE or CENTER is purposely left somewhat vague so that the term central tendency can refer to a wide variety of measures. The three most common measures of central tendency are the mean, median, and mode.
- (i) **Mode:** The mode is the most frequently occurring value in a set of discrete data. There can be more than one mode if two or more values are equally common.
- (j) **Range – the range of a sample (or a data set):** It is a measure of the spread or the dispersion of the observations. It is the difference between the largest and the smallest observed value.
- (k) **Weighted arithmetic mean (Weighted Average):** A method of computing a kind of arithmetic mean of a set of numbers in which some elements of the set carry more importance (weight) than others.
- (l) **Standard deviation:** Standard deviation is a measure of the spread or dispersion of a set of data. It is defined as the square root of the variance.

2. PROBLEMS

- (a) Ten patients at a doctor's surgery wait for the following lengths of times to see their doctor.
5 mins, 17 mins, 8 mins, 2 mins, 55 mins, 9 mins, 22 mins, 11 mins, 16 mins, 5 mins,
What are the mean, median and mode for these data? What measure of central tendency would you use here?
Answer: Mean = 15 min, Median = 10 min, Mode = 5 min
- (b) If the mean of 9, 8, 10, x , 12 is 15, find the value of x .
Answer: $x = 36$
- (c) The mean of 40 numbers was found to be 38. Later on, it was detected that a number 56 was misread as 36. Find the correct mean of given numbers.
Answer: Correct Mean = 38.5

- (d) The mean of the heights of 6 boys is 152 cm. If the individual heights of five of them are 151 cm, 153 cm, 155 cm, 149 cm and 154 cm, find the height of the sixth boy.

Answer: Height = 150 cm

- (e) Calculate the mean, median and standard deviation of the following set of data. Birth-weight of ten babies (in kilograms)

2.977, 3.155, 3.920, 3.412, 4.236, 2.593, 3.270, 3.813, 4.042, 3.387

Answer: Mean = 3.4805 kg, Median = 3.3995 kg, Standard deviation = 0.4893 kg

- (f) Which of the following lists has the greatest standard deviation?

i. 98 99 100 101 102

ii. 2 4 6 8 10

iii. 2 10

Answer: (iii)

- (g) In an experiment, 50 people were asked to estimate the length of a rod to the nearest centimetre. The results were recorded.

Length (cm)	20	21	22	23	24	25	26	27	28	29
Frequency	0	4	6	7	9	10	7	5	2	0

- i. What is the range of the data set?

Answer: Range = 9

- ii. Find the value of the median and the value of the mode.

Answer: Median = 24, Mode = 25

- iii. Calculate the mean and the standard deviation.

Answer: Mean = 24.32, Standard Deviation = 1.88

- iv. In a second experiment another 50 people were asked to estimate the length of the same rod. The most common estimate was 23 cm. The range of the estimates was 13 cm. Make two comparisons between the results of the two experiments.

SOURCES:

1. 'Introduction to Descriptive Statistics', Jackie Nicholas, Mathematics Learning Centre, University of Sydney, NSW 2006
2. <http://www.intermath-uga.gatech.edu/dictionary>
3. <http://www.cimt.org.uk/cmmss/S1/Text.pdf>
4. <http://www.cimt.org.uk/cmmss/S2/Text.pdf>