DESCRIBING DATA WITH AVERAGES

MODE

The mode reflects the value of the most frequently occurring score. In other words, A mode is defined as the value that has a higher frequency in a given set of values. It is the value that appears the greatest number of times. Example: In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is 5 since it has appeared in the set twice.

Types of Modes

Bimodal, Trimodal & Multimodal (More than one mode)

- When there are two modes in a data set, then the set is called **bimodal**
- For example, The mode of Set A = {2,2,2,3,4,4,5,5,5} is 2 and 5, because both 2 and 5 is repeated three times in the given set.
- When there are three modes in a data set, then the set is called **trimodal.** For example, the mode of set A = {2,2,2,3,4,4,5,5,5,7,8,8,8} is 2, 5 and 8
- When there are four or more modes in a data set, then the set is called **multimodal**

Example: The following table represents the number of wickets taken by a bowler in 10 matches. Find the mode of the given set of data.

Match No.	1	2	3	4	5	6	7	8	9	10
No. of Wickets	2	1	1	3	2	3	2	2	4	1

It can be seen that 2 wickets were taken by the bowler frequently in different matches. Hence, the mode of the given data is 2.

MEDIAN

The median reflects the middle value when observations are ordered from least to most. The median splits a set of ordered observations into two equal parts, the upper and lower halves.

Finding the Median

- Order scores from least to most.
- If the total number of observations given is odd, then the formula to calculate the median is:

Median = {(n+1)/2}th term / observation

- If the total number of observations is even, then the median formula is:

Median = 1/2[(n/2)th term + {(n/2)+1}th term]

Example 1: Find the median of the following: 4, 17, 77, 25, 22, 23, 92, 82, 40, 24, 14, 12, 67, 23, 29

Solution:

n= 15

When we put those numbers in the order we have:

4, 12, 14, 17, 22, 23, 23, 24, 25, 29, 40, 67, 77, 82, 92

Median = $\{(n+1)/2\}^{th}$ term

= (15+1)/2

=8

The 8^{th} term in the list is 24 The median value of this set of numbers is 24.

MEAN

The mean is found by adding all scores and then dividing by the number of scores. Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

$$Mean = \frac{sum\ of\ all\ scores}{number\ of\ scores}$$

Types of means

- Sample mean
- Population mean

Sample Mean

The sample mean is a central tendency measure. The arithmetic average is computed using samples or random values taken from the population. It is evaluated as the sum of all the sample variables divided by the total number of variables.

SAMPLE MEAN

$$\bar{X} = \frac{\sum X}{n}$$

Population Mean

The population mean can be calculated by the sum of all values in the given data/population divided by a total number of values in the given data/population.

POPULATION MEAN

$$\mu = \frac{\sum X}{N}$$

AVERAGES FOR QUALITATIVE AND RANKED DATA

Mode

The mode always can be used with qualitative data.

Median

The median can be used whenever it is possible to order qualitative data from least to most because the level of measurement is ordinal.

DESCRIBING VARIABILITY

RANGE

The range is the difference between the largest and smallest scores. The range in statistics for a given data set is the difference between the highest and lowest values. For example, if the given data set is {2,5,8,10,3}, then the range will be 10 – 2 = 8.

Example 1: Find the range of given observations: 32, 41, 28, 54, 35, 26, 23, 33, 38, 40.

Solution: Let us first arrange the given values in ascending order.

23, 26, 28, 32, 33, 35, 38, 40, 41, 54

Since 23 is the lowest value and 54 is the highest value, therefore, the range of the observations will be;

Range (X) = Max (X) – Min (X)

= 54 – 23

= 31

VARIANCE

Variance is a measure of how data points differ from the mean. A variance is a measure of how far a set of data (numbers) are spread out from their mean (average) value.

Formula

$$\sigma = \Sigma(x-\mu)2 \text{ or}$$

Variance = (Standard deviation)2= σ2 = > $\sigma\ 2 = \Sigma(x-\mu)2\ /n$

The values of all scores must be added and then divided by the total number of scores.

Example

X = 5, 8, 6, 10, 12, 9, 11, 10, 12, 7

Solution

Mean = sum (x)/ n

n= 10

sum (x) = 5+8+6+10+12+9+11+10+12+ 7

=90

Mean=> μ = 90 / 10 = 9

Deviation from mean

$x-\mu$ = -4, -1, -3, 1, 3, 0, 2,1,3,-2

$(x-\mu)^2$ = 16,1,9,1,9,0,4,1,9,4

$\Sigma(x-\mu)^2$ = 16+1+9+1+9+0+4+1+9+4 =54

$\sigma^2 = \Sigma(x-\mu)^2 /n$

=54/10

= 5.4

STANDARD DEVIATION

The standard deviation, the square root of the mean of all squared deviations from the mean, that is,

Standard deviation = √variance

Standard Deviation:

A rough measure of the average (or standard) amount by which scores deviate.

Standard Deviation: A Measure of Distance The mean is a measure of position, but the standard deviation is a measure of distance (on either side of the mean of the distribution).

Sum of Squares (SS)

Calculating the standard deviation requires that we obtain first a value for the variance. However, calculating the variance requires, in turn, that we obtain the sum of the squared deviation scores. The sum of squared deviation scores or more simply the sum of squares, symbolized by SS

SUM OF SQUARES (*SS*) FOR POPULATION (DEFINITION FORMULA)

$$SS = \Sigma(X - \mu)^2$$

"The sum of squares equals the sum of all squared deviation scores." You can reconstruct this formula by remembering the following three steps:

1. Subtract the population mean, μ, from each original score, X, to obtain a deviation score, $X - \mu$.
2. Square each deviation score, $(X - \mu)^2$, to eliminate negative signs.
3. Sum all squared deviation scores, $\Sigma (X - \mu)^2$.

VARIANCE FOR POPULATION

$$\sigma^2 = \frac{SS}{N}$$

STANDARD DEVIATION FOR POPULATION

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{SS}{N}}$$

Sum of Squares Formulas for Sample

Sample notation can be substituted for population notation in the above two formulas without causing any essential changes:

SUM OF SQUARES (*SS*) FOR SAMPLE (DEFINITION FORMULA)

$$SS = \Sigma(X - \overline{X})^2$$

VARIANCE FOR SAMPLE

$$s^2 = \frac{SS}{n-1}$$

STANDARD DEVIATION FOR SAMPLE

$$s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}}$$

DEGREES OF FREEDOM (*df*)

- Degrees of freedom (df) refers to the number of values that are free to vary, given one or more mathematical restrictions, in a sample being used to estimate a population characteristic.
- Degrees of freedom are the number of independent variables that can be estimated in a statistical analysis. These values of these variables are without constraint, although the values do import restrictions on other variables if the data set is to comply with estimate parameters.
- Degrees of Freedom (df) The number of values free to vary, given one or more mathematical restrictions.

Formula Degree of freedom

df = n-1

INTERQUARTILE RANGE (IQR)

The interquartile range (IQR), is simply the range for the middle 50 percent of the scores. More specifically, the IQR equals the distance between the third quartile (or 75th percentile) and the first quartile (or 25th percentile), that is, after the highest quarter (or top 25 percent) and the lowest quarter (or bottom 25 percent) have been trimmed from the original set of scores. Since most distributions are spread more widely in their extremities than their middle, the IQR tends to be less than half the size of the range.

Simply, The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.