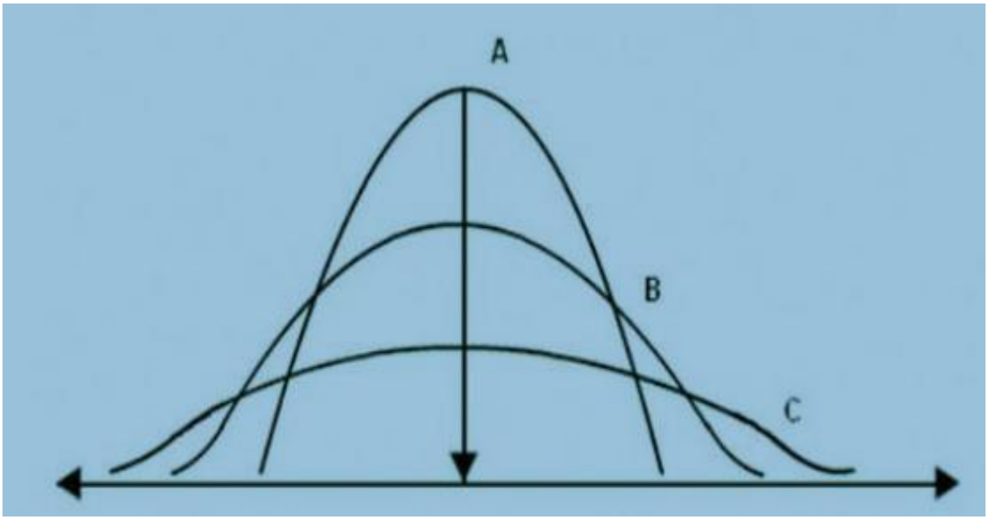
4.1 Measures of Dispersion

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which a numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.



In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms.

4.2 Significance of Measuring Dispersion

1.To determine the reliability of an average.

2.To facilitate comparison.

3.To facilitate control.

4. To facilitate the use of other statistical measures

4.3 Types of Measures of Dispersion

There are two main types of dispersion methods in statistics which are:

• Absolute Measure of Dispersion

• Relative Measure of Dispersion

Absolute Measure of Dispersion A absolute measure of dispersion contains the same unit as the original data set. Absolute dispersion method expresses the variations in terms of the average of deviations of observations. The types of absolute measures of dispersion are: **Range, Variance, Standard Deviation, Quartiles and Quartile Deviation, Mean Deviation.**

.

Relative Measure of Dispersion: The relative measures of depression are used to compare the distribution of two or more data sets. This measure compares values without units. Common relative dispersion methods include:

1.Coefficient of Variation

2.Coefficient of Standard Deviation

3.Coefficient of Mean Deviation.

Range

The simplest measure of variability for a set of data is the range and is defined as the difference between the largest and smallest values in the set.

Range = Largest Value-Smallest Value

R= H - L

Example: Find the range for the sample observations: 13, 23, 11, 17, 25, 18, 14, 24

Solution:

We see that the largest observation is 25 and the smallest observation is 11.

The range is 25-11=14.

Example: Calculate range from following date

Class Limits	1-10	11-20	21-30	31-40	41-50
frequency	7	9	12	8	5

L= Upper boundary of the highest class = 75

S = Lower boundary of the lowest class = 60

Solution:

Range = L – S = 75 – 60 = 15

Variance

The variance is the mean sum of the squares of the deviations of the data from the arithmetic mean of the data. The best estimate of this (take a good statistics class to find out how best is defined) is the sample variance, obtained by taking the sum of the squares of the differences of the data values from the sample mean and dividing this by the number of data points minus one,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 ,$$

where n is the number of data points in the data set, xi is the *i*th data point in the data set x, and $\bar{x}$ is the arithmetic mean of the data set x.

Standard Deviation

The variance has square units, so it is usual to take its square root to obtain the standard deviation,

$$s = \sqrt{\text{variance}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Example: In a summer ecology research program, Jane is asked to count the number of trees per hectare in five different sampling locations in King's Canyon National Park in California. Each sampling location is referred to as a plot, and each plot is a different size. Here are the data she collected:

Plot Size (hectares)	No. of Trees in Plot
1.50	20
2.30	31
1.75	43
3.10	58
2.65	29

Given the data Jane collected, (a) construct the data set that represents the number of trees per hectare for each of the five plots and then calculate the (b) range, (c) variance, and (d) standard deviation of the data set you constructed.

Solution:

(a) For each plot, the number of trees per hectare is

$$\frac{\#\text{ trees in plot}}{\text{plot size}}.$$

For example, the first plot has $20/1.5 = 13.3$ trees/hectare. Thus, the data set that represents the number of trees per hectare for each of the five plots is

$$x = \{13.3, 13.5, 24.6, 18.7, 10.9\}.$$

(b) To calculate the range, we need to know x_{max} and x_{min} (the maximum and minimum values of the data set x). Looking at the data set constructed in (a), $x_{min} = 10.9$ and $x_{max} = 24.6$. Thus,

$$\text{range} = 24.6 - 10.9 = 13.7.$$

(c) Recall that to calculate the variance of a data set, you must first know the arithmetic mean of that data set. For the data set constructed in (a),

$$\bar{x} = \frac{13.3 + 13.5 + 24.6 + 18.7 + 10.9}{5} = 16.2.$$

Then, the variance is

$$\begin{aligned} s^2 &= \frac{1}{5-1}\left[(13.3-16.2)^2 + (13.5-16.2)^2 + (24.6-16.2)^2 \right. \\ &\qquad \left. +(18.7-16.2)^2 + (10.9-16.2)^2\right] \\ &= \frac{1}{4}\left[(-2.9)^2 + (-2.7)^2 + (8.4)^2 + (2.5)^2 + (-5.3)^2\right] \\ &= \frac{1}{4}[8.41 + 7.29 + 70.56 + 6.25 + 28.09] \\ &= \frac{1}{4}[120.6] \\ &= 30.15. \end{aligned}$$

(d) Recall that the standard deviation of a data set is the square root of the variance of that data set. Thus, the standard deviation is

$$s = \sqrt{30.15} = 5.491.$$

Mean Deviation

The mean absolute deviation is defined exactly as the words indicate. The word "deviation" refers to the deviation of each member from the mean of the population. The term "absolute deviation" means the numerical (i.e. positive) value of the deviation, and the "mean absolute deviation" is simple.

$$\textbf{Mean Value (M.D)} = \frac{\sum_{i=1}^{n} |xi - x'|}{n}$$

To calculate mean absolute deviation it is necessary to take following steps:

1.Find x'

2.Find and record the signed differences

3.Find and record the absolute differences

4.Find $\Sigma|xi - x'|ni = 1$.

5.Find the mean absolute deviation.

Example 1: Suppose that sample consists of the observations (21, 17, 13, 25, 9, 19, 6, and 10) Find the mean deviation.

Solution: Perhaps the best manner to display the computations in steps 1, 2, 3, and 4 is to make use of a table composed of three columns:

xi	$xi - x'$	$\lvert xi - x' \rvert$
21	21-15 = 6	6
17	17-15 = 2	2
13	13-15 = -2	2
25	25-15 = 10	10
9	9-15 = -6	6
19	19-15 = 4	4
6	6-15 = -9	9
10	10-15 = -5	5

$$x' = \frac{120}{8} = 15 \qquad \sum 44$$

$$\text{Mean Value (M.D)} = \frac{\sum_{i=1}^{n} |xi - x'|}{n}$$

$$= \frac{44}{8} = 5.5$$

Example 2: Suppose that sample consists of the observations 4, 6, 2, 0, 3, 5 and 8 Find the mean absolute deviation.

Solution:

xi	$xi - x'$	$\lvert xi - x' \rvert$
4	4-4 = 0	0
6	6-4 = 2	2
2	2-4 = -2	2
0	0-4 = -4	4
3	3-4 = -1	1
5	5-4 = 1	1
8	8-4 = 4	4

$$x' = \frac{28}{7} = 4 \qquad \sum 14$$

Mean Value (M.D) = $\frac{\sum_{i=1}^{n} |xi - x'|}{n}$

$$= \frac{14}{7} = 2$$