Measures of Dispersions:

Dispersion refers to the spread or variability in the data.
If the dispersion is small, it indicates high uniformity of the observations in the distribution. Absence of dispersion in the data indicates perfect uniformity. This situation arises when all observations in the distribution are identical.

Purposes of measures of dispersions:
First, it is one of the most important quantities used to characterize a frequency distribution.
Second, it affords a basis of comparison between two or more frequency distributions.

Measures of dispersion include the following: range, mean deviation, variance, and standard deviation.

Range: The simplest measure of dispersion is the range. It is the difference between the largest and the smallest values in a data set.
Range = Largest value – Smallest value
Let $x_1, x_2, \ldots\ldots\ldots, x_n$ are the values of observations in a sample, then range (R) of the variable X is given by:

$$R = X_{\max} - X_{\min}$$

Merits and limitations: Among all the methods of studying variation, range is the simplest to understand and the easiest to compute.
Range cannot tell us anything about the character of the distribution within two extreme observations.

Mean Deviation: The arithmetic mean of the absolute values of the deviations from the arithmetic mean.i.e.mean deviation is obtained by calculating the absolute deviations of each observation from mean and then averaging these deviations by taking arithmetic mean.

Mean deviation for ungroup data:

$$MD(\bar{x}) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Example: The weights of a sample of crates containing books for the bookstore (in pounds) are:
103, 97, 101, 106, 103
Find the mean deviation.
Sol:
$\bar{x}$ =102
The mean deviation is:

$$MD(\bar{x}) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$= \frac{|103-102| + |97-102| + |101-102| + |106-102| + |103-102|}{5}$$

$$= \frac{1+5+1+4+1}{5}$$

=2.4

Merits and limitations: The advantage of the mean deviation is its relative simplicity and it is less affected by the values of extreme observation.
The greatest limitation of this method is that algebraic sings are ignored while taking the deviations of the items.

Variance: The arithmetic mean of the squared deviations from the mean.
Standard deviation: The square root of the variance.

Population variance: let $x_1, x_2, ..., x_n$ are the values of observations. The population variance is define as

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Where:

σ^2 is the population variance
X is the value of an observation in the population
μ is the arithmetic mean of the population
N is the number of observations in the population

Population Standard deviation: The square root of the population variance is the population standard deviation.
That is we can write according to the definition

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Sample variance: Let $x_1, x_2, ..., x_n$ are the values of observations in a sample,
Then sample variance is define as

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

Where:
s^2 is the sample variance
x_i is the value of an observation in the sample
$\overline{x}$ is the mean of the sample
n is the number of observations in the sample.

Sample Standard deviation: The square root of the population variance is the population standard deviation.
That is we can write according to the definition

$$s = \sqrt{s^2}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

Example: The hourly wages earned by a sample of five students are:
\$7, \$5, \$11, \$8, \$6.
Find the sample variance and standard deviation.

Sol: $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{37}{5} = 7.40$

$$s^2 = \frac{\Sigma\left(x_i - \overline{x}\right)^2}{n-1} = \frac{(7-7.4)^2 + ... + (6-7.4)^2}{5-1}$$

$$= \frac{21.2}{5-1} = 5.30$$

$$s = \sqrt{s^2} = \sqrt{5.30} = 2.30.$$

Sample Standard Deviation for Group Data: Let $x_1, x_2, ..., x_n$ are the values of observations in a sample and $f_1, f_2, ..., f_n$ are the class frequencies then SD is defined as

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i\left(x_i - \overline{x}\right)^2}{n-1}}; \quad \text{Here } \overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Where:
s is the symbol for the sample standard deviation
x_i is the mid value of the class.
f_i is the class frequency.
n is the number of observations in the sample.

$\bar{x}$ is the mean of the sample

Example: Find Sample Standard Deviation of Group Data

x_i	f_i
3	2
5	3
7	2
8	2
9	1

Sol:

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
3	2	6	-3	9	18
5	3	15	-1	1	3
7	2	14	1	1	2
8	2	16	2	4	8
9	1	9	3	9	9
Total	10	60	-	-	40

We know that,

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{60}{10} = 6$$

So the sample standard deviation is:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{40}{9}} = \sqrt{4.44} = 2.107$$

Merits of Standard Deviation:
Among all measures of dispersion Standard Deviation is considered superior because it possesses almost all the requisite characteristics of a good measure of dispersion. It has the following merits:

1) It is rigidly defined.

2) It is based on all the observations of the series and hence it is representative.

3) It is amenable to further algebraic treatment.

4) It is least affected by fluctuations of sampling.

Demerits:

1) It is more affected by extreme items.

2) It cannot be exactly calculated for a distribution with open-ended classes.

3) It is relatively difficult to calculate and understand.

Coefficient of Variation:

The Coefficient of Variation expresses the standard deviation as a percentage of the mean. A coefficient of variation is computed as a ratio of the standard deviation of the distribution to the mean of the same distribution.

It is used in such problems where we want to compare the variability of two or more than two series. That series (or group) for which the coefficient of variation is greater is said to be more variable or less consistent, less uniform or less homogeneous. On the other hand, the series for which Coefficient of variation is less is said to be less variable or more consistent, more uniform or more homogeneous.

The population Coefficient of Variation is

$$CV = \frac{\sigma}{\mu} \times 100$$

The sample Coefficient of Variation is

$$CV = \frac{s}{\bar{x}} \times 100$$

Example: Comments on Children in a community

	Height	weight
Mean	40 inch	10 kg
SD	5 inch	2 kg
CV	12.5	20

Sol:

Since the coefficient of variation for weight is greater than that of height, we would tend to conclude that weight has more variability than height in the population.