**Measures of Dispersion**

Dispersion (also called variability, scatter, or spread) is the extent to which a distribution is stretched or squeezed. Common examples of measures of statistical dispersion are the variance, standard deviation, and interquartile range.

Dispersion is contrasted with location or central tendency, and together they are the most used properties of distributions.

A measure of statistical dispersion is a nonnegative real number that is zero if all the data are the same and increases as the data become more diverse.

Most measures of dispersion have the same units as the quantity being measured. In other words, if the measurements are in metres or seconds, so is the measure of dispersion. Examples of dispersion measures include:

**Measure # 1. Range:**

Range is the interval between the highest and the lowest score. Range is a measure of variability or scatteredness of the variants or observations among themselves and does not give an idea about the spread of the observations around some central value.

**Symbolically R = Hs – Ls. Where R = Range**

**Example 1:-**

The scores of ten boys in a test are: 17, 23, 30, 36, 45, 51, 58, 66, 72, 77.

the highest score is 77 and the lowest score is 17.

So the range is the difference between these two scores:

... Range = 77 – 17 = 60

**Example 2 :-** The scores of ten girls in a test are:

48, 49, 51, 52, 55, 57, 50, 59, 61, 62.

Range = 62 – 48 = 14

Here we find that the scores of boys are widely scattered. Thus the scores of boys vary much But the scores of girls do not vary much (of course they vary less). Thus the variability of the scores of boys is more than the variability of the scores of girls.

Computation of Range (Grouped data):

**Example3:-** Find the range of data in following distribution:



**Solution:**

In this case, the upper true limit of the highest class 70-79 is Hs = 79.5 and the lower true limit of the lowest class 20-29 is Ls = 19.5

Therefore, Range R = Hs – Ls

= 79.5 – 19.5 = 60.00

**Advantages:**

1. Range can be calculated quite easily.

2. It is a simplest measure of dispersion.

3. It is computed when we want to make a rough comparison of two or more graphs of variability.

**Limitations:**

1. Range is not based on all the observations of the series. It takes into account only the most extreme cases.

2. It helps us to make only a rough comparison of two or more groups of variability.

3. The range takes into account the two extreme scores in a series.

Thus when N is small or when there are large gaps in the frequency distribution, range as a measure of Scores of Group B – 3, 5, 8, 11, 20, 22, 27, 93

Here range = 93 – 3 = 90.variability is quite unreliable.

**Measure # 2. Quartile Deviation:**

Range is the interval or distance on the scale of measurement which includes 100 percent cases. The limitations of the range are due to its dependence on the two extreme values only.

There are some measures of dispersion which are independent of these two extreme values. Most common of these is the quartile deviation which is based upon the interval containing the middle 50 percent of cases in a given distribution.

Quartile deviation is one-half the scale distance between the third quartile and the first quartile. It is the Semi-interquartile range of a distribution:

Before taking up the quartile deviation, we must know the meaning of quarters and quartiles.



For example a test results 20 scores and these scores are arranged in a descending order. Let us divide the distribution of scores into four equal parts. Each part will present a ‘quarter’. In each quarter there will be 25% (or 1/4th of N) cases.

As scores are arranged in descending order. The top 5 scores will be in the 1st quarter. The next 5 scores will be in the 2nd quarter. The next 5 scores will be in the 3rd quarter. And the lowest 5 scores will be in the 4th quarter.

With a view to having a better study of the composition of a series, it may be necessary to divide it in three, four, six, seven, eight, nine, ten or hundred parts.

Usually, a series is divided in four, ten or hundred parts. One item divides the series in two parts, three items in four parts (quartiles), nine items in ten parts (deciles), and ninety-nine items in hundred parts (percentiles).

There are, thus, three quartiles, nine deciles and ninety-nine percentiles in a series. The second quartile, or 5th decile or the 50th percentile is the median.

A student must clearly distinguish between a quarter and a quartile. Quarter is a range; but quartile is a point on the scale. Quarters are numbered from top to bottom (or from highest score to lowest score), but quartiles are numbered from the bottom to the top.

The Quartile Deviation (Q) is one half the scale distance between the Third Quartile (Q3) and the First Quartile (Q1):



L = Lower limit of the c.i. where Q3 lies, 3N/4= 3/4 of Nor 75% of N.

F = total of all frequencies below ‘L’, fq = Frequency of the c.i. upon which Q3 lies and i = size or length of the c.i.



L = Lower limit of the c.i. where Q1 lies, N/4 = One fourth (or 25%) of N,

F = total of all frequencies below ‘L’, fq = frequency of the c.i. upon which Q1 lies, and i = size or length of c.i.

**Inter-Quartile Range:**

The range between the third quartile and the first quartile is known as the inter-quartile range. Symbolically inter-quartile range = Q3 – Q1.

**Semi-Interquartile Range:**

It is half the distance between the third quartile and the first quartile.

Thus, S I R. = Q3 – Q1/4

Q or Quartile Deviation is otherwise known as semi-interquartile range (or S.I.R.)

Thus, Q = Q3 – Q1/2

If we will compare the formula of Q3 and Q1 with the formula of median the following observations will be clear:

i. In case of Median we use N/2 whereas for Q1 we use N/4 and for Q3 we use 3N/4.

ii. In case of median we use fm to denote the frequency of c.i., upon which median lies; but in case of Q1 and Q3 we use fq to denote the frequency of the c.i. upon which Q1 or Q3 lies.

Example 5:

Find out Q of the following scores 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 . There are 20 scores.

25% of N = 20/4 = 5

Q1 is a point below which 25% of cases lie. In this example, Q1 is a point below which 5 cases lie. From the mere inspection of ordered data it is found that below 24.5 there are 5 cases. Thus Q1 = 24.5

Likewise Q3 is a point below which 75% of cases lie.

75% of N = 3/4 x 20 = 15

We find that below 34.5,15 cases lie

Thus Q3 = 34.5.



**Merits of Q:**

1. It is a more representative and trustworthy measure of variability than the overall range.

2. It is a good index of score density at the middle of the distribution.

3. Quartiles are useful in indicating the skewness of a distribution.

4. Like the median, Q is applicable to open-end distributions.

5. Wherever median is preferred as a measure of central tendency, quartile deviation is preferred as measure of dispersion.

**Limitations of Q:**

1. However, like median, quartile deviation is not amenable to algebraic treatment, as it does not take into consideration all the values of the distribution.

2. It only calculates the third and the first quartile and speaks us about the range. From Q’ we cannot get a true picture about how the scores are dispersed from the central value. That is ‘Q’ does not give us any idea about the composition of scores. ‘Q’ of two series may be equal, yet series may be quite dissimilar in composition.

3. It roughly gives an idea of dispersion.

4. It ignores the scores above the third quartile and the scores below the first quartile. It simply speaks us about the middle 50% of the distribution.

**Uses of Q:**

1. When the median is a measure of a central tendency;

2. When the distribution is incomplete at either end;

3. When there are scattered or extreme score which would disproportionately influence the SD;

4. When the concentration around the median — the middle 50% of cases is of primary interest.

**Measure # 3. Average Deviation (A.D.) or Mean Deviation (M.D.):**

As we have already discussed the range and the ‘Q’ roughly gives us some idea of variability. The range of two series may be the same or the quartile deviation of two series may be same, yet the two series may be dissimilar. Neither the range nor the ‘Q’ speaks of the composition of the series. These two measures do not take into consideration the individual scores.

The method of average deviation or ‘the mean deviation’, as it is called sometimes, tends to remove a serious shortcoming of both methods (Range and ‘Q’). The average deviation is also called the first moment of dispersion and is based on all the items in a series.

Average deviation is the arithmetic mean of the deviations of a series computed from some measure of central tendency (mean, median or mode), all the deviations being considered positive. In other words the average of the deviations of all the values from the arithmetic mean is known as mean deviation or average deviation. (Usually, the deviation is taken from the mean of the distribution.)



Where ∑ is sum total of; X is the score; M is the mean; N is the total number of scores.

And ‘d’ means the deviation of individual scores from the mean.

Computation of Mean Deviation (Ungrouped data):

**Example 8:**

Find mean deviation for the following set of variants:

X = 55, 45, 39, 41, 40, 48, 42, 53, 41, 56

**Solution:**

In order to find mean deviation we first calculate mean for the given set of observations.

The deviations and the absolute deviations are given in Table 4.2:





**Example 9:**

Find the mean deviation for the scores given below:

25, 36, 18, 29, 30, 41, 49, 26, 16, 27

The mean of the above scores was found to be 29.7.

For calculating the mean deviation:

Computation of Mean Deviation (grouped data):

**Example 10:**

Find the mean deviation for the following frequency distribution:



Here, in column 1, we write the c.i. ‘s, in column 2, we write the corresponding frequencies, in column 3, we write the mid-points of the c.i. ‘s which is denoted by ‘X’, in column 4, we write the product of frequencies and mid-points of the c.i. ‘s denoted by X, in column 5, we write the absolute deviations of mid-points of c.i. from the mean which is denoted by |d| and in column 6, we write the product of absolute deviations and frequencies, denoted by |fd|.



**Merits of Mean Deviation:**

1. Mean deviation is the simplest measure of dispersion that takes into account all the values in a given distribution.

2. It is easily comprehensible even by a person not well versed in statistics.

3. It is not very much affected by the value of extreme items.

4. It is the average of the deviations of individual scores from the mean.

**Limitations:**

1. Mean deviation ignores the algebraic signs of the deviations and as such it is not capable of further mathematical treatment. So, it is used only as a descriptive measure of variability.

2. In fact, M.D. is not in common use. It is rarely used in modern statistics and generally dispersion is studied by standard deviation.

**Uses of M.D:**

1. When it is desired to weigh all the deviations according to their size.

2. When it is required to know the extent to which the measures are spread out on either side of the mean.

3. When extreme deviations unduly influence the standard deviation.

**Measure # 4. Standard deviation:-**

the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range.

Standard deviation may be abbreviated SD, A useful property of the standard deviation is that unlike the variance, it is expressed in the same unit as the data. When only a sample of data from a population is available, the term standard deviation of the sample or sample standard deviation can refer to either the above-mentioned quantity as applied to those data, or to a modified quantity that is an unbiased estimate of the population standard deviation (the standard deviation of the entire population).

**“Standard deviation or S.D. is the square root of the mean of the squared deviations of the individual scores from the mean of the distribution.”**

To be more clear, we should note here that in computing the S.D., we square all the deviations separately. Find their sum, divide the sum by total number of scores and then find the square root of the mean of the squared deviations.

 

Where d = deviation of individual scores from the mean;

(Some authors use ‘x’ as the deviation of individual scores from the mean)

∑ = sum total of; N = total number of cases.

There are two ways of computing S.D. for ungrouped data:

(a) Direct method.

(b) Short-cut method.

**Example:-** Find the standard deviation for the scores given below:

X = 12, 15, 10, 8, 11, 13, 18, 10, 14, 9

**This method uses formula (18) for finding S.D. which involves the following steps:**

Step 1:Calculate arithmetic mean of the given data: 

Step 2:

Write the value of the deviation d i.e. X – M against each score in column 2. Here the deviations of scores are to be taken from 12. Now you will find that ∑d or ∑ (X – M) is equal to zero. Think, why is it so? Check it. If this is not so, find out the error in computation and rectify it.

Step 3:

Square the deviations and write the value of d2 against each score in column 3. Find the sum of squared deviations. ∑d2 = 84 .



The required standard deviation is 2.9.

Step 4:

Calculate the mean of the squared deviations and then find out the positive square root for getting the value of standard deviation i.e. σ.

Using formula (19), the Variance will be σ2 = ∑d2/N = 84/10 = 8.4

**Merits of Standard Deviation**

* Squaring the deviations overcomes the drawback of ignoring signs in mean deviations
* Suitable for further mathematical treatment
* Least affected by the fluctuation of the observations
* The standard deviation is zero if all the observations are constant
* Independent of change of origin

**Demerits of Standard Deviation**

* Not easy to calculate
* Difficult to understand for a layman
* Dependent on the change of scale

(**b) Short-cut Method:**

In most of the cases the arithmetic mean of the given data happens to be a fractional value and then the process of taking deviations and squaring them becomes tedious and lime consuming in computation of S.D.

To facilitate computation in such situations, the deviations may be taken from an assumed mean. The adjusted short-cut formula for calculating S.D. will then be,



where,

d = Deviation of the score from an assumed mean, say AM; i.e. d = (X – AM).

d2 = The square of the deviation.∑d = The sum of the deviations.

∑d2 = The sum of the squared deviations.N = No. of the scores or variants.

The computation procedure is clarified in the following example:

**Example 11:**

Find S.D. for the scores given in table 4.5 of X = 12, 15, 10, 8, 11, 13, 18, 10, 14, 9. Use short-cut method.

**Solution:**

Let us take assumed mean AM = 11.

The deviations and squares of deviations needed in formula are given in the following table:



Putting the values from table in formula, the S.D.



The short-cut method gives the same result as we obtained by using direct method in previous example. But short-cut method tends to reduce the calculation work in situations where arithmetic mean is not a whole number.