Measures of Dispersion

The observation deviating from the central value is different in different set of values of character. This property of deviation of the values from the average is called variation or dispersion. The degree of variation is indicated by measure of dispersion. The various measures of the dispersion are as follows:

- Range
- Mean deviation
- Standard deviation
- Quartile deviation

Among these measures of the deviation, range and standard deviation are the commonly used measures of dispersion.

Range

It is the difference between highest and lowest value in the data. If H is the highest and L is the lowest value, then

Range (R) = H-L

For example, calculate the range for the following data.

3,5,7,8, 9, 12, 15, 17, 19, 20, 21, 23, 26. Range (R) = 26-3= 23

Therefore, range of this data is 23.

MERITS OF RANGE

- Range is very simple to understand.
- It is also easy to calculate.

DEMERITS OF RANGE

- It is not suitable for deep analysis.
- It is not suitable in case of extreme values.

Standard Deviation (SD)

It is the most prominently used measure of dispersion. It is denoted by 'SD

DEFINITION

Standard deviation is the positive square root of mean of the squared deviations of values from the arithmetic mean.

Formula for the calculation of the SD from discrete data is:

$$SD = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

Formula for the calculation of the SD from continuous data is:

$$SD = \sqrt{\frac{\sum (x - x)^2 \times f}{n}}$$

Write down the problems as we discussed in class.

NORMAL PROBABILITY CURVE (Z SCORE)

If large values are collected for any character and a frequency table is prepared with small class interval, the frequency curve of this data will give a bell-shaped symmetrical curve, which is known as Gaussian or normal curve. The shape of this curve depends on the mean and SD of data. If the standard deviation (variation) is Very high, the width of the curve is also more. If we move both sides from the midpoint (x = mean), the height of the curve decreases. If the mean and standard deviation are 0 and 1, respectively, then normal curve is called standard normal curve. If the total area under the curve is considered as unity, the normal curve is called normal probability curve.

The normal distribution is

In this formula:

μ= mean distribution (it also may be median and/or more)

δ= standard deviation, and its variance is therefore o?

Properties of Normal Probability Curve

- It is a normal bell-shaped curve.
- Normal probability curve is continuous-type probability curve.
- The curve is symmetrical and asymptotic (i.e. touches at infinity).
- All the measures of central tendency are equal and stable on the highest peak axis, i.e. mean= median = mode.
- The total area under the curve is equal to unity. The quartiles Q_1 and Q_3 are equidistant from the mean μ and $Q_1 = \mu - 0.67\delta$, $Q3 = \mu + 0.67\delta$ approximately.
- The normal curve has two parameters, i.e. mean (μ) and standard deviation (δ).

On the basis of mean (μ) and standard deviation (δ), the area of normal curve is distributed as:

- $\mu \pm 1\ \delta = 68.27\%$
- $\mu \pm 2\ \delta = 95.45\%$
- $\mu \pm 3\ \delta = 99.73\%$
- $\mu \pm 0.67\ \delta = 50\%$
- $\mu \pm 1.96\ \delta = 95\%$
- $\mu \pm 2.58\ \delta = 99\%$

❖ Normal curve is mesokurtic.

Importance of Normal Probability Curve

Normal distribution plays an important role in the sampling theory. It has been found that

Data obtained from biological measurements approximately follow normal distribution.

Binomial and Poisson distribution can be approximated to normal distribution.

For a large sample, any statistics (1.e. sample mean, sample standard deviation) approximately follow normal distribution, and as such it can be studied with the help of a normal curve.

Normal curve is used to find confidence limits of the population parameters.

Normal distribution also forms the basis for the tests of significance.

This relationship or association between two quantitatively measured or continuous variables is called Correlation.

The extent or degree of relationship between two sets of figures is measured in terms of another parameter called correlation coefficient. It is denoted by r.

Correlation determines the relationship between two variables but it does not prove that one particular variable alone causes change in other ($-1 \leq r \leq 1$).

Types of Correlation Coefficient

1. Perfect positive correlation ($r = +1$)

2. Perfect negative correlation ($r =-1$)

3. Moderately positive correlation ($0 < r < 1$)

4. Moderately negative correlation ($-1 < r < 1$)

5. Absolutely no correlation ($r = 0$)

Perfect Positive Correlation ($r = +1$)

X is directly proportional to Y

Both variables rise and fall in same proportion.

For example, designation and salary, height and weight

Perfect Negative Correlation ($r =-1$)

Here X and Y are inversely proportional to each other.

$r =-1$

X is inversely proportional to Y.

For example, leniency of teacher and discipline among students.

Moderately Positive Correlation ($0 < r <$)

For example, the correlation between infant mortality rate (IMR) and overcrowding, and correlation between the plasma value and circulating albumin in grams.

Moderately Negative Correlation ($-1 < r <$)

For example, age and vital capacity in adults, IMR, and income.

Absolutely No Correlation r = 0

For example, habit of smoking and housing facility of person.

Methods of Computing the Correlation

The correlation between variables can be studied by following ways:

- Karl Pearson's correlation coefficient
- Spearman's rank correlation coefficient

Karl Pearson's Correlation Coefficient

It is used to measure the degree of linear relationship between two variables. It is also called product moment correlation coefficient. It is denoted by r. It can be computed using the following

$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 \sum (y - \overline{y})^2}}$$

Spearman's Rank Correlation Coefficient

It is a method of finding the correlation between two variables by taking their ranks. This method of finding correlation is especially useful in dealing with qualitative data. It can be used when the actual magnitude of characteristics under consideration is not known, but relative position or rank of the magnitude is known. It is denoted by p.

There are two cases for calculating rank correlation. In the first case, there is no tie of allotted ranks, i.e. no two numbers are same; in the second case there is tie in between allotted ranks.

Case 1-Ranks correlation with no tie among allotted ranks: In this case, any one of the values in x or y series is not repeated. So we can use following steps for calculating rank correlation coefficient with no tie among allotted ranks. Then we can calculate Spearman's rank correlation coefficient through steps given below:

- Rank one the highest value, rank two to next highest value, and so on.
- Rank x series values and y series values separately.
- Calculate the difference of ranks in each pair of values or ($d = R_x - R_y$)
- Calculate sum of the square of the difference of ranks ($\sum d^2$).

Finally calculate the correlation coefficient using following formula:

$$\rho = 1 - \frac{6 \sum d^2}{n\ (n^2 - 1)}$$