PGDISM - SEMESTER - 2,PMIR Department,
P.U. [e-Content]

Course / Paper Code - 06 (201)

Safety Statistics And Accident Inspection

Unit-3

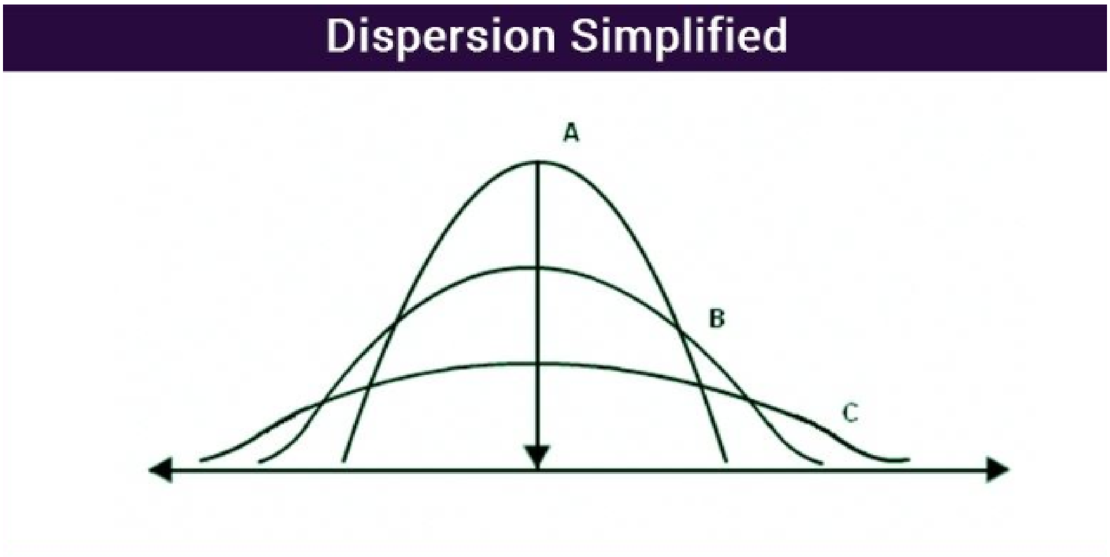
MR. RAJEEV KUMAR
[GUEST FACULTY]
PMIR DEPARTMENT
PATNA UNIVERSITY
Mobile No. +91-6287858781
Email:-rajeevk.patna@gmail.com

CONTENTS

Dispersion and Measures of Dispersion

What is Dispersion in Statistics?

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.



Dispersion and Measures of Dispersion in Statistics

Measures of Dispersion

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogenous the data is. In simple terms, it shows how squeezed or scattered the variable is.

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

An 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 like standard or means deviations. It includes range,standard deviation, quartile deviation, etc.

The types of absolute measures of dispersion are:

1. Range: It is simply the difference between the maximum value and the minimum value given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6
2. Variance: Deduct the mean from each data in the set then squaring each of them and adding each square and finally dividing them by the total no of values in the data set is the variance. Variance $(\sigma^2)=\sum(X-\mu)^2/N$
3. Standard Deviation: The square root of the variance is known as the standard deviation i.e. S.D. = $\sqrt{\sigma}$.
4. Quartiles and Quartile Deviation: The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.
5. Mean and Mean Deviation: The average of numbers is known as the mean and the arithmetic mean of the absolute deviations of the observation from a measure of central tendency is known as the 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 Range
2. Coefficient of Variation
3. Coefficient of Standard Deviation
4. Coefficient of Quartile Deviation
5. Coefficient of Mean Deviation

Range

A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the two extreme observations then

$$\text{Range} = L - S$$

R=Range,L=Largest Value,S=Smallest Value

Coefficient of Range=L-S/L+S

Merits of Range

- It is the simplest of the measure of dispersion
- Easy to calculate
- Easy to understand
- Independent of change of origin

Demerits of Range

- It is based on two extreme observations. Hence, get affected by fluctuations
- A range is not a reliable measure of dispersion
- Dependent on change of scale

Quartile Deviation

The quartiles divide a data set into quarters. The first quartile, (Q_1) is the middle number between the smallest number and the median of the data. The second quartile, (Q_2) is the median of the data set. The third quartile, (Q_3) is the middle number between the median and the largest number.

Quartile deviation or semi-inter-quartile deviation is

$$Q = \frac{1}{2} \times (Q_3 - Q1)$$

Merits of Quartile Deviation

- All the drawbacks of Range are overcome by quartile deviation
- It uses half of the data
- Independent of change of origin
- The best measure of dispersion for open-end classification

Demerits of Quartile Deviation

- It ignores 50% of the data
- Dependent on change of scale
- Not a reliable measure of dispersion

Mean Deviation

Mean deviation is the arithmetic mean of the absolute deviations of the observations from a measure of central tendency. If $x_1, x_2, \dots, x_n$ are the set of observation, then the mean deviation of x about the average A (mean, median, or mode) is

Mean deviation from average

$$A = 1/n \left[\sum_i |x_i - A|\right]$$

For a grouped frequency, it is calculated as:

Mean deviation from average

$$A = 1/N \left[\sum_i f_i |x_i - A|\right], \qquad (N = \sum f_i)$$

Here, x_i and f_i are respectively the mid value and the frequency of the i_{th} class interval.

Merits of Mean Deviation

- Based on all observations
- It provides a minimum value when the deviations are taken from the median
- Independent of change of origin

Demerits of Mean Deviation

- Not easily understandable
- Its calculation is not easy and time-consuming
- Dependent on the change of scale
- Ignorance of negative sign creates artificiality and becomes useless for further mathematical treatment

Standard Deviation

A standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by a Greek letter sigma, σ. It is also referred to as root mean square deviation. The standard deviation is given as

$$\sigma = [(\Sigma_i (y_i - \bar{y}) ⁄ n]^{½} = [(\Sigma_i y_i^2 ⁄ n) - \bar{y}^2]^{½}$$

For a grouped frequency distribution, it is

$$\sigma = [(\Sigma_i f_i (y_i - \bar{y}) ⁄ N]^{½} = [(\Sigma_i f_i y_i^2 ⁄ n) - \bar{y}^2]^{½}$$

The square of the standard deviation is the variance. It is also a measure of dispersion.

$\sigma^2 = [(\Sigma_i (y_i - \bar{y}) / n]^{½} = [(\Sigma_i y_i^2 ⁄ n) - \bar{y}^2]$

For a grouped frequency distribution, it is

$\sigma^2 = [(\Sigma_i f_i (y_i - \bar{y}) ⁄ N]^{½} = [(\Sigma_i f_i x_i^2 ⁄ n) - \bar{y}^2]$.

If instead of a mean, we choose any other arbitrary number, say A, the standard deviation becomes the root mean deviation.

Variance of the Combined Series

If σ_1, σ_2 are two standard deviations of two series of sizes n_1 and n_2 with means $\bar{y}_1$ and $\bar{y}_2$. The variance of the two series of sizes $n_1 + n_2$ is:

$\sigma^2 = (1/ n_1 + n_2) \div [n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2)]$

where, $d_1 = \bar{y}_1 - \bar{y}$, $d_2 = \bar{y}_2 - \bar{y}$, and $\bar{y} = (n_1 \bar{y}_1 + n_2 \bar{y}_2) \div (n_1 + n_2)$.

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

Coefficient of Dispersion

Whenever we want to compare the variability of the two series which differ widely in their averages. Also, when the unit of measurement is different. We need to calculate the coefficients of dispersion along with the measure of dispersion. The coefficients of dispersion (C.D.) based on different measures of dispersion are

- Based on Range = $(X_{max} - X_{min}) / (X_{max} + X_{min})$.
- C.D. based on quartile deviation = $(Q_3 - Q1) / (Q_3 + Q1)$.
- Based on mean deviation = Mean deviation/average from which it is calculated.
- For Standard deviation = S.D. ⁄ Mean

Coefficient of Variation

100 times the coefficient of dispersion based on standard deviation is the coefficient of variation (C.V.).

$$C.V. = 100 \times (S.D. / Mean) = (\sigma/\bar{y}) \times 100.$$