10 CHAPTER Statistics

AIEEE Syllabus

Calculation of mean, median and mode of grouped and ungrouped data. Calculation of standard deviation, variance and mean deviation for grouped and ungrouped data

Statistics is a branch of mathematics which deals with collection, grouping, analysis of the data and then to derive the results which are representative of the data collected.

Single value of the variable describing the data is called an average. Average lies generally in the middle of data so such values are called measures of central tendency. Common measures are

MEAN, MEDIAN AND MODE

The degree to which the values in a set of data are spread around an average are called dispersion of variation. The different measures of dispersion are range, quartile deviation, mean deviation, standard deviation and variance.

Statistics finds wide application in finance, economics, various surveys, planning, demography and various other analysis. All future planning of a country depends upon statistics of growth or fall in population, industrial production, crop production etc.

Mean

Also called arithmetic mean or simply average.

Unclassified data : The mean of n observations x_1, x_2 .... x_n is defined as

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots\dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Grouped data : Let x_1, x_2 .... x_n be n observation with corresponding frequencies as f_1, f_2, ..., f_n respectively, then

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \dots\dots + f_nx_n}{f_1 + f_2 + \dots\dots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Shortcut method : Let the assumed mean (usually the middle term of data) be A. The deviation from assumed mean, $d_i = (x_i - A)$ for each term

$$\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

Step deviation method : When the data is grouped in equal class intervals with class size as h, then to simplify the calculation each $d_i = (x_i - A)$ is divided by h.

$$\bar{x} = A + \frac{\Sigma f_i d_i}{\Sigma f_i} \times h$$ here A = assumed mean and $d_i = \frac{x_i - A}{h}$

Weighted mean : If w_1, w_2.........., w_n are the weights assigned to the values x_1, x_2.........., w_n respectively then the weighted mean is given by

$$\text{Weight mean} = \frac{w_1x_1 + w_2x_2 + \dots\dots\dots + w_nx_n}{w_1 + w_2 + \dots\dots\dots + w_n}$$

Combined mean : Let individual means of two sets of data are $\bar{x}_1$ and $\bar{x}_2$ with the size of data as n_1 and n_2 respectively, then

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Properties of Mean

1. If each of the n given observations are increased (decreased) by same quantity a then the new mean of data is also increased (decreased) by a.
2. If each of the n given observation are multiplied (divided) by same quantity $a(a \neq 0)$, the mean of data is also multiplied (divided) by a.
3. The sum of the deviations of individual values from mean is always zero.

$$\sum_{i=1}^{n} f_i(x_i - \bar{x}) = 0$$

4. The sum of squares of deviation from mean is least.

$$\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 \text{ is least}$$

Median

It is the middle most or central value of the variate in a set of observations, when the observations are arranged either in ascending or in descending order of their magnitudes. Median divides this arranged series in two equal parts.

Median of individual series : If there are total n observations arranged in ascending or descending order, then

(i) If n is odd, median = value of $\left(\frac{n+1}{2}\right)^{th}$ observation.

(ii) If n is even, median = mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations.

Median of a discreet series : Prepare a cumulative frequency after arranging the data in ascending or descending order.

(i) If n is odd, then median = size of the $\left(\frac{n+1}{2}\right)^{th}$ term

(ii) If n is even, then median = $\dfrac{\text{size of}\left(\frac{n}{2}\right)^{th}\text{term} + \text{size of}\left(\frac{n}{2}+1\right)^{th}\text{term}}{2}$

Mode

It is the value of variate occurring most frequently or the variate having maximum frequency.

Mode of individual series is the value which is repeated maximum number of times.

Mode of discrete series is the value corresponding to maximum frequency.

Mode of continuous series : Modal class is the class having maximum frequency.

$$\text{Mode} = l + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}} \times h$$

l = lower limit of modal class

h = width of modal class

f_{m-1} = frequency of class preceding modal class

f_m = frequency of modal class

f_{m+1} = frequency of class succeeding modal class

If mode falls in other class than modal class then, mode = $l + \frac{f_{m+1}}{f_{m-1} + f_{m+1}} \times h$

When mean median and mode coincide, such distribution is called symmetrical distribution. Otherwise

Mode = 3 median – 2 mean for moderately skew distribution

VARIANCE

If $x_1, x_2, x_2, \ldots\ldots. x_n$ be n observations then the variance of there is given by

$$\text{Var(x)} = \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

STANDARD DEVIATION (σ)

The positive square root of the mean of squares of deviations from the mean of data.

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

Standard deviation for a frequency distribution is $\text{S.D.} = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}}$

***Note :** When data is grouped in classes then respective mid-points are taken as x_i.*

Coefficient of S.D. (C.V.)

Two or more series can be compared for variability by C.V. (the coefficient of variation)

$$\text{C.V.} = \frac{\sigma}{\bar{x}} \times 100 \text{ percent}$$

Variance : Square of S.D. (σ^2) is called variances of distribution.

Direct method for S.D.

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

Shortcut method for S.D.

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$ for ungrouped data, $d = (x_i - A)$. A is the assumed mean.

$$\sigma = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$$

Step-deviation method

$$\sigma = h\sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2}$$ where $d' = \frac{x_i - A}{h}$

SOLVED EXAMPLES

Example 1 :

The sum of squares of deviation for n observations taken from their mean x is y. The coefficient of variation is

(1) $\sqrt{\frac{y}{nx}} \times 100\%$ *(2)* $\frac{1}{nx}\sqrt{y} \times 100\%$

(3) $\frac{1}{x}\sqrt{\frac{y}{n}} \times 100\%$ *(4)* $\frac{1}{n}\sqrt{\frac{y}{n}} \times 100\%$

Solution :

$$\sigma = \sqrt{\frac{y}{n}}$$

$$\therefore \text{C.V.} = \frac{\sigma}{x} \times 100 = \frac{1}{x}\sqrt{\frac{y}{n}} \times 100\%$$

Hence answer (3) is correct

Example 2 :

The mean of six observations is 7 and their variance is $\frac{25}{3}$*. If four observations are 5, 6, 8, 9, then the median of all observations are*

Solution :

$\bar{x} = 7, N = 6, \Sigma\frac{(x-\bar{x})^2}{N} = \frac{25}{3}$ and x_1 and x_2 are other two observations, then

$x_1 + x_2 = 42 - 5 - 6 - 8 - 9 = 14$ ... (i)

$\Sigma(x - \bar{x})^2 = 50$

$(5 - 7)^2 + (6 - 7)^2 + (8 - 7)^2 + (9 - 7)^2 + (x_1 - 7)^2 + (x_2 - 7)^2 = 50$

$(x_1 - 7)^2 + (x_2 - 7)^2 = 40$

$$(x_1 - 7)^2 + (14 - x_1 - 7)^2 = 40$$

$$(x_1 - 7)^2 + (7 - x_1)^2 = 40$$

$$(x_1 - 7)^2 = 20$$

$$x_1 - 7 = \pm\sqrt{20}$$

$$x_1 = 7 \pm \sqrt{20}$$

$$x_1 = 7 + \sqrt{20}$$

$$x_2 = 7 - \sqrt{20}$$

Example 3 :

If the data is less scattered about its mean then

(1) Its C.V. is large

(2) Its σ is large

(3) Its variance is small

(4) Its mean is small

Solution :

Less scattered data about mean, indicates less variation from mean, hence all measures of dispersion would be small. So variance is small is the correct statement.

Example 4 :

A data given below has its S.D. = 2, then

$$a = \underbrace{-a, -a, -a, \ldots\ldots\ldots, -a}_{n \text{ times}} \quad \underbrace{a, a, a, \ldots\ldots\ldots, a}_{n \text{ times}}$$

Solution :

$$\bar{x} = 0$$

$$\therefore \sigma = 2 = \sqrt{\frac{n(-a-0)^2 + n(a-0)^2}{2n}}$$

$$\Rightarrow 2 = \sqrt{\frac{2na^2}{2n}}$$

$$\Rightarrow a^2 = 4$$

$$\Rightarrow a = \pm 2$$