Basic Statistics

Lesson 3: Measures of Dispersion: Standard Deviation and Coefficient of Variation

Dispersion:

The degree to which numerical values tend to spread about an average value is called variation or dispersion of the data.

The measures of dispersion can be of two types: Absolute and Relative. We shall discuss about Standard Deviation as the absolute measure and Coefficient of Variation as the relative measure of dispersion.

Standard Deviation:

(a) Simple Series/Ungrouped Data:

Let us consider that n observations x_i $(i = 1, 2, \ldots.., n)$ are given. Then the standard deviation is given by

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

Here $\bar{x}$ is the A.M of the distribution. In another way, we can write

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

(b) Discrete/Simple Frequency Distribution

Let us consider that n observations are given with n frequencies f_i $(i = 1, 2, \ldots.., n)$. Then the standard deviation is given by

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

In the alternative way, the standard deviation for this case is given by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - \bar{x}^2$$

(c) Continuous/Grouped Frequency Distribution:

Let us consider that n classes or intervals are given with n frequencies f_i $(i = 1, 2, \dots, n)$. Then the standard deviation is given by

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

Hence the alternative way of writing the standard deviation will be similar to the previous case as

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - \bar{x}^2$$

Coefficient of Variation:

This is the percentage variation in the mean where standard deviation has been considered as the total variation in the mean. It is mathematically given by

$$\text{Coefficient of variation (C.V)} = 100 \times \frac{\sigma}{\bar{x}}\ \%$$

For comparing the variability of two distributions, we calculate the C.V's of the two distributions. The distribution having greater C.V is more variable and the distribution with less C.V is said to be more consistent (or homogeneous).

Problems:

Ex.1. Calculate the standard deviation for the following frequency distribution:

Age (in years)	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of members	3	61	132	153	140	51	2

Solution: Let us make the table for the calculation of the standard deviation:

Age (in years)	20-30	30-40	40-50	50-60	60-70	70-80	80-90	Total
No. of members	3	61	132	153	140	51	2	542
Class mark	25	35	45	55	65	75	85	-
$x_i f_i$	75	2135	5940	8415	9100	3825	170	29660
${x_i}^2$	625	1225	2025	3025	4225	5625	7225	-
$f_i {x_i}^2$	1875	74725	267300	462825	591500	286875	14450	1699550

The A.M is given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i = \frac{29660}{542} = 54.72 \text{ years}$$

The standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - \bar{x}^2} = \sqrt{\frac{1699550}{542} - (54.72)^2} = \sqrt{3135.7 - 2994.28}$$
$$= \sqrt{141.42} = 11.89 \text{ years.}$$

Ans.

Ex.2. For a group of 200 candidates, the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on, it was discovered that the scores 43 and 35 were misunderstood by 34 and 53 respectively. Find the corrected mean and standard deviation.

Solution: We are given n = 200 for the distribution x_i (say). We also know that the incorrect mean $\bar{x} = 40$ and the incorrect s.d. σ = 15. From these expression of mean, we can write

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{or} \quad \sum_{i=1}^{n} x_i = n \times \bar{x} = 200 \times 40 = 8000$$

This is the incorrect sum of the scores. In order to rectify the sum, we add the correct score and subtract the incorrect score from this sum. That gives us

$$\sum_{i=1}^{n} x_i \Big|_{\text{corrected}} = 8000 - 34 - 53 + 43 + 35 = 7991$$

Hence the corrected mean of the scores = $\frac{7991}{200} = 39.955$

Similarly, from the expression of the incorrect s.d, we have

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2 \quad \text{or, } 15^2 = \frac{1}{200}\sum_{i=1}^{n} {x_i}^2 - 40^2 \quad \text{or, } \sum_{i=1}^{n} {x_i}^2 = 365000$$

This is again the incorrect sum of the squares of the scores. Then the corrected form of this sum will be

$$\sum_{i=1}^{n} {x_i}^2 {}_{|\text{corrected}} = 365000 - 34^2 - 53^2 + 43^2 + 35^2 = 364109$$

Therefore, the corrected s.d. of the scores = $\sqrt{\frac{364109}{200} - (39.955)^2} = 14.97$

Ans.

Ex.3. Two sets of variables x_i and y_i are related by the relation $y_i = 10 - 3x_i$ (i = 1, 2, …, n). If the standard deviation of the x distribution is 4, what is the standard deviation of the y distribution?

Solution: We have, $y_i = 10 - 3x_i$. Then $\sum_{i=1}^{n} y_i = 10n - 3\sum_{i=1}^{n} x_i$. This gives,

$$\bar{y} = 10 - 3\bar{x}$$

Now,

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(10 - 3x_i) - (10 - 3\bar{x})]^2}$$

$$= \sqrt{\frac{1}{n}\,9\sum_{i=1}^{n}(x_i - \bar{x})^2} = 3\sigma_x = 12$$

Ans.

Ex.4. The monthly average electricity charges is Rs. 2460 and standard deviation is Rs.120, whereas the monthly average direct wages is Rs. 42000 and standard deviation is Rs. 1200. State which one is more variable?

Solution: We can write,

Coefficient of variation (electricity charges) = $\frac{120\times100}{2460} = 4.87\ \%$

Coefficient of variation (direct wages) = $\frac{1200\times100}{42000} = 2.86\ \%$

Since the C.V of electricity charges is more, therefore the distribution of electricity charges is more variable.

Ans.

Ex.5. In a company, the coefficient of variation of wages of male and female workers were 55% and 70% respectively. The standard deviations were Rs. 22 and Rs. 15.40 respectively. Calculate the combined average wages for all workers if 80% of the workers were male.

Solution: Let $\overline{x_1}$: average wages of male workers

$\overline{x_2}$: average wages of female workers

σ_1 : standard deviation of wages for male workers = 22 Rs.

σ_2 : standard deviation of wages for female workers = 15.4 Rs.

It is also given that

$$\frac{\sigma_1}{\overline{x_1}}\times 100 = 55 \rightarrow \overline{x_1} = 40 \text{ Rs.} \quad \text{and} \quad \frac{\sigma_2}{\overline{x_2}}\times 100 = 70 \rightarrow \overline{x_2} = 22 \text{ Rs.}$$

If we now assume that the total number of workers in the company is n, then the number of male and female workers are given by $\frac{80n}{100}$ and $\frac{20n}{100}$.

Then the combined average wages of all the workers will be given by

$$\bar{x} = \frac{\frac{80n}{100}\times 40 + \frac{20n}{100}\times 22}{n} = 36.4 \text{ Rs.}$$

Ans.