Bachelor of Architecture

Mathematics

Lecture 15

In this lecture we are going to see the Standard deviation and variance, Regression and correlation and then the summary.

Standard deviation and Variance:

The most important and the most powerful measure of dispersion is the standard deviation (S.D) or represented by the symbol (σ). It is computed as the square roots of the mean of the squares of the differences of the variant values form their mean. Thus standard deviation (S.D) is given by,

$$\sigma = \sqrt{\left[\frac{\sum f_i (x_i - x)^2}{N}\right]}$$

Where N is the total no of the frequency $\sum f_i$. If however the deviation is measured from any other value say A instead of x it is called the root-mean-square deviation.

Variance:

The square of the standard deviation is known as the variance,

$$\sigma^2 = \sqrt{\left[\frac{\sum f_i (x_i - x)^2}{N}\right]}$$

Next let us move to the calculation of the standard deviation method.

Calculation of S.D:

The change of origin and the change of scale considerably reduces the labour in the calculation of standard deviation. The formula for the computation of σ are as follows,

Short-cut method:

$$\sigma = \sqrt{\left[\frac{\sum f_i d_i^2}{\sum f_i} - \left(\frac{\sum f_i d_i}{\sum f_i}\right)^2\right]}$$

Step deviation method:

$$\sigma = h\sqrt{\left[\frac{\sum f_i d'^2_i}{\sum f_i} - \left(\frac{\sum f_i d'_i}{\sum f_i}\right)^2\right]}$$

Where $d_i = x_i - A \,\&\, d'_i = \dfrac{x_i - A}{h}$ being the assumed mean and h is the equal class interval.

Observance:

The root mean square deviation is least when measured from the mean. The root mean square deviation is given by

$$s^2 = \frac{\sum f_i {d_i}^2}{\sum f_i}$$ and

$$\frac{\sum f_i d_i}{\sum f_i} = \left[A + \frac{\sum f_i {d_i}^2}{\sum f_i}\right] - A = x - A$$

From the short cut method of standard deviation $s^2 = \sigma^2 + (x - A)$. This shows that $s^2 = \sigma^2$. This occurs when A=x.

We have learned two methods to solve this standard deviation, one is short-cut method and the other one is step deviation method. Then next thing is Co-efficient of variation.

Co-efficient of Variation:

The ratio of the standard deviation to the mean i.e., $\sigma\sqrt{x}$ is known as the co-efficient of variation. As this is a ratio having no dimension it is used for comparing the variations between the two groups with different means.

Example 1:

The following are scores of two batsmen A and B in a series of innings.

A	12	115	6	73	7	19	119	36	84	29
B	47	12	16	42	4	51	37	48	13	0

Who is the better score getter and who is more consistent?

Solution:

Let us first we calculate the values as the x denotes score of A and y that of B taking 51 as the origin we proceed the following table.

X	d(=x-15)	d^2	Y	$\delta(=y-51)$	δ^2
12	-39	1521	47	-4	16
115	64	4096	12	-39	1521
6	-45	2025	16	-35	1225
73	22	484	42	-9	81
7	-44	1936	4	-47	2209
19	-32	1024	51	0	0
1119	68	4624	37	-14	196

36	-15	225	48	-3	9
84	33	1089	13	-38	1444
29	-22	484	0	-51	2601
Total	-10	17508		-240	9302

The arithmetic mean for the batsmen A is,

$$\bar{x} = 51 + \frac{\sum d}{n}$$

$$= 51 + \frac{10}{10}$$

$$= 50$$

Then the standard deviation for batsmen A will be,

$$\sigma_1 = \sqrt{\left[\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2\right]}$$

$$= \sqrt{1750.8 - (-1)^2}$$

$$== 41.8$$

Co-efficient of batmen A is,

$$= \frac{\sigma_1}{x} \times 100 = \frac{41.8}{50} \times 100$$

$$= 83.6$$

The arithmetic mean for the batsmen B is,

$$\bar{y} = 51 + \frac{\sum \delta}{n}$$

$$=51-\frac{240}{10}$$

$$=27$$

Then the standard deviation for batsmen B will be,

$$\sigma_2=\sqrt{\left[\frac{\sum\delta^2}{n}-\left(\frac{\sum\delta}{n}\right)^2\right]}$$

$$=\sqrt{930.2-(-24)^2}$$

$$=18.8$$

Co-efficient of batmen B is, $=69.6$

So the co-efficient of batman B is lesser that the co-efficient of batman A, hence the batman B is more consistent than batman A. Then coming to the S.D for batman A the S.D is 41.8 whereas for batman B it is 18.8 so the batman A is better scorer that the batman B.

Correlation:

There are many phenomena where the changes in one variable are related to the changes in the other variable. For instance the yield of a crop varies with the amount of rainfall. Such a simultaneous variation that is when the changes in one variable are associated by changes in the other is called Correlation.

This correlation is of two types. That is positive correlation and the negative correlation. Then the data involved in the two variables is called bivariate population. If an increase (or decrease) in the value of one variable corresponding to an increase (or decrease) in the other correlation is positive.

If an increase (or decrease) in the values of one variable corresponds to an decrease (or increase) in the other the correlation is negative.

Co-efficient of correlation:

The numerical measure of correlation is called the co-efficient of correlation and is given by,

$$r = \frac{\sum xy}{n\sigma_x \sigma_y}$$

Where,

x is deviation from the mean x=x-x,

y is deviation from the mean y=y-y.

σ_x is the S.D of x-Series.

σ_y is the S.D of y-Series.

n is the number of values of two variables.

Lines of regression:

In the scatter diagram the dots tends to cluster along a well-defined direction which suggest a linear relationship between the variables x and y. Such a line of best-fit for the given distribution of dots is called the line of regression.

The best possible mean value of y for each values of x is the line of regression of y on x. The best possible mean value of x for each value of y is the line of regression of x on y.

The regression co-efficient of x is,

$$x = r\frac{\sigma_x}{\sigma_y}.$$

The regression co-efficient of y is,

$$y = r\frac{\sigma_x}{\sigma_y}$$

Summary:

In this lecture we have learnt that the standard deviation is the most powerful measure of dispersion. The square of the standard deviation is known as the variance. And the co-efficient of variation is the ratio of the standard deviation to the mean.

Then we learnt about the correlation that it is the phenomena connecting one variable with other. Line of best fit for the given distribution of dots is called line of regression. After listening to this lecture you can answer the following questions.

Questions:

1. Define standard deviation. Give its formula.
2. Explain the correlation concept.
3. What do you mean regression?