Quadrant II – Transcript and Related Materials

Programme: S. Y. B. Com

Subject: Mathematics

Paper Code: CAG102

Paper Title: Business Statistics-II (GE 5)

Unit: I Correlation and Regression analysis

Module Name: Karl Pearson's coefficient of correlation: direct method and changing the scale method

Module No: 05

Name of the Presenter: Asmita Atchut Gaude

Notes:

Karl Pearson's Coefficient of Correlation

Direct Method

As we have already studied that Karl Pearson's coefficient is used to find the relation between the two variables. It tells us about whether it is positive correlation or negative correlation.

So let us look at the direct method formula

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

Where, N is the No. of observations.

PROBLEM:

Calculate the Karl Pearson's coefficient of correlation for the following data using direct method

X	35	40	45	50	55
Y	70	60	65	45	45

Solution: -

X	Y	X^2	Y^2	XY
35	70	1225	4900	2450
40	60	1600	3600	2400
45	65	2025	4225	2925
50	45	2500	2025	2250
55	45	3025	2025	2475
$\sum X = 225$	$\sum Y = 285$	$\sum X^2 = 10375$	$\sum Y^2 = 16775$	$\sum XY = 12500$

Now Substitute all the values in the direct method formula.

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{5(12500) - (225)(285)}{\sqrt{5(10375) - (225)^2}\sqrt{5(16775) - (285)^2}}$$

$$r = \frac{62500 - 64125}{\sqrt{51875 - 50625}\sqrt{83875 - 81225}}$$

$$r = \frac{-1625}{\sqrt{1250}\sqrt{2650}}$$

$$r = \frac{-1625}{(35.36)(51.48)}$$

$$r = \frac{-1625}{1820.33}$$

$$r = -0.89$$

Hence the Karl Pearson's coefficient of correlation is -0.89 which shows negative correlation.

Changing the scale Method

Change of scale method is used to simplify the calculations. Here, we use same formula of r i.e.

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

But the difference is, here we are changing original value of x to $X = \frac{x - x_0}{c}$ and y to $Y = \frac{y - y_0}{d}$, where N is the no. of observations, x_0 and y_0 are assumed mean or middle value of x and y respectively, c and d are common difference/ divisor of x and y respectively.

PROBLEM:

Calculate the Karl Pearson's coefficient of correlation for the following data using changing the scale method

x	35	40	45	50	55
y	70	60	65	45	45

x	y	$X = \frac{x - x_0}{c} = \frac{x - 45}{5}$	$Y = \frac{y - y_0}{d} = \frac{y - 60}{5}$	X^2	Y^2	XY
35	70	-2	2	4	4	-4
40	60	-1	0	1	0	0
45	65	0	1	0	1	0
50	45	1	-3	1	9	-3
55	45	2	-3	4	9	-6
		$\sum X = 0$	$\sum Y = -3$	$\sum X^2 = 10$	$\sum Y^2 = 23$	$\sum XY = -13$

Substitute all the values in the formula

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{5(-13) - (0)(-3)}{\sqrt{5(10) - (0)^2}\sqrt{5(23) - (-3)^2}}$$

$$r = \frac{-65 - 0}{\sqrt{50 - 0}\sqrt{115 - 9}}$$

$$r = \frac{-65}{\sqrt{50}\sqrt{106}}$$

$$r = \frac{-65}{(7.07)(10.3)}$$

$$r = \frac{-65}{72.82}$$

$$r = -0.89$$

Hence the Karl Pearson's coefficient of correlation is -0.89 which shows negative correlation.

We will get same answer in both the methods.