Frequently Asked Questions

1. Write the regression equation of X on Y
   Answer:
   If X is dependent variable and Y is independent variable, then equation for regression line X on Y is given by,
   $(x-\overline{x}) = b_{xy}(y-\overline{y})$

2. Write the regression equation of Y on X
   Answer:
   In a bivariate data, if X is independent variable and Y is dependent variable, the equation of regression line, Y on X is given by $(y-\overline{y}) = b_{yx}(x-\overline{x})$

3. Write the formula to find the regression coefficient in terms of coefficient of correlation.
   Answer: The formula for finding the regression coefficients in terms of correlation coefficient is,
   $b_{xy} = \frac{r.\sigma_x}{\sigma_y}$ or $b_{yx} = \frac{r.\sigma_y}{\sigma_x}$

4. Write the formula for coefficient of correlation in terms of regression coefficients.

   Answer: The formula for coefficient of correlation in terms of regression coefficients is,
   $$r = \pm\sqrt{b_{xy}b_{yx}}$$

5. Write the formula to find regression coefficients when we have raw data and tabulated data.

   Answer:
   When we have raw data,
   $b_{xy} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$ and $b_{yx} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$
   For tabulated data,
   $b_{xy} = \frac{N\Sigma\Sigma fxy - \Sigma fx\Sigma fy}{N\Sigma fy^2 - (\Sigma fy)^2}$ and $b_{yx} = \frac{N\Sigma\Sigma fxy - \Sigma fx\Sigma fy}{N\Sigma fx^2 - (\Sigma fx)^2}$

6. How do you find mean and r form the given regression equations?
   Answer:
   Given the regression equations, X on Y and Y on X, we can find the means by solving the two simultaneous equations.
   We can identify the equations X on Y and Y on X, and can find the regression coefficients and in turn we can find correlation coefficient between the two variables X and Y

7. Given the two regression equations, how to identify the equation which is X on Y and which is Y on X?

Answer:

If when we consider arbitrarily the given equations as X on Y and Y on X and we get r greater than 1 then interchange the equations. i.e., the equation X on Y should be considered as Y on X and the equation Y on X should be considered as X on Y

8. How do you obtain the graph of the regression equations?

Answer:

To obtain the graph of the regression equation,

- We find two points on the regression line
- These points are plotted on a graph sheet
- The points are joined by straight line

The resulting line is regression line of x on y

9. In a bivariate data on x and y the means are respectively 15 and 27 and the variances are respectively 25 and 9. The coefficient of correlation is -0.3. Then what would be the expected value of y when x=8

Answer:

Here we have given $\bar{x}$=15, $\bar{y}$ =27,

σ_x=√25=5, σ_y=√9=3 and r=-0.3

$$b_{yx} = \frac{r.\sigma_y}{\sigma_x} = \frac{-0.3\times 3}{5} = -0.18$$

Therefore

$$(y-\bar{y}) = b_{yx}(x-\bar{x})$$

y=-0.18x+29.7

When x = 8, the estimate of y is

y=-0.18(8) +29.7=28.26

10. A survey of children revealed the following information regarding IQ of child and age of the mother at the time of giving birth to child.

	IQ of child	Age of mother
Mean	98	28 years
Standard deviation	2	4 years

Coefficient of correlation r=-0.24

Estimate IQ of a child whose mother was aged 47 at the time of giving birth to the child.

Answer:

Let x denote IQ of a child and y denote the age of the mother at the time of giving birth to the child.

Given $\bar{x}$ = 98, $\bar{y}$ = 28, r=-0.24, SD(X) =2, SD(y) =4. We assume that IQ depends on age of the mother. And so for the estimation of x given y we write down regression equation of x on y.

$$b_{xy} = \frac{r.\sigma_x}{\sigma_y} = \frac{-0.24\times 2}{4} = -0.12$$

The regression equation of x on y is,

$$(x-\bar{x}) = b_{xy}(y-\bar{y})$$

$$(x-98) = -0.12(y-28)$$

x=-0.12y+101.36

Estimate of IQ of a child when y=47 is

x=-0.12(47) +101.36=95.56

11. The following are heights of 8 persons and one each of their sons. From the data, estimate the height of a person whose father is 150 cms. Tall

Height of father	164	176	178	184	175	167	173	180
Height of son	168	174	175	181	173	166	173	179

Answer:

Let x and y respectively denote the heights of fathers and the sons. Then the value of y corresponding to x=150 has to be estimated. For this, regression of y on x should be found and the estimation should be made.

$\bar{x} = \frac{\Sigma x}{n} = \frac{1397}{8} = 174.63$, $\bar{y} = \frac{\Sigma y}{n} = \frac{1389}{8} = 173.63$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{8 \times 242776 - 1397 \times 1389}{8 \times 244255 - (1397)^2} = 0.7302$$

Thus regression of y on x is,

$(y - \bar{y}) = b_{yx}(x - \bar{x})$

Y - 173.63 = 0.7302 (x-174.63)

Y = 0.7302x + 46.12

This is the regression equation of y on x. to estimate the value of y when x =150, in this equation the value of x=150 is substituted.

Thus, the estimate of son's height is, y=0.7302(150) +46.12=155.65 cms.

12. In a bivariate data, $\sum x=20$, $\sum y=400$, $\sum x^2=196$, $\sum y^2=46500$, $\sum xy=850$ and n=10. Estimate the value of y corresponding to the value of x=5

Answer:

Here regression equation of y on x should be found.

$\bar{x} = \frac{\Sigma x}{n} = \frac{30}{10} = 3$, $\bar{y} = \frac{\Sigma y}{n} = \frac{400}{10} = 40$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{10 \times 850 - 30 \times 40}{10 \times 196 - (30)^2} = -3.3$$

The regression equation of y on x is

$(y - \bar{y}) = b_{yx}(x - \bar{x})$

(y-40)=-3.3(x-3)

y=-3.3x+49.9

The estimate of y when x=5 is

y=-3.3(5) +49.9=33.4

13. If b_{xy}=-1/3 and b_{yx}=-3/4, find the value of r_{xy}

Answer:

The coefficient of correlation is numerically equal to the geometric mean of the regression coefficients. Therefore the coefficient of correlation is

$$r = \pm\sqrt{b_{xy}b_{yx}} = \pm\sqrt{\left(-\frac{1}{3}\right)\left(-\frac{3}{4}\right)} = -0.5$$

Since the regression coefficients are negative, the coefficient of correlation is also negative.

Thus, coefficient of correlation is r=-0.5

14. In a bivariate data, the two regression equations are 3x+4y=1 and 3x +y=4. Find the means and coefficient of correlation.

Answer

Since the regression line intersect at $(\overline{x}, \overline{y})$ the values x and y are obtained by solving the regression equations for x and y. The equations are

3x+4y=1 ------------ (1)

3x +y=4------------- (2)

On subtracting we get,

3y=-3 or y=-1

Putting y=-1 in equation 1 we get,

3x-4=1 or 3x=5 or x=5/3

Thus $\overline{X}$ =5/3 and $\overline{Y}$ =-1

Now let us consider the two equations as y on x and x on y.

Let regression equation of y on x be 3x+4y=1

and regression equation of x on y be 3x +y=4

The equations are re-written as y= (-¾) x + ¼

And x= (-⅓) y+ (4/3)

From above equations, b_{yx}= (- ¾) and b_{xy}= (-⅓)

We know that r is geometric mean of regression coefficients. i.e.

$$r = \pm\sqrt{b_{xy}b_{yx}} = \pm\sqrt{(-1/3)\times(-3/4)} = -1/2$$

r is negative, since both the regression coefficients are negative.

15. In a bivariate data, the regression coefficients are 7.3 and 0.11. Find the coefficient of correlation

Answer:

The coefficient of correlation is numerically equal to the geometric mean of the regression coefficients. Therefore the coefficient of correlation is

$$r = \pm\sqrt{b_{xy}b_{yx}} = \pm\sqrt{7.3\times 0.11} = 0.8961$$

Since the regression coefficients are positive, the coefficient of correlation is also positive.

Thus coefficient of correlation is r=0.8961