Online-Class 22-3-2021

Probability, Statistics and Reliability (MAT3003)

SLOT: B21 + B22 + B23

MODULE - 3

Topic: Regression Lines

Reference: Probability and Statistics for Engineers & Scientists
by R.E.Walpole, Pearson (2012).

Equation of the Regression Lines

Equation of the regression line of Y on X is assumed as:

$$y = ax + b, \qquad \text{or} \qquad y - \bar{y} = r\frac{\sigma_Y}{\sigma_X}(x - \bar{x})$$

where

$\dfrac{r_{XY}\sigma_Y}{\sigma_X}$ *is called the regression coefficient of Y on X and denoted by* b_1 *or* b_{YX}.

Equation of the Regression Lines

Equation of the regression line of X on Y is assumed as:

$$x = cy + d, \qquad \text{or} \qquad x - \bar{x} = r\frac{\sigma_X}{\sigma_Y}(y - \bar{y})$$

where

$\dfrac{r_{XY}\sigma_X}{\sigma_Y}$ *is called the regression coefficient of X on Y and denoted by* b_2 *or* b_{XY}.

Clearly, $b_1 b_2 = r_{XY}^2$, *i.e.,* r_{XY} *is the geometric mean of* b_1 *and* b_2.

$\therefore \qquad r_{XY} = \pm\sqrt{b_1\, b_2}$

The sign of r_{XY} *is the same as that of* b_1 *or* b_2, *as* $b_1 = r_{xy}\dfrac{\sigma_Y}{\sigma_X}$ *and* $b_2 = r_{XY}\dfrac{\sigma_Y}{\sigma_X}$

have the same sign as r_{XY} ($\because$ σ_X *and* σ_Y *are positive).*

Also, $$\frac{b_1}{b_2} = \frac{\sigma_Y^2}{\sigma_X^2}$$

Question 1

Find the angle between the two lines of regression. Deduce the condition for the two lines to be (a) at right angles, and (b) coincident.

Solution

The equations of the regression lines

are $$y - \bar{y} = r\frac{\sigma_Y}{\sigma_X}(x - \bar{x}) \qquad (1)$$

and $$x - \bar{x} = r\frac{\sigma_X}{\sigma_Y}(y - \bar{y}) \qquad (2)$$

$$\text{Slope of line (1)} = r\frac{\sigma_Y}{\sigma_X} = m_1\text{, say}$$

$$\text{Slope of line (2)} = \frac{\sigma_Y}{r\sigma_X} = m_2\text{, say}$$

If θ is the acute angle between the two lines, then

$$\tan\,\theta = \frac{|\,m_1 - m_2\,|}{1 + m_1 m_2}$$

$$= \frac{\left| r\,\dfrac{\sigma_Y}{\sigma_X} - \dfrac{\sigma_Y}{r\,\sigma_X} \right|}{1 + \dfrac{\sigma_Y^2}{\sigma_X^2}}$$

$$= \frac{\left| r - \dfrac{1}{\cdot r} \right| \sigma_X\,\sigma_Y}{\sigma_X^2 + \sigma_Y^2} = \frac{(1 - r^2)}{|\,r\,|}\,\frac{\sigma_X\,\sigma_Y}{\sigma_X^2 + \sigma_Y^2}$$

Reference: Probability and Statistics for Engineers & Scientists
by R.E.Walpole, Pearson (2012).

The two regression lines are at angles when $\theta = \frac{\pi}{2}$, i.e., tan $\theta = \infty$

i.e., $r = 0$

$\therefore$ when the linear correlation between X and Y is zero, the two lines of regression will be at right angles.

The two regression lines are coincident, when $\theta = 0$, i.e., when tan $\theta = 0$

i.e., when $r = \pm 1$.

$\therefore$ when the correlation between X and Y is perfect, the two regression lines will coincide.

Remarks:

When there is perfect linear correlation between X and Y, viz., when $r_{XY} = \pm 1$, *the two regression lines coincide.*

The point of intersection of the two regression lines is clearly the point whose co-ordinates are $(\overline{x}, \overline{y})$.

When there is no linear correlation between X and Y, viz., when $r_{XY} = 0$, *the equations of the regression lines become* $y = \overline{y}$ *and* $x = \overline{x}$, *which are at right angles.*

Question 2

From the equation of the two regression lines, $4x+3y+7=0$ and $3x+4y+8=0$, find the mean of x and y.

Solution

The equations of the two lines of regression are

$4x + 3y + 7 = 0$ ..... (i)

$3x + 4y + 8 = 0$ ..... (ii)

Solving (i) and (ii) simultaneously, we get,

$x = -4/7$ and $y = -11/7$

Hence $\overline{x} = -4/7$ and $\overline{y} = -11/7$

Question 3

In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible: Variance of X equals to 1. The regression equations are

$$3x + 2y = 26 \text{ and } 6x + y = 31.$$

Determine the following

(a) Mean values of X and Y?

(b) Correlation coefficient between X and Y?

(c) Standard deviation of Y?

Solution

(a) Since the lines of regression intersect at $(\bar{x}, \bar{y})$ we have $3\bar{x} + 2\bar{y} = 26$ and $6\bar{x} + \bar{y} = 31$
Solving these equations, we get $\bar{x} = 4$ and $\bar{y} = 7$.

(b) Which of the two equations is the regression equation of Y on X and which one is the regression equation of X on Y are not known.

Let us tentatively assume that the first equation is the regression line of X on Y and the second equation is the regression line of Y on X. Based on this assumption, the first equation can be re-written as

$$x = -\frac{2}{3}y + \frac{26}{3} \qquad (1)$$

and the other as $\quad y = -6x + 31 \qquad (2)$

Then $b_{XY} = -\frac{2}{3}$ and $b_{YX} = -6$

$\therefore \quad r_{XY}^2 = b_{XY} \times b_{YX} = 4$

$\therefore \quad r_{XY} = -2$, which is absurd.

Hence, our tentative assumption is wrong.

$\therefore$ the first equation is the regression line of Y on X and re-written as

$$y = -\frac{3}{2}x + 13 \tag{3}$$

The second equation is the regression line of X on Y and re-written as

$$x = -\frac{1}{6}y + \frac{31}{6} \tag{4}$$

Hence, the correct $b_{YX} = -\dfrac{3}{2}$ and the correct $b_{XY} = -\dfrac{1}{6}$

$$\therefore \qquad r_{XY}^2 = b_{YX} \cdot b_{XY} = \frac{1}{4}$$

$$\therefore \qquad r_{XY} = -\frac{1}{2} \qquad (\because \text{ both } b_{YX} \text{ and } b_{XY} \text{ are negative})$$

(c) Now $$\frac{\sigma_Y^2}{\sigma_X^2} = \frac{b_{YX}}{b_{XY}} = \frac{-\frac{3}{2}}{-\frac{1}{6}} = 9$$

$\therefore$ $$\sigma_Y^2 = 9 \times \sigma_x{}^2 = 9$$

$\therefore$ $$\sigma_Y = 3$$

Question 4 (For Students)

The equations of lines of regression are given by $x + 2y - 5 = 0$ and $2x + 3y - 8 = 0$ and variance of X is 12. Compute the values of $\bar{x}$, $\bar{y}$, σ_Y^2 and r_{XY}.

Solution

Ans. $\bar{x} = 1,\ \bar{y} = 2,\ \sigma_y^2 = 4$, and $r = -0.866$

Question 5 (For Students)

The equations of the two lines of regression are:

$$3x + 2y - 26 = 0 \text{ and } 6x + y - 31 = 0.$$

Find (a) the means of X and Y ,

(b) Correlation coefficient between X and Y.

Solution

Ans. $\bar{x} = 4,\ \bar{y} = 7,$ and $r = -0.5$

Practice Questions

THANK YOU