# Correlation Analysis-

Dr. Uday Kashid.

## Karl pearson's coefficient of correlation [or] product moment coefficient of correlation:- ($r$)

(i) $$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{N \cdot \sigma_x \cdot \sigma_y}$$

where $\bar{x} = \frac{\Sigma x}{N}$, $\bar{y} = \frac{\Sigma y}{N}$

$\sigma_x^2$ = variance of $x = \frac{\Sigma(x-\bar{x})^2}{N}$

$\sigma_y^2$ = variance of $y = \frac{\Sigma(y-\bar{y})^2}{N}$.

and $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{V(x)}$ = S.D of $x$

$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{V(y)}$ = S.D of $y$.

**or**

(ii) $$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\,\sqrt{\Sigma(y-\bar{y})^2}}$$

[Use when $\bar{x}$ & $\bar{y}$ are in integer Form]

**or** Direct Method:-

(iii) $$r = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{N\Sigma(x^2)-[\Sigma(x)]^2}\,\sqrt{N\Sigma(y^2)-(\Sigma y)^2}}$$

**or**

(iv). $$r = \frac{\Sigma(xy) - N(\bar{x})(\bar{y})}{\sqrt{\Sigma(x^2)-N(\bar{x})^2} \cdot \sqrt{\Sigma(y^2)-N(\bar{y})^2}}$$

**or**

(v) Assumed mean Formula:

If A & B are assumed means of x & y respectively then

$$r = \frac{\Sigma(x-A)(y-B) - \frac{\Sigma(x-A) \cdot \Sigma(y-B)}{N}}{\sqrt{\Sigma[(x-A)^2] - \frac{[\Sigma(x-A)]^2}{N}}\,\sqrt{\Sigma[(y-B)^2] - \frac{[\Sigma(y-B)]^2}{N}}}$$

Note: (i) $-1 \leq r \leq 1$

(ii) If $r = 1$ (perfect +ve correlation) (iii) If $r = -1$ (perfect (-ve) correlation)

(11)

| x | 57 | 42 | 38 | 42 | 45 | 42 | 44 | 40 | 46 | 44 | 43 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 10 | 26 | 41 | 29 | 27 | 27 | 19 | 18 | 19 | 31 | 29 | 33 |

Ans $r = -0.74$

Direct calculating method of $r$

$$r = \frac{\Sigma xy - N(\bar{x})(\bar{y})}{N\sqrt{\frac{\Sigma x^2}{N} - (\bar{x})^2}\sqrt{\frac{\Sigma y^2}{N} - (\bar{y})^2}}$$

Spearman's Rank correlation coefficient : (R)

$$R = 1 - \left[\frac{6\Sigma d_i^2}{n^3 - n}\right]$$

where $d_i = (R_1 - R_2)$ = [difference betn Ranks of $i^{th}$ item]

$$-1 \leq R \leq 1$$

Find Rank correlation coeff.

(1)

| Industry | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rank (profit) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Rank (capital) | 13 | 16 | 14 | 15 | 10 | 12 | 4 | 11 | 5 | 9 | 8 | 3 | 1 | 6 | 7 | 2 |

Ans $\boxed{R = -0.8176}$

(For repeated values)

(2)

| x | 85 | 74 | 85 | 50 | 65 | 78 | 74 | 60 | 74 | 90 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 78 | 91 | 78 | 58 | 60 | 72 | 80 | 55 | 68 | 70 |

$R = 0.45$

(3) Growth of employment in Lakhs in India betn 1988–1995

| Year | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|---|---|---|
| Public sector | 98 | 101 | 104 | 107 | 113 | 120 | 125 | 128 |
| Private sector | 65 | 65 | 67 | 68 | 68 | 69 | 68 | 68 |

Note: For repeated ~~Rect~~ items Rank $= \frac{i^{th} + j^{th}}{2}$ for Two repeated items

If $m_1, m_2, \ldots$ are repeated nos of items, Then spearman Rank correlation coeff $= R = 1 - \left\{\frac{6\left[\Sigma d_i^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots\right]}{n^3 - n}\right\}$

Dec-21 MCQ. (4) Rank correlation coefficient of the following data is [ans $R = 1$]

| x | 23 | 25 | 27 | 29 | 31 | 33 |
|---|---|---|---|---|---|---|
| y | 43 | 45 | 47 | 49 | 51 | 53 |

Ans:– $R = 1 - \frac{6\Sigma(R_1 - R_2)^2}{N^3 - N} = 1 - 0$

For Not repeated (5)

| x | 35 | 38 | 43 | 30 | 54 | 68 | 70 | 92 | 44 | 56 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 51 | 37 | 48 | 62 | 93 | 73 | 56 | 72 | 70 | 92 |

Find Rank correlation coeff. $R = 0.59$ $N = 10$

For Not repeated (6)

| x | 105 | 110 | 112 | 108 | 111 | 116 | 120 | 104 | 115 | 125 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 39 | 41 | 45 | 38 | 48 | 58 | 60 | 35 | 54 | 69 |

$R = 0.9636$

For repeated (7)

| x | 98 | 101 | 104 | 107 | 113 | 120 | 125 | 128 |
|---|---|---|---|---|---|---|---|---|
| y | 65 | 65 | 67 | 68 | 68 | 69 | 68 | 68 |

$R = 0.98$

(8) If $r = 0.4$, $cov(x, y) = 1.6$, $\sigma_y^2 = 25$, find $\sigma_x$. (Ans = 0.8)

# REGRESSION



Regression line y on x is given by

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

Uday Kashid.

But $r\frac{\sigma_y}{\sigma_x}$ is called slope of line & called as Regression coeff. y on x & denoted by byx.

$\therefore \quad byx = r\frac{\sigma_y}{\sigma_x}$

Hence $\quad y-\bar{y} = byx(x-\bar{x})$

similarly Regression line x on y is

$$x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$x-\bar{x} = bxy(y-\bar{y}) \quad \text{where } bxy = r\frac{\sigma_x}{\sigma_y}$$

(i) $\quad byx \cdot bxy = r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y}$

$$byx \cdot bxy = r^2$$

$$\therefore \quad r = \sqrt{byx \cdot bxy}$$

(ii) $\quad byx = r\frac{\sigma_y}{\sigma_x} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}\frac{\sigma_y}{\sigma_x} = \frac{cov(x,y)}{\sigma_x^2}$

$$\Rightarrow byx = \frac{\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\cancel{N}}}{\frac{\Sigma(x-\bar{x})^2}{\cancel{N}}} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma[(x-\bar{x})^2]}$$

$\therefore$ similarly

$$bxy = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma[(y-\bar{y})^2]}$$

Q(1) Show that Arithmetic mean of coeff. of regression is greater than or equal to coeff. of correlation.

$\rightarrow$ for $\sigma_x$ & $\sigma_y$ we can write

$$(\sigma_x - \sigma_y)^2 \geq 0$$

$$\sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y \geq 0$$

$$\therefore \quad \sigma_x^2 + \sigma_y^2 \geq 2\sigma_x\sigma_y$$

$$\frac{\sigma_x^2 + \sigma_y^2}{2\sigma_x\sigma_y} \geq 1$$

$$\frac{1}{2}\left[\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x}\right] \geq 1$$

multiply by r

$$\frac{1}{2}\left[r\frac{\sigma_x}{\sigma_y} + r\frac{\sigma_y}{\sigma_x}\right] \geq r$$

$$\frac{1}{2}[bxy + byx] \geq r \quad \rightarrow \text{proved.}$$

# Regression lines

प्रा. उदय काशिद

Q 1) A panel of two Judges Mrs. Madhuri (Dipika) Dixit-Nene & Mrs. ~~Kajol~~ (Anushka, Rashmika Mandana) graded dramatic performances by independently awarding marks as follows.

| Performance NO | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Marks By ~~Madhuri~~ Dipika | 36 | 32 | 34 | 31 | 32 | 32 | 34 |
| Marks by Anushka | 35 | 33 | 31 | 30 | 34 | 32 | 36 |

The eighth performance however which Judge Anushka could not attend, got 38 marks by Judge Dipika. If Judge Anushka had also been present, how many marks would she be expected to have awarded to the eight performance.

→ method I: $\bar{x} = \frac{\Sigma x}{N} = 33$, $\bar{y} = \frac{\Sigma y}{N} = 33$, $b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = 0.72$

$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow$ at $x = 38$, $\rightarrow y = 36.6 \simeq 37$

method II: $y = a + bx \rightarrow$ eqn of required line.

$\Sigma y = aN + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$

Q 2) Find the coeff. of regression lines and hence eqn of regression lines for the following data.

| X | 78 | 36 | 98 | 25 | 75 | 82 | 90 | 62 | 65 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 84 | 51 | 91 | 60 | 68 | 62 | 86 | 58 | 53 | 47 |

Estimate value of Y when $X = 50$ and value of X when $Y = 90$.

Dec-21

Ex 3) obtain the equations of line of regression for the following data.

| x | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 |
|---|---|---|---|---|---|---|---|---|
| y | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |

→ $n = 8$, $\Sigma x = 544$, $\Sigma y = 552$, $\Sigma xy = 37560$, $\Sigma x^2 = 37028$, $\Sigma y^2 = 38132$

(I) Regression line y on x is $y = ax + b \Rightarrow \Sigma y = a\Sigma x + bn$, $\Sigma 1 = n$

$\Sigma xy = a\Sigma x^2 + b\Sigma x$ → By solving $a = 0.66$ & $b = 23.67$

$\boxed{y = (0.66)x + 23.67}$

(II) Regression line x on y ⇒ $x = ay + b \Rightarrow \Sigma x = a\Sigma y + bn$

$\Sigma xy = a\Sigma y^2 + b\Sigma y \Rightarrow$ By solving $a = 0.55$ & $b = 30.36$.

$\boxed{x = (0.55)y + 30.36}$

Dec-21

Ex 4) Fit the straight line of the form $y = ax + b$ to the following data.

| x | 1 | 3 | 5 | 7 | 8 | 10 |
|---|---|---|---|---|---|---|
| y | 8 | 12 | 15 | 17 | 18 | 20 |

उत्तर $n = 6$, $\Sigma y = a\Sigma x + b$ & $\Sigma xy = a\Sigma x^2 + b\Sigma x$

$\Sigma x = 34$, $\Sigma y = 90$, $\Sigma xy = 582$, $\Sigma x^2 = 248$

$a = 1.30$, $b = 7.63$

Regression lines

Ex ① Find (i) the lines of regression (ii) coeff of correlation for

(A)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| x | ~~30~~ | 10 | 12 | 13 | 16 | 17 | 20 | 25 |
| y | ~~67~~ | 19 | 22 | 24 | 27 | 29 | 33 | 37 |

Ans:- $y = 0.8x + 13.23$
$x =$
$r =$

Ex ② The heights in cms of fathers (x) and of the eldest sons (y) are given below. Then find the lines of regression. Also estimate the height of the eldest son if the height of the father is 172 cms. and the height of father if the height of son is 173 cm. Also find coeff of correlation (r).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| x | 165 | 160 | 170 | 163 | 173 | 158 | 178 | 168 | 173 | 170 | 175 | 180 |
| y | 173 | 168 | 173 | 165 | 175 | 168 | 173 | 165 | 180 | 170 | 173 | 178 |

{Ans:- (i) $y = (1.016)x - 5.123$ (ii) $x = (0.476)y + 98.98$ (iii) 169.97
(iv) 173.45 (v) $r = 0.696$}

---

## Fitting of parabola [second degree curve]

Note ⇒
We know second degree curve i.e. (fitting of parabola) equation is $y = ax^2 + bx + c$ —(I)

$\Sigma y = a\Sigma x^2 + b\Sigma x + c\Sigma 1$ But $\Sigma 1 = N$.

$\Sigma y = a\Sigma x^2 + b\Sigma x + cN$ —(II)

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$ —(III)

$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$ —(IV)

By solving eq (II), (III) & (IV) obtain values of a, b, & c.

Ex ① By the method of least square method find the best values of a, b, c in the second degree curve is. $y = ax^2 + bx + c$ to fit the following data

| | | | | | |
|---|---|---|---|---|---|
| x | -2 | -1 | 0 | 1 | 2 |
| y | -3.150 | -1.390 | 0.620 | 2.880 | 5.378 |

Ans:- $y = (0.1233)x^2 + (2.1326)x + (0.621)$

Ex ② Fit a second degree curve (parabola) to the following data and estimate y when $x = 80$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x | 10 | 20 | 30 | 40 | 50 | 60 | 70 |
| y | 20 | 60 | 70 | 80 | 90 | 100 | 100 |

Ans:- Put $u = (x-40)/10$
$v = y/10$
$y = (-0.2381)x^2 + (1.2943)x + (8.381)$

at $x = 80$, ⇒ $y = 94.286$

Ex ③ Fit the second degree curve (parabola) to the following data

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| year (x) | 1965 | 1966 | 1967 | 1968 | 1969 | 1970 | 1971 | 1972 |
| profit in crores Rs (y) | 125 | 140 | 165 | 195 | 200 | 215 | 220 | 230 |

Ans:- put $X = (x-1968.5)/2$, $Y = y$, $Y = (-0.40)x^2 + (7.68)x + 194.68$