

Practice Problems: Correlation and Regression

① Given:

$$r_{X,Y} = 0.28; \text{Cov}(X,Y) = 7.6; \text{Var}(X) = 9$$

Find s.d. of Y

Ans: 9.048

② Given:

$$r_{X,Y} = 0.8; \sum(xy) = 60; \sigma_y^2 = \text{s.d. of } Y = 2.5;$$

$$\sum(x^2) = 90$$

Find n [in usual notations]

Hint:

$$r = \frac{\text{Cov}(X,Y)}{\text{s.d. of } X \times \text{s.d. of } Y}$$

Ans: n = 10

$$\text{Cov}(X,Y) = \frac{1}{n} \sum(xy); \text{s.d. of } X = \sqrt{\frac{1}{n} \sum(x^2)}$$

$$\text{s.d. of } Y = \sqrt{\frac{1}{n} \sum(y^2)}$$

$$0.8 = \frac{\frac{1}{n} \sum(xy)}{\sqrt{\frac{1}{n} \sum(x^2)} \sqrt{\frac{1}{n} \sum(y^2)}}$$

$$\Rightarrow \sqrt{n} = \frac{(2.5) \sqrt{\frac{1}{n} \sum(y^2)}}{0.8} = \sqrt{10}$$

$$\Rightarrow n = 10$$

③ Given:

$$n = 50; \sum X = 70; \sum Y = 80; \sum(X^2) = 130; \sum(Y^2) = 140$$

$\sum(XY) = 120$, find $r_{X,Y}$

④ Given:

$$n = 10; \sum X = 140; \sum Y = 150; \sum(X-10)^2 = 180;$$

$$\sum(Y-15)^2 = 215; \sum(X-10)(Y-15) = 60$$

Find $r_{X,Y}$

Ans: 0.915

(Hint: Let $A = 10; B = 15; dx = X-A; dy = Y-B$)

$$\text{Note: } \sum(dx) = \sum(X-A) = \sum X - nA; \sum(dy) = \sum(Y-B) = \sum Y - nB$$

⑤ Given:

$$r_{X,Y} = 0.3; n = 20; \bar{X} = 15; \bar{Y} = 20; \text{s.d. of } X = 4$$

$$\text{s.d. of } Y = 5$$

At the time of calculation, one entry 27 was wrongly

taken as 17 in case of X series and as 35 in case of Y-series.

Find the correct correlation co-efficient.

Ans: 0.1832 (Verify)

⑥ Given: Reg. line of X on Y is

$$X = 4Y + 5 \text{ and } Y \text{ on } X \text{ is } Y = kX + 4$$

Show that (i) $0 \leq k \leq \frac{1}{4}$ (ii) Find the means of X & Y and $r_{X,Y}$

$$\text{When } k = \frac{1}{16}$$

$$\text{Ans: } 28, \frac{23}{4}, \frac{1}{2}$$

(P.T.O.)

Solution

⑨ Regn. Line of Y on X is [it passes through $(4, 0)$ & $(-14, 3)$
 (x_1, y_1) & (x_2, y_2)]

$$\frac{Y - y_1}{y_1 - y_2} = \frac{X - x_1}{x_1 - x_2} \Rightarrow \boxed{X + 6Y - 4 = 0} \rightarrow \textcircled{I}$$

$$\text{or } Y = \frac{1}{6}(4 - X)$$

$$\Rightarrow \boxed{b_{yx} = -\frac{1}{6}}$$

Given: $r = -\frac{1}{2} \Rightarrow r^2 = \frac{1}{4} = b_{xy} \cdot b_{yx}$

$$\Rightarrow \boxed{b_{xy} = -\frac{3}{2}}$$

Regn. Line of X on Y is $X - \bar{X} = b_{xy}(Y - \bar{Y})$

\Rightarrow slope of the line is

$$m = \frac{1}{b_{xy}} \quad \left[\text{if } Y = mX + c \right. \\ \left. m \rightarrow \text{slope} \right]$$

Given: It passes through $(1, -1) \Rightarrow$ its eqn is
 (x_3, y_3)

$$Y - y_3 = m(X - x_3)$$

$$\Rightarrow Y + 1 = (-2/3)(X - 1)$$

$$\Rightarrow \boxed{2X + 3Y + 1 = 0} \rightarrow \textcircled{II}$$

Solving \textcircled{I} & \textcircled{II} we get $\{\bar{X} = -2, \bar{Y} = 1\}$

⑩ Regn. Line of Y on X is

$$Y - \bar{Y} = b_{yx}(X - \bar{X}) = \left(r \frac{\sigma_y}{\sigma_x}\right)(X - \bar{X})$$

$X = 70 \Rightarrow Y = ?$

$$Y = \bar{Y} + \left(r \frac{\sigma_y}{\sigma_x}\right)(X - \bar{X})$$

$$= 67 + (0.8)\left(\frac{3.5}{2.5}\right)(70 - 65)$$

$$= 67 + 5.6 = 72.6$$

PARTICULARS		USE
Amount as per Contract Note	-18,804.51	NORMAL
CM/Derrial Charges	20.00	
Service Tax	3.00	
Amount as per Bill	-18,781.51	

- Q2) The 2 regression lines are $3X + 2Y = 26$ and $6X + Y = 31$. Find \bar{X} , \bar{Y} & r_{XY} and Ans: 4; 7; -0.5
- Q3) The 2 regression lines are $5X - 6Y = 24$ and $1000Y - 768X = 3608$. Find the ratio of the variances of X and Y. [Hint: ① is regⁿ line of X on Y & ② is Y on X] Ans: $r = 0.96$ and $\sigma_x/\sigma_y = 5/4$

- Q4) Given: ① The value of $r_{XY} = -\frac{1}{2}$

- ① Regⁿ line of Y on X passes through $(1, -1)$.

- ② other regⁿ line passes through $(1, -1)$. Determine the 2 regression lines and hence find \bar{X} & \bar{Y} .

Ans: Regⁿ Line of Y on X is $X + 6Y - 4 = 0$ & $b_{YX} = -\frac{1}{6}$
 X on Y is $2X + 3Y + 1 = 0$
 $\bar{X} = -2; \bar{Y} = +1$

Hint: Eqⁿ of line passing through (x_1, y_1) & (x_2, y_2) is $y - y_1 = m(x - x_1)$ where $m = \frac{y_2 - y_1}{x_2 - x_1}$

$\frac{Y - y_1}{y_1 - y_2} = \frac{X - x_1}{x_1 - x_2}; b_{YX} = -\frac{1}{6}$
 $b_{XY} = -\frac{3}{2}$

$r^2 = b_{XY} \cdot b_{YX} \Rightarrow b_{XY} = -\frac{3}{2}$
 (1/4)
 Regⁿ Line of X on Y is $X - \bar{X} = b_{XY}(Y - \bar{Y})$
 $\Rightarrow Y = \frac{1}{b_{XY}}(X - \bar{X} + \bar{Y}) = mX + c$
 $\Rightarrow m = \frac{1}{b_{XY}}; x_1 = 1, y_1 = -1$
 \Rightarrow Eqⁿ of R.L. of X on Y is $Y - (-1) = m(X - 1)$
 $\Rightarrow Y + 1 = m(X - 1)$
 $r = 0.8$

Q10) Given:

	Chennai (X)	Mumbai (Y)
Mean	20.65 (\bar{X})	20.67 (\bar{Y})
S.d	2.5 (σ_X)	3.5 (σ_Y)

(Price)

Find the most likely price in Mumbai corresponding to the price of Rs 70 at Chennai (Use regⁿ line of Y on X) Ans: Rs 72.6

Q11) Given:

	Sales Index	91	97	108	121	67	124	51	73	111	57
Purchase Index	71	75	69	97	70	91	39	61	80	47	

- ① Estimate the purchase index if sale index is 100 \rightarrow Ans: 76.13
- ② Obtain Pearson's Corⁿ Co-eff^t. using ④ Product Moment method
- ③ Obtain Spearman's Rank Corⁿ Co-eff^t. $\rightarrow r = 0.89$ (21/2) = 18

Solution

⑩ Given: Wrong $P = 0.8$; $n = 10$
Pb ⑩ Wrong $d_i = 7$; correct $d_i = 9$

$$P = 1 - 6 \left[\frac{\sum (d_i^2)}{n(n^2 - 1)} \right]$$

$$\text{Wrong } \sum (d_i^2) = \frac{(990)(0.2)}{6} = 33$$

$$\Rightarrow \text{Correct } \sum (d_i^2) = 33 - 49 + 81 = 65$$

$$\Rightarrow \text{Correct } P = 1 - \left[\frac{(6)(65)}{990} \right] = \frac{160}{165} = 0.606060\ldots \\ = 0.6061 \text{ app}$$

End of Contract

- 12) Following data gives the ranks assigned by 3 judges in a beauty contest:-

Judge ①	1	6	5	10	3	2	4	9	7	8
②	3	5	8	4	7	10	2	1	6	9
③	6	4	9	8	1	2	3	10	5	7

Determine which pair of judges will have the nearest P -values approach to beauty.

Ans: $\sum(d_{12}^2) = 200$ $\bar{0.212}$
 $\sum(d_{23}^2) = 214$ $\bar{0.297}$
 $\sum(d_{13}^2) = 60$ $\bar{0.636}$ ✓

- 13) Find the rank correlation $P_{X,Y}$ & write down its significance.

X: Supply	124	100	105	112	102	93	99	123	104	99	113	121	103	101
Y: Cost	80	100	102	91	100	111	109	89	104	111	102	98	111	123

Ans: -0.764

- 14) Find $r_{X,Y}$ between

X: Fertilizer used (mt. tonnes)	15	18	20	24	30	35	40	50
Y: Productivity (mt. tonnes)	85	93	95	105	120	130	150	160

Ans: 0.9917

- 15) Given: Marks assigned to 7 debators by 2 judges are as follows:-

Marks by Judge A	40	34	28	30	44	38	31
Marks by Judge B	32	39	26	30	38	34	28

An eighth debator was awarded 36 marks by judge A, while judge B was not present. Estimate the most likely mark that would have been awarded by judge B to the eighth debator.

Ans: 33.015 (Verify the answer)

- 16) The rank corⁿ. co. eff^t. of marks obtained by 10 students in 2 subjects was found to be 0.8 . It was later observed that the difference in ranks in ^{the} 2 subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct rank corⁿ. [ranks are not repeated]

Ans: 0.606

Extra Problems:

- ① Given: Mean of $X = 5$; Mean of $Y = 10$.
Line of regⁿ of Y on X is parallel to the line

$$20Y = 9X + 40.$$

Ans: $Y = 22.25$

Estimate Y when $X = 30$

Hint: Use Regⁿ line of Y on X ; $b_{yx} = \frac{9}{20}$

- ② If the tangent of the angle made by the regⁿ line of Y on X is 0.6 and $\sigma_y = 2\sigma_x$, find $r_{X,Y}$

Note: If $Y = mX + c$ is the eqⁿ of the line, then
 $m = \text{slope of the line} = \tan \theta$; $\theta \rightarrow$ angle made by the line with +ve x -axis.

Ans: $r = 0.3$

Hint: Given: $b_{yx} = 0.6$

- ③ Given: Angle between the lines of regⁿ is $\tan^{-1}(3)$ & $\sigma_x = \sigma_y$

Find $r_{X,Y}$

Hint: Use the formula: $\tan \theta = \left(\frac{1-r^2}{r} \right) \left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right)$ (*)

$\theta \rightarrow$ angle between 2 regⁿ lines; $-1 \leq r \leq 1$
Given: $\theta = \tan^{-1}(3)$; Let $\sigma_x = \sigma_y = \sigma$, say

Ans: $r = -0.17$

- ④ Given: $\sum X = 15000$; $\sum Y = 6800$; $\sum(X^2) = 2272500$;

$$\sum(Y^2) = 463025; \sum(XY) = 1022250; b_{yx} = 0.1$$

Find the value of n

Ans: $n = 100$

- ⑤ If $n = 50$, $\sum X = 75$; $\sum Y = 80$; $\sum(XY) = 120$;

show that the 2 regⁿ lines are perpendicular to

each other. (Hint: Using (*), $\theta = \frac{\pi}{2}$ when $r = 0$; show that

$$\{ \text{Cov}(X,Y) = \frac{1}{n} \sum XY - \bar{X}\bar{Y}; \bar{X} = \frac{\sum X}{n}; \bar{Y} = \frac{\sum Y}{n} \} \Rightarrow r = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}; \text{Cov}(X,Y) \Leftrightarrow r = 0 \text{ in this problem}$$

Given: $n = 36$; $\sum X = 120$; $\sum Y = 90$; $\sum(X^2) = 600$; $\sum(Y^2) = 250$; $\sum(XY) = 356$

Later it was found that entries were copied as

Find the correct value of r . (Ans: 0.0545)

8	10
12	7
X	Y

instead of

8	12
10	8
X	Y

Ans
Corrected
$\sum X = 118$
$\sum Y = 93$
$\sum XY = 368$
$\sum X^2 = 558$
$\sum Y^2 = 309$

corrected
66
 $r = \frac{1306.2339}{1306.2339}$

Question Bank (A.M.V)
Correlation Co-efficient (Karl Pearson)
&
Regression Analysis

① The following table enumerates the marks obtained by students in Statistics I & II papers.

St-I	80	45	55	56	58	60	65	68	70	75	85
St-II	82	56	50	48	60	62	64	65	70	74	90

Find (i) Corⁿ coeff^t r by (a) Product Moment Method
 (b) Using Regression Co-efficients.

Ans
 $r = 0.9$

(ii) Most likely mark in Stat-II if mark in Stat-I is 86

② Find the 2 regⁿ lines; hence or otherwise find 'r' between X & Y:-

X	21	23	30	54	57	58	72	78	87	90
Y	60	71	72	83	110	84	100	92	113	135

Ans:-
 $r = 0.876$

③ Marks in English	44	42	40	52	39	32	24	46	41	50
" Maths	24	25	28	29	32	35	36	41	45	50

Ans
 $r = 0.0808$
 $P = 0.0182$

Calculate: (i) Karl Pearson Corⁿ Co-eff^t $r_{X,Y}$
 (ii) Spearman's Rank Corⁿ Co-eff^t $P_{X,Y}$

④ Prod ⁿ of tea (in crores of kgs)	34	27	31	38	38	36	39	40
Price (Rs/kg)	3.75	4.62	4.25	4.12	4.28	4.32	4.21	4.05

Ans
 $r = -0.4$

Find (i) 'r' (ii) Most likely price if production of tea is 35 (crores of kgs)

⑤ Price (in Rs)	100	98	85	92	90	84	88	90	93	95
Sales (in lakhs)	500	610	700	630	670	800	800	750	700	690

Ans
 $r = 0.8179$
 $P = 0.8$

Find (i) $r_{X,Y}$ (ii) $P_{X,Y}$

(R.T.O)

Hts of sons (inches) \downarrow	56	55	58	58	57	56	60	54	59	57
Hts of fathers	68	67	67	70	65	68	70	66	68	66

Find (i) $r_{x,y}$ using two methods

(ii) the two regⁿ. eq^{ns}

(iii) Best estimated height of father whose son's height is 62 inches

(iv) Rank Corⁿ Coeff^t $r_{x,y}$

(v) Write down the significance of $r_{x,y}$

Ans: $Y = 39 + \underbrace{0.5X}_{b_{yx}} ; X = 15.67 + \underbrace{0.6122Y}_{b_{xy}} ; r = 0.5533 ; r^2 = 0.4876$ (verify)

$X = 62 \Rightarrow Y = 70$ inches.

Marks by A	36	32	34	31	32	32	35
Marks by B	35	33	31	30	34	32	36

given to 7 participants in a beauty contest. Estimate the approximate mark given by A if corresponding mark given by B to the eighth participant is 37. Ans: 35

(8) Given: $r_{x,y} = 0.8 ; \sum(xy) = 60 ; \sum(x^2) = 30 ; \sigma_y^2 = 2.5 = \text{s.d. of } y$

Find n $[x = x - \bar{x}, y = y - \bar{y}]$ $r = \frac{\sum(xy)}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$ Ans: $n = 30$

(9) Given: $\sum X = 250 ; \sum Y = 300 ; \sum(XY) = 7900 ; \sum(X^2) = 6500 ;$

$\sum(Y^2) = 10000 ; n = 10.$

Find (i) $r_{x,y}$ (ii) b_{xy} (iii) b_{yx} (iv) 2 regⁿ. lines

Ans: $X = \underbrace{0.4Y + 13}_{b_{yx}} ; Y = \underbrace{1.6X + 10}_{b_{xy}} ; r = 0.8$

(10) For 50 students in a class, the regⁿ. line of X on Y is $3Y - 5X + 180 = 0$. If mean of X is 62.4 & $\text{var}(X) = \frac{9}{16} \text{Var}(Y)$, find (i) mean of Y

(ii) $r_{x,y}$

Ans: $\bar{Y} = 44 ; r = 0.8$

(11) Given: Regⁿ Line of X on Y is $X = 4Y + 5$; Regⁿ Line of Y on X is $Y = kX + 4$

(i) S.T. $0 \leq k \leq \frac{1}{4}$ (ii) If $k = \frac{1}{16}$, find \bar{X}, \bar{Y} & $r_{x,y}$

Ans: $\bar{X} = 28 ; \bar{Y} = \frac{23}{4} ; r = 0.5$ (P.T.O)

Given: Regⁿ. Line of Y on X is $X + 2Y - 5 = 0$
 " " " X on Y is $2X + 3Y - 8 = 0$

$\text{Var}(X) = 12$.

Find (i) \bar{X} (ii) \bar{Y} (iii) $r_{X,Y}$ (iv) σ_y^2 (or $\text{Var}(Y)$)

Ans: 1 ; 2 ; -0.866 ; 4

(13) Given:

	Calcutta	Mumbai	$r_{X,Y} = 0.8$
Mean Price	65	57	
S.D	2.5	3.5	

Find the most likely price in Mumbai if price is Rs 70 at Calcutta.

[Ans: Rs 72.6]

(14) For 2 variables X and Y with the same mean, the

2 regⁿ. equations are $Y = aX + b$ & $X = \alpha Y + \beta$.

S.T. $\frac{b}{\beta} = \frac{1-a}{1-\alpha}$. Also find the common mean.

Proof: The 2 regⁿ. lines are

$Y - \bar{Y} = b_{YX} (X - \bar{X})$ & $X - \bar{X} = b_{XY} (Y - \bar{Y})$; Given: $\bar{X} = \bar{Y} = m$, say

$\Rightarrow b_{YX} = a; b_{XY} = \alpha; (\bar{X} = m; \bar{Y} = m)$

$\Rightarrow Y - m = a(X - m)$ & $X - m = \alpha(Y - m)$

$\Rightarrow Y = aX - am + m$ & $X = \alpha Y - \alpha m + m$

$\Rightarrow Y = aX + m(1-a)$ & $X = \alpha Y + m(1-\alpha)$

$\Rightarrow b = m(1-a); \beta = m(1-\alpha)$

$\Rightarrow \frac{b}{\beta} = \frac{(1-a)}{(1-\alpha)}$ & $m = \frac{b}{1-a} = \frac{\beta}{1-\alpha}$ [the means of X & Y]

(15) P.T. $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2r_{X,Y} \sigma_X \sigma_Y$

S.T. $\tan \theta = \frac{(1-r^2) \sigma_X \sigma_Y}{\sigma_X^2 + \sigma_Y^2}$; 'θ' is the angle between the two regⁿ. lines.

Find θ when $r=0$

Fitting of Curves [Principle of Least Squares]

[Use Normal Equations] Get values upto 4 decimals

- ① Fit a st. line of the form $Y = a + bx$ for the foll. data:-

X	0	1	2	3	4	5
Y	1	2	3	4.5	6	7.5

Ans: $Y = 0.70 + 1.32x$
(app)

②

X	100	120	140	160	180	200
Y	0.45	0.55	0.60	0.70	0.80	0.85

Using the above data and suitable st. line trend, estimate

Y when $X = 220$

Ans: $Y = 0.041 + 0.0041X$
 $X = 220 \Rightarrow Y = 0.943$
(app)

- ③ Estimate the prodⁿ in 2020, using

(i) st. line trend (ii) parabolic trend for the following data:-

Year :	1951	1961	1971	1981	1991
Prod ⁿ (tons)	10	12	8	10	15

(Use: 1971 as mid. year & $X = \text{Year} - 1971$)

Ans: $Y = -146.68 + 0.08X$

(i) $X = 2020 - 1971 = 49$

$Y = ?$ (tons)

- ④ Fit a parabola to the following data:

X	-2	-1	0	1	2
Y	1.0	1.8	1.3	2.5	6.3

Ans: $Y = 1.48 + 1.13X + 0.55X^2$

- ⑤ Fit a curve of the form $Y = ab^X$ for the foll. data:-

X	0	2	4
Y	5.012	10.0	31.62

Ans: $Y = 4.68(1.58)^X$