K. J. Somaiya College of Engineering

(A Constituent College of Somaiya Vidyavihar University)

Question Bank

Correlation

1. Calculate the coefficient of correlation for the following data:

Values of X	12	9	8	10	11	13	7
Values of Y	14	8	6	9	11	12	3

[Ans: 0.949]

2. Calculate the coefficient of correlation between density of population and death rate

Cities	Area in Sq. miles	Population in thousand	No. of deaths
A	150	30	300
B	180	90	1440
C	100	40	560
D	60	42	840
E	120	72	1224
F	80	24	312

[Ans: 0.988]

3. Find the coefficient of correlation for the following series:

Price (in Rs.)	10	11	12	13	14	15	16	17	18	19
Demand (in Kg)	420	410	400	310	280	260	240	210	210	200

[Ans: -0.96]

4. The following data gives the growth of employment in organized sector in India between 1988 and 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

Find the coefficient of correlation between the employment in public and private sector. [Ans: 0.772]

5. The no. of pairs of observation of x and y are 1000. $\sigma_x = 4.5$; $\sigma_y = 3.6$; $\sum (x - \bar{x})(y - \bar{y}) = 4800$. Calculate the coefficient of correlation between x and y series. [Ans: 0.296]

6. From the following table, find the missing values and calculate the coefficient of correlation by Karl Pearson's method.

X	6	2	10	4	?
Y	9	11	?	8	7

Arithmetic means of x and y series are 6 and 8 respectively.
[Ans: 8, 5, and -0.92]

7. A computer while calculating the correlation coefficient between two variables, x and y obtained the following results:

n=25, $\sum x = 125$, $\sum x^2 = 650$, $\sum y = 100$, $\sum y^2 = 460$, $\sum xy = 508$

It was, however, discovered at the time of checking that it had copied down two pairs of observations as (6,14) and (6,8) in place of correct pairs (8,12) and (8,6). Obtain the correct value of coefficient of correlation. [Ans:0.67]

8. For ten pairs of values of x and y the following results were obtained

variable	x	y
mean	30.1	47.8
SD	6.2	9.5

Coefficient of correlation between x and y is 0.72. Later, it is found that one pair was wrongly taken as (35,47) while actual values were (53,74). Determine the correct value of r. [Ans:0.862]

9. In sets of variables x and y with 50 observations each, the following data were obtained.

$\bar{x} = 10, \sigma_x = 3, \bar{y} = 6, \sigma_y = 2, r_{xy} = 0.3$.

But on subsequent scrutiny it was found that one value of x viz. x=10 and one value of y viz. y=6 was inaccurate and hence deleted. Find the correlation coefficient of remaining 49 pairs of observations. [Ans: 0.3]

Rank Correlation

10. The rankings of 10 individuals at the start and at the finish of a course of trainings are as below

Individuals	A	B	C	D	E	F	G	H	I	J
Rank before(X)	1	6	3	9	5	2	7	10	8	4
Rank after(y)	6	8	3	7	2	1	5	9	4	10

Calculate rank correlation coefficient. [Ans:0.394]

11. Calculate the coefficient of rank correlation for the following series:

Series (x)	57	59	62	63	64	65	55	58	57
Series (y)	113	117	126	126	130	129	111	116	112

[Ans: 0.967]

12. Ten competitors in a beauty contest are ranked by three judges in the following order.

competitor	A	B	C	D	E	F	G	H	I	J
First judge(X)	1	6	5	10	3	2	4	9	7	8
Second judge(y)	3	5	8	4	7	10	2	1	6	9
Third judge (z)	6	4	9	8	1	2	3	10	5	7

(i) Use rank correlation coefficient to discuss which pair of judges have the nearest approach to common tastes in beauty?

(ii) What does the obtained value of R indicate?

(iii) If instead of calculating values of rank correlation coefficient you calculate Pearson's coefficient of correlation between ranks, would your results be the same or different?

Regression

1. The following table shows the arithmetic mean and standard deviation of the advertising. Expenditure (X) and sales of the company (Y)for the year 2008-2009.

Statistical measures	Advertising Expenditure(X) (Rs. Lakhs)	Sales(Y) (Rs. Lakhs)
Arithmetic mean	20	100
Standard deviation	3	12

Coefficient of correlation between X and Y is 0.8

(i) Find the equation of two lines of regression

(ii) What would be the expected sales of the company if the advertising expenditure is Rs. 32 lakhs? [y=3.2x+36, x=0.2y,138.4 lakhs]

2. Given below are data relating to prices of a share in Kolkata and Mumbai Stock exchanges:

Particulars	Kolkata(X) Rs.	Mumbai (Y) Rs.
Average Price	130	134
Standard deviation	5	7

Coefficient of correlation between prices of the shares in Kolkata and Mumbai is 0.8. Fit the two lines of regression and find the most likely price in Mumbai Stock Exchange corresponding to a price of Rs.140 in Kolkata. [y=1.12x-11.6,7x=4y+37 4, Rs.145.2]

3. A panel of two judges A and B, graded seven T.V. serial performances by awarding marks independently as shown in the following table:

Performance	1	2	3	4	5	6	7
Marks by A	46	42	44	40	43	41	45
Marks by B	40	38	36	35	39	37	41

The eighth T.V. performance which judge B could not attend, was awarded 37 marks by judge A. If the judge B had also been present, how many marks would be expected to have been awarded by him to the eighth T.V. performance? [y=0.75x+5.75,33.5]

4. A department store gives in-service training to its salesmen which is followed by a test. It is considering whether it should terminate the service of any salesman who does not do well in the test. The following data give the test scores and sales made by 9 salesmen during a certain period:

Test Scores	14	19	24	21	26	22	15	20	19
Sales ('00 Rs.)	31	36	48	37	50	45	33	41	39

(i) Calculate the coefficient of correlation between the test scores and the sales.

(ii) Does it indicate that the termination of services of low-test scores is justified?

(iii) If the firm wants a minimum sales volume of Rs.3000, what is the minimum test score that will ensure continuation of service?

(iv) Also estimate the most probable sales volume of a salesman making a score of 28. [0.9476; high degree of correlation, yes, 14.42, Rs.5287]

5. Given n = 100, Σx = 12500, Σy = 8000, Σx^2 = 1585000, Σy^2 = 648100, Σxy = 1007425 obtain the regression line of y on x. Also find value of x at y = 10.

6. The two regression lines are $4x - 5y + 33 = 0; 20x - 9y = 107$ and variance of x = 25.

Find i) mean of x & y

ii) Coefficient of correlation

iii) Variance of y

7. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible. Variance of x = 9, regression equations $8x - 10y + 66 = 0; 40x - 18y = 214$

8. Find i) Find $\overline{x}, \overline{y}$

ii) The S.D. of y

iii) The coefficient of correlation between x & y

[***Ans:*** $\overline{x} = 13, \overline{y} = 17, \sigma_y = 4, r = 0.6$]

9. The heights in cms of fathers (x) and of the eldest sons(y) are given below

X	165	160	170	163	173	158	178	168	173	170
Y	173	168	173	165	175	168	173	165	180	170

Estimate the height of the eldest son if the heights of the fathers are 172 cm. and the height of the father if the height of the eldest son is 173cm.Also find the coefficient of correlation between the heights of fathers and sons.

10. Find i) the lines of regressions, ii) find y if x=71, iii) find x if y=70, iv) Coefficient of correlation for the following data. v) Find angle θ between two lines of regressions.

X	65	66	67	67	68	69	70	72
Y	67	68	65	66	72	72	69	71

11. Find i) the lines of regressions, ii) find y if x=75, iii) find x if y=177, iv) Coefficient of correlation for the following data. v) Find angle θ between two lines of regressions.

X	70	72	74	76	78	80
Y	163	170	179	188	196	220

12. The regression lines of a sample are $X + 6y = 6$ and $3X + 2y = 10$

Find i) sample mean $\bar{X}\&\bar{Y}$,ii) coefficient of correlation between X and Y

Example-5) Given$-5x + 6y = 90, 15x - 8y = 130$, $\sigma_x^2 = 16$

Find i) $\bar{x}\&\bar{y}$,ii) r_{xy},iii) σ_y^2

13. Given

	Mathematics	English
Mean	80	50
S.D.	15	10

Coefficient of correlation r=0.4. i)Estimate the marks of the students in Mathematics who scored 60 marks in English, ii) Estimate the marks of the students in English who scored 80 marks in Mathematics.

14. Given

	X	Y
Mean	36	85
S.D.	11	8

Coefficient of correlation r=0.66.

i) Find the two regression equations. Ii) Estimate the value of X when Y=75, iii) Estimate the value of Y when X=30.