Correlation and Regression (CO-1)

S.No	QUESTION
	Correlation
1.	Calculate Karl Pearson's coefficient of correlation.

x:	23	27	28	29	30	31	33	35	36	39
y:	18	22	23	24	25	26	28	29	30	32

2. Obtain the rank correlation coefficient from the following data

x:	10	12	18	18	15	40
y:	12	18	25	25	50	25

3. Calculate rank correlation coefficient and coefficient of correlation for the following data And interpret your result.

x:	12	17	22	27	32
y:	113	119	117	115	121

4. From the data calculate Karl Pearson's correlation between x & y.

x:	36	56	20	42	33	44	50	15	60
y:	50	35	70	58	75	60	45	80	38

5. From the data calculate Spearmen's rank correlation between x & y.

x:	36	56	20	42	33	44	50	15	60
y:	50	35	70	58	75	60	45	80	38

6. Determine the coefficient of rank correlation from the following data-

x:	68	64	75	50	64	80	75	40	55	64
y:	62	58	68	45	81	60	68	48	50	70

7. Determine the Karl Pearson's coefficient of correlation from the following data-

x:	68	64	75	50	64	80	75	40	55	64
y:	62	58	68	45	81	60	68	48	50	70

8. Find Karl Pearson's coefficient of correlation for the indices of supply and price of an article.

Supply Index:	124	100	112	102	93	99	104	99	113	103	101
Price Index:	80	100	91	100	111	109	104	111	102	111	123

9. Calculate the coefficient of correlation between the indices of business activity (X) and employment (Y) from the following data.
X : 100, 102, 108, 111, 115, 116, 118.
Y : 100, 100, 104, 108, 112, 119, 120.

10. .For 10 pairs of values of x and y the following values are determined:
Later on it was found that one pair of values was taken as (34, 47) instead of (43, 74)
.Determine the correct value of the coefficient of correlation if
Mean(X) = 30.1, Mean(Y) = 47.8, S.D.(X)=6.2, S.D.(Y)=9.5, r= 0.72

11. The coefficient of rank correlation of the marks obtained by 10 students in Physics and chemistry was found to be 0.5. It was later discovered that the differences in ranks in the two subjects obtained by one of the students was wrongly taken as 3 instead of 7. Find the correct coefficient of rank correlation.

12. Soil temperature (x) and germination (y) for winter wheat in 12 places are as follows.
Determine the correct value of the coefficient of correlation

x (^0F)	57	42	38	42	45	42	44	40	46	44	43	40
y(days)	10	26	41	29	27	27	19	18	19	31	29	33

13 Find the coefficient of correlation 'r' between the heights of father (x) and sons (y) from the following data.

x:	65	66	67	68	69	70	71	67
y:	67	68	64	72	70	67	70	68

14 The panel of two judges A & B graded dramatic performance by independently awarding marks as follows. eight performance, however, which judge B could not attend, got 38 marks by judge A. if judge B had also present, how many marks would he be expected to have awarded to the eight performance?

Performance No.	1	2	3	4	5	6	7
Marks by A	36	32	34	31	31	32	35
Marks by B	35	33	31	30	34	32	36

15 Compute Spearman's rank correlation coefficient from the following data.
X : 85, 74, 85, 50, 65, 78, 74, 60, 74, 90.
Y : 78, 91, 78, 58, 60, 72, 80, 55, 68, 70.

16 Calculate the Spearman's rank correlation coefficient from the following data.
X : 23, 27, 28, 29, 30, 31, 33, 35, 36, 39.
Y : 18, 22, 23, 24, 25, 26, 28, 29, 30, 32.

17 Calculate the correlation coefficient from the following data.
X : 23, 27, 28, 29, 30, 31, 33, 35, 36, 39.
Y : 18, 22, 23, 24, 25, 26, 28, 29, 30, 32.

18 A sample of 25 pairs of values of x and y lead to the following results. $\sum x = 127$, $\sum y = 100$, $\sum x^2 = 760$, $\sum y^2 = 449$, $\sum xy = 500$. Later on it was found that two pairs of values were taken as (8,14) and (8,6) instead of correct values (8,12) and (6,8). Find corrected correlation coefficient between x and y.

19 Calculate the coefficient of correlation between the indices of business activity (X) and employment (Y) from the following data.
X : 100, 102, 108, 111, 115, 116, 118.
Y : 100, 100, 104, 108, 112, 119, 120.

20 Compute Spearman's rank correlation coefficient from the following data.
X : 32, 55, 49, 60, 43, 37, 43, 49, 10, 20.
Y : 40, 30, 70, 20, 30, 50, 72, 60, 45, 25.

21 The values of demand and the corresponding price of a commodity are given in the following table. Find Karl Pearson's coefficient of correlation.

Demand in quintals	65	66	67	67	68	69	70	72
Price in Rs per kg	67	68	69	68	72	72	68	71

22 In two sets of variables x and y with 50 observations each gave the results $\bar{x} = 10$, $\bar{y} = 6$, $\sigma_x = 3$ $\sigma_y = 2$, $r = 0.3$. But on subsequent verification it was found that one value of x=10 and one value of y=6 were inaccurate and were discarded. With the remaining 49 pairs of values how is the original value of 'r' affected?

23 Calculate coefficient of correlation using change of scale and origin

X	1200	1300	1400	1500	1600	1700	1800	1900
Y	5445	6145	6645	7045	7445	7845	8545	8945

24 Obtain the rank correlation coefficient from the following data.
X : 18, 20, 34, 52, 12
Y : 39, 23, 35, 18, 46.

25 Compute Rank correlation coefficient from the following data

x	105	104	102	101	100	99	98	96	93	92
y	101	103	100	98	95	96	104	92	97	94

26 Following data gave the growth of employment in lakhs in organised sector of India

between 1988 and 1995. Find the correlation coefficient between the employment in Public and Private sector and give the comment.

year	88	89	90	91	92	93	94	95
Public sector	98	101	104	107	113	120	125	128
Private sector	65	65	67	68	68	69	68	68

27 Following table shows the marks obtained by 8 students in Accountancy and Statistics. Spearman's coefficient of rank correlation

Student No	1	2	3	4	5	6	7	8		
Accountancy	45	70	65	30	90	40	50	57		
Statistics	35	90	70	40	95	40	60	80		

(i)Will the result change if the marks of the two subjects of all the students are increased by 5 (ii) Will the result change if the marks of the two subjects of all the students are halved ?

28 Calculate the correlation coefficient from the following data

$N=9,\ \sum X=45,\quad \sum Y=108,\quad \Sigma X^2=285,\quad \Sigma Y^2=1356,\quad \Sigma XY=597$

29 Ten competitors in a beauty contest are ranked by three judges in the following order.Use the method of rank correlation coefficient to determine

First Judge	1	6	5	10	3	2	4	9	7	8
Second Judge	3	5	8	4	7	10	2	1	6	9
Third Judge	6	4	9	8	1	2	3	10	5	7

which pair of judges has the nearest approach to common taste in beauty?

30 Calculate rank correlation coefficient of the following data

Subject 1	40	46	54	60	70	80	82	85	87	90	95
Subject 2	45	46	50	43	40	75	55	72	65	42	70

31 Calculate coefficient of correlation for the ages of husbands and their respective wives:

Age of husbands	23	27	28	29	30	31	33	35	36	39
Age of wives	18	22	23	24	25	26	28	29	30	32

32 Calculate the coefficient of correlation between *X* and *Y* series from the following data.Summation of product deviations of *X* and *Y* series from their respective arithmetic means is 122.

	No of pairs of observations	Arithmetic mean	Standard Deviation	Sum of squares of deviations from the arithmetic mean
X	15	25	3.01	136
Y	15	18	3.02	138

33 Calculate coefficient of correlation using change of scale and origin

X	230	270	280	290	300	310	330	350	360	390
Y	1890	2290	2390	2490	2590	2690	2890	2990	3090	3290

34 An examination of 11 applicants for a accountant post was taken by a finance company. The marks obtained by the applicants in the reasoning and aptitude tests are given below. Calculate Spearman's rank correlation coefficient from

the data given above.

Applicant	A	B	C	D	E	F	G	H	I
Reasoning Test	20	50	28	25	70	90	76	45	30
Aptitude Test	30	60	50	40	85	90	56	82	42

35 A random sample of recent repair jobs was selected and estimated cost and actual cost were recorded. Calculate the value of spearman's correlation coefficient.

Estimated Cost	300	450	800	250	500	975	475	400
Actual Cost	273	486	734	297	631	872	396	457

36 Find lines of regression for the following data to estimate y corresponding to x=155 estimate x corresponding to y= 75

x	100	110	120	130	140	150	160	170	180	190
y	45	51	54	61	66	70	74	78	85	89

37 Find the equations of two regression lines for the following data

X : 78, 36, 98, 25, 45, 82, 90, 62, 65, 39.

Y : 84, 51, 91, 60, 68, 62, 86, 58, 53, 47.

Estimate the value of Y when X is 50 and value of X when Y is 90.

38 From the data: i) Find the two regression equations, ii) estimate the value of X when Y = 75.

	X	Y
Arithmetic mean	36	85
Standard deviation	11	08
Correlation coeff	0.66	

39 The heights in cms of fathers(x) and of the eldest sons (y) are given below.
Estimate the height of the eldest son if the height of the father is 172 cms. and the height of the father if the eldest son is 173 cm. also find the coefficient of correlation between the heights of father and son.

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

40 Obtain two lines of regression and coefficient of correlation from the following data-

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

41 Obtain two lines of regression and coefficient of correlation

x	62	64	65	69	70	71	72	74
y	126	125	139	145	165	152	180	208

42 The given data indicates weights(x) and heights(y) of 1000 men.
$\bar{x} = 150\ lbs,\ \bar{y} = 68\ inches,\ \sigma_x = 20\ lbs,\ \sigma_y = 2.5\ inches,\ r = 0.6$ John weights 200 lbs Smith is 5ft tall. Estimate the height of John and Weight of Smith. From the value of height of John estimate his weight. Why it is different from 200?

43 Given the following information about the marks of 60 students, estimate the marks of a student in Mathematics who score 60 marks in Physics and the marks of a student in Physics who scored 70 in mathematics:

	Mathematics	Physics	r
Mean	80	50	0.4
S.D.	15	10	

44 From 8 observations the following results were obtained

$$\sum x = 59, \quad \sum y = 40, \quad \sum x^2 = 524, \quad \sum y^2 = 256, \quad \sum xy = 364.$$

Find the equations of the lines of regression and the coefficient of correlation.

45 State whether the following statement is true or false with reasoning – (i) The lines of regression between x and y are parallel to the lines of regression between 2x and 2y. (ii) The coefficient of regression between x and y are same as the coefficient of regression between (2x+5) and (2y-7).

46 A chemical engineer is investigating the effect of process operating temperature X on product yield Y. Find the regression line to predict yield on the basis of temperature. Also verify that the sum of the coefficients of regression is greater than 2r.

X	120	130	140	150	160	170	180	190
Y	54	61	66	70	74	78	85	89

47 Obtain two lines of regression and coefficient of correlation

x	65	66	67	67	68	69	70	72
y	67	68	35	66	72	72	69	71

48 Find Coefficients of regression and hence the regression lines for the following data

x	78	36	39	65	62	90	75	30	98	85
y	84	51	47	53	58	86	68	60	91	70

49 If $R_{xy} = 0.143$ and the sum of squares of the differences between the ranks is 48, find N.

50 If $r_{xy} = 0.4,\ cov(x,y) = 2.4,\ \sigma_y{}^2 = 36.$ Find σ_x

51 The equations of the two regression lines are $3x + 2y = 26$ and $6x + y = 31$. Find **(i)** The means of x and y and
(ii) Coefficient of correlation between x and y.

52 It is given that the mean x and y are 5 and 10. If the line of regression of y on x is parallel to the line $20y = 9x + 40$. Estimate the value of y for $x = 30.$

53 The regression equations of y on x and of x on y are $y = x$ *and* $4x - y = 3$ respectively and the second moment of x about the origin is 2. Find (i) the mean of x & mean of y.(ii)correlation coeff (iii) standard deviation of x & y.

54 The equations of the two lines of regression for a bivariate data are $9x + 10y - 67 = 0,$ *and* $5x + 2y - 23 = 0$ Find i) mean values of x and y, ii) regression coefficient, iii) correlation coefficient.

55 Given $6y = 5x + 90, 15x = 8y + 130, \sigma_x^2 = 16$.
Find i) $\bar{x}$ *and* $\bar{y}$ ii) Correlation coefficient, iii) σ_y^2

56 In a partially destroyed laboratory record of analysis of correlation data, following results are legible. Variance of x=9, equations of the lines of regression $4x - 5y + 33 = 0,\ 20x - 9y - 107 = 0.$ Find (i) the mean values of x and y, (ii) the standard deviation of y, (iii) coefficient of correlation .

57 The regression lines of a sample are $x + 6y = 6$ *and* $3x = 2y = 10$ Find (i) $\bar{x}$ *and* $\bar{y}$ (ii) correlation coefficient. Also estimate y when $x = 12$. Also verify that the sum of the coefficients of regression is greater than $2r$

58 State true or false with reasoning: " $2x + y = 3$ *and* $x = 2y + 3$ cannot be the lines of regression."

59 If the tangent of the angle made by the lines of regression of y 0n x is 0.6 and $\sigma_y = 2\sigma_x$, find the correlation coefficient between x and y.

60 In a regression analysis, it is found that $b_{yx} = 0.87$, $b_{xy} = 1.55.$ Can these values be regarded as consistent values and why ?

61 (i) Let $r_{xy} = 0.4,\ Cov(x,y) = 1.6,\ \sigma_y^2 = 25.$ find σ_x. (ii) If $R_{x,y} = 0.143$ and the sum of the squares of the differences between the ranks is 48, find n.

62 If $\sigma_x = \sigma_y = \sigma$ and the angle between the lines of regression is $\tan^{-1} 3$, find the coefficient of correlation.

63 Find the regression coefficients & the coefficient of correlation
$N = 12, \sum x = 120, \sum y = 432, \sum xy = 4992, \sum x^2 = 1392, \sum y^2 = 18252.$

64

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Find the lines of regression.?

65 In logistic model people's sex as male or female from their long hair , then the Y=1 could be female , check whether a person with length of hair 5 cm is a female . (given b0=-10, b1=0.2) Ans:0.99987

66 The relative risk of developing cardiovascular disease (CVD) for people with low- and high-salt diets was estimated

Developed CVD	salt in diet		Total
	Low	High	
Yes	88	112	200
No	1081	1246	2215
Total	1169	1246	2415

a) For each salt level, find the probability of developing CVD.
(b) Convert each of the probabilities that you found in part (a) to odds.
(c) Find the log of each of the odds that you found in part (b).

67 A survey was conducted for some students to find the number of hours each student spent in daily studying and whether they passed or failed. Using Logistic regression, it was found that Log(odds of passing exam)= 1.5046.hours-4.0777 Find
(i) Odds of passing the exam
(ii) Probability of passing exam
(iii) Slope of odds ratio
(iv) Probability of passing exam if a student studies 2 hrs daily.

68 A survey was conducted for some people regarding Facebook

People	Use of Facebook		Total
	Yes	No	
Women	283	312	595
men	1682	1258	2940
Total	1965	1570	3535

a) For men and women, find the probability of using Facebook
(b) Convert each of the probabilities that you found in part (a) to odds.
(c) Find the log of each of the odds that you found in part (b).

69 A data for selection of students regarding placement is as

students	Selection in placement		Total
	Yes	No	
boys	753	102	855
girls	382	158	540
Total	1145	250	135

a) For boys and girls, find the probability of selection of students
(b) Convert each of the probabilities that you found in part (a) to odds.
(c) Find the log of each of the odds that you found in part (b).

70