Module: 3: Correlation and Regression
(Correlation and Regression – Rank Correlation – Partial and Multiple Correlation – Multiple Regression)

1. Find the coefficient of correlation between industrial production and export using the following data: **Ans: 0.603**

Production(x)	55	56	58	59	60	60	62
Export(y)	35	38	37	39	44	43	44

2. Calculate the coefficient of correlation and obtain the least square regression line of y on x for the following data:

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

Also obtain an estimate of y which should correspond on the the average to x= 6.2
Ans: 0.95; y = 0.95 x+7.25; 13.14

3. The two lines of regression are $4x - 5y + 33 = 0$; $20x - 9y - 107 = 0$. The variance of X is 25. Find (i) the mean values of x and y; (ii) the variance of Y and (iii) correlation coefficient between X and Y. **Ans: (i) 13,17 (ii) 6.67 (iii) 0.6**

4. Find the co-efficient of correlation for the following table. **Ans: 0.898**

x	10	14	18	22	26	30
y	18	12	24	6	30	36

5. Obtain the line of regression from the following data

x:	25	28	35	32	31	36	29	38	34	32
y:	43	46	49	41	36	32	31	30	33	39

Test the validity of the linear regression model.
Ans: x=-0.2337 y+40.8806; y= -0.6643x+59.2576

6. The ranks of some 7 students in Mathematics and Physics are as follows. Calculate rank correlation coefficients for proficiency in Mathematics and Physics. **Ans: 0.6429**

Ranks in Maths X	1	2	3	4	5	6	7
Ranks in Physics Y	4	3	1	2	6	5	7

7. A sample of 12 fathers and their eldest sons have the following data about their height in inches.

Fathers(X)	65	63	67	64	68	62	70	66	68	67	69	71
Sons(Y)	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the rank correlation coefficient. **Ans: 0.722**

8. From the data relating to yield of dry bark (X_1), height (X_2) and girth (X_3) for 18 Cinchona plants the following simple correlation coefficients were obtained:
$$r_{12} = 0.77, r_{13} = 0.73 \text{ and } r_{23} = 0.52.$$
Find the partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$
Ans: 0.62 and 0.8564 (multiple correlation coefficient is always non negative).

9. The sale of a product in lakhs of rupees(Y) is expected to be influenced by two variables namely the advertising expenditure X_1 (in '000 Rs) and the number of sales persons(X_2) in a region. Sample data on 8 regions of a state has given the following results

Area	Y	X_1	X_2
1	110	30	11
2	80	40	10
3	70	20	7
4	120	50	15
5	150	60	19
6	90	40	12
7	70	20	8
8	120	60	14

Develop a multiple regression model for the above data.
Ans: The regression model is Y = 16.834- 0.2442 X_1 + 7.8488 X_2

10. From the following data, obtain $(R_{1.23})^2$.

X_1	65	72	54	68	55	59	78	58	57	51
X_2	56	58	48	61	50	51	55	48	52	42
X_3	9	11	8	13	10	8	11	10	11	7

Ans: r_{12} = 0.80; r_{13}=0.64; r_{23}=0.79; $R^2_{1.23}$= 0.63

① 0.896

② $Y = 0.95x + 7.25$

$r = 0.95$

$x = 6.2 \implies y = 13.14$

③ $E(x) = 13$

$E(Y) = 17$

$Var(Y) = 44.44$

$Cor(x,y) = 0.6$

④ $r = 0.6$

⑤ $r = \pm\sqrt{b_{yx} \, b_{xy}}$

$r = -0.39$

$X = -0.2337Y + 40.ff6$

$Y = -0.6643X + 59.2576$

$b = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2}$

$b_{yx} = \dfrac{}{N\Sigma x^2 - (\Sigma x)^2}$

6) 0.6429

7) 0.722

8) $r_{12.3} = 0.66$

$R_{1.23} = 0.86$

9) $Y = 16.834 - 0.2442x_1 + 7.8488x_2$

10) $r_{12} = 0.7899, r_{13} = 0.6059, r_{23} = 0.7946$,

$(R_{1.23})^2 = 0.6275$