Chapter → 7

Linear Correlation

		1	1	1	1	equal diff.
x	1	2	3	4	5	
y	2	4	6	8	10	
		2	2	2	2	

equal diff.

CORRELATION AND LINEAR REGRESSION

Page No.

Correlation is the relation between two or more than two variable or it measures the degree of relationship between two or more than two variable. If we study the relationship between only two variables then such a correlation is called simple and if we study the relation between three or more than three variables then such a correlation is called either multiple or partial. For example: The ~~corre~~ relation between income and expenditure, Relation between height and weight, Relation between sales and population etc.

Not imp

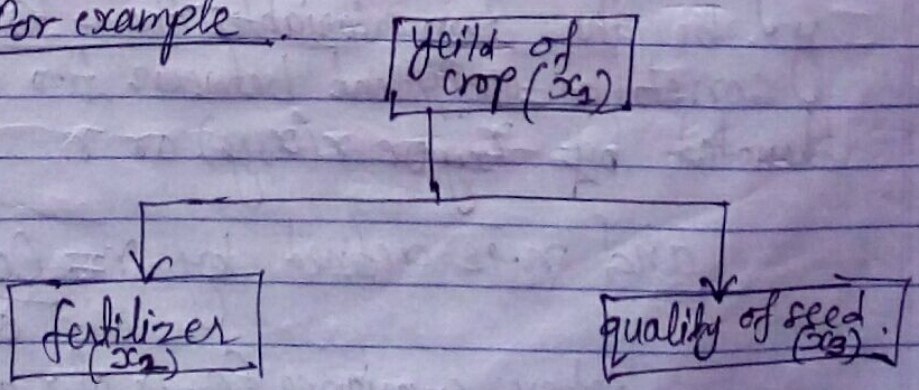
Types of correlation:

i) Positive correlation. → both are increasing or decreasing
ii) Negative correlation. → One is increasing and another decreasing
iii) Linear correlation. → If X increases by equal difference in each case then Y increase by equal difference in each case.
iv) Non-linear correlation. → Alternate.
v) Simple, multiple and partial correlation.

Multiple and partial:

Multiple → Multiple correlation is the correlation between dependent and joint effect of independent variable.

For example.



Here, the correlation between x_1 and joint effect of x_2 and x_3 is called multiple correlation and its coefficient

is denoted by $R_{1.23}$. Similarly if x_2 is dependent & x_1 & x_3 are independent then multiple correlation coefficient $R_{2.13}$

Partial → Partial correlation is the correlation between one dependent and one independent variable when other independent variable are taken constant.

From previous example the correlation between x_1 & x_2 where x_3 is taken as constant is a partial correlation & its coefficient is denoted by $r_{12.3}$.

Imp (1) Not imp

Methods of calculating Correlation Coefficient.

Imp
- i) Karl Pearson's Correlation Coefficient.
- ii) Spearman's Rank Correlation.
- iii) Scatter diagram or graphic method.

i) Karl Pearson's Correlation Coefficient:

Karal Pearson's Correlation measures the degree of relationship between two variables. Let x and y be the two variables then Karl Pearson's correlation coefficient between two variables is denoted by r_{xy} or $r(x,y)$ or r.

and r is defined by $r = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$

where covariance (Cov) of x and y is

$$Cov(x,y) = \frac{1}{n} \Sigma (x-\bar{x})(y-\bar{y})$$

Date.
Page No.

σ_x = st. dev. of $X = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2}$

σ_y = st. dev of $Y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2}$

Here, n is no. of pair of observation.

Then, $$r = \frac{\frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2}\sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2}}$$

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

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

where, $x = (X-\bar{x})$ and $y = (Y-\bar{y})$.

OR

$$\boxed{r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\cdot\Sigma y^2 - (\Sigma y)^2}}}$$

By using deviation or change of origin:

Let $U = X - A$

$V = Y - B$

where, A and B are assumed value of series x and y respectively

Then, (Replace x by U and y by V)

$$r = \frac{n\Sigma UV - \Sigma U \Sigma V}{\sqrt{n\Sigma V^2 - (\Sigma V)^2}\sqrt{n\Sigma V^2 - (\Sigma V)^2}}$$

for discrete series → common factor
for continuous series → ~~class~~ size of class interval.

Date:
Page No.

By using step deviation (or change of origin or scale):

Let $U' = \frac{X-A}{h}$

$V' = \frac{Y-B}{k}$

where h and k are common factors or size of class interval.

Then,

$$r = \frac{n\Sigma U'V' - \Sigma U' \cdot \Sigma V'}{\sqrt{n\Sigma U'^2 - (\Sigma U')^2}\sqrt{n\Sigma V'^2 - (\Sigma V')^2}}$$

Properties of Karl Pearson's Correlation Coefficient:

i) Correlation coefficient lies between −1 to +1.
i.e. $-1 \leq r \leq +1$

ii) Correlation coefficient is symmetrical i.e. $r_{xy} = r_{yx} = r$.

iii) Correlation coefficient is independent of change of origin and scale.

iv) Correlation coefficient is the geometric mean of the two regression coefficients. $r = \pm\sqrt{b_{yx} \times b_{xy}}$

v) Correlation coefficient has no unit because of relative measure.

Interpretation of calculated value of r:

i) If $r = 0$, then there is no correlation.

ii) If $r = +1$, then there is perfect positive correlation

iii) If $r = -1$, then there is perfect negative correlation

iv) If r lies between 0.001 to 0.499, then there is low degree of positive correlation.

v) If r lies between 0.5 to 0.699, then there is moderate positive correlation.

Date.
Page No.

vi) If r lies between 0.70 to 0.999 high degree positive correlation.

vii) If r lies between −0.999 to −0.001, there is low degree of negative correlation.

viii) If r lies between −0.699 to −0.500 there is moderate Negative Correlation.

ix) If r lies between −0.699 to −0.999 there is high degree of Negative Correlation.

Significant test of correlation coefficient (r):

For testing the significance of correlation coefficient (r) we need to find the probable error and hence probable error (P.E) $= 0.6745 \times \frac{1-r^2}{\sqrt{n}}$

where r is the correlation coefficient and n is the no. of pairs of observation.

Following are the condition of the significance of correlation coefficient.

i) If $|r| < P.E$ then correlation coefficient (r) is not significant.

ii) If $|r| > 6 \times P.E$ then the correlation coefficient (r) is significant.

In other situations nothing can be concluded certainly. The probable error of correlation is used to determine the limits within which the population correlation coefficient may be expected to lie of population correlation coefficient $= r \pm P.E$

regression is different from nonlinear regression?

12. What do you mean by coefficient of determination? How can you interpret it?

Numerical Questions

1. If the covariance between x and y variable is 36 and variance of X and Yare 36 and 100 respectively, find the coefficient of correlation between them. [Ans: 0.6]

2. Find the correlation coefficient between x and y series

	X	Y
Number of observation	10	10
Standard deviation	2.05	2.06

And $\Sigma(X - \bar{X})(Y - \bar{Y}) = 40$ [Ans: 0.947]

3. For 10 observations on two variables X and Y, the following information are as follows:
$\Sigma X = 666$, $\Sigma Y = 663$, $\Sigma X^2 = 44{,}490$, $\Sigma Y^2 = 44{,}061$, $\Sigma XY = 44{,}224$
Compute, Karl Pearson's coefficient of correlation. [Ans: 0.576]

4. From the following information find the total number of pair of observations given that $r = 0.8$, $\Sigma xy = 60$, S.D. of Y = 60 and $\Sigma x^2 = 90$, where x and y are deviations taken from their respective means. [Ans: 1636]

5. From the following data examine whether there exists any correlation between X and Y

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

[Ans: 0.95]

6. The following table gives the distribution of items and also defective items among them according to size groups. Find the correlation coefficient between size and defect in quality.

Size group	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45
No of items	200	270	340	360	400	300
No of defective items	150	162	170	180	180	120

[Ans: -0.939]

7. Following figure give the age in years of newly married husbands and wives (25,17) (26,18) (27,19) (25, 17) (26,19) (28,20) (25,17) (25,17) (24,18) (26,18) (26,20) (27,18) (27,19) (28,19) (25,18) (25,19) (26,18) (25,18) (27,20). Find Karl Pearson's Correlation coefficient. Test its significance. [Ans: 0.654]

8. From 20 pairs of X and Y variables the following results obtained
$\Sigma X = 127$, $\Sigma Y = 100$, $\Sigma X^2 = 860$, $\Sigma Y^2 = 549$, $\Sigma XY = 674$
at the time of verification, the following wrong values of X and Ywere taken as X = 10, 8 and Y= 14,6 instead of correct values X = 8,6 and Y= 12, 8. Find correct value of correlation. [Ans: 0.47]

9. A student calculates the value of r as 0.795 when the number of items is 100 and concludes that r is highly significant. Is his conclusion correct? [Ans: Yes]

10. Correlation coefficient between two variables with the pair of 10 observations is 0.81. Discuss if the value of r be significant or not. Also determine the limits of population correlation coefficient. [Ans: Significant, 0.736 to 0.883]

11. The information given below are related with ages of husband(X) and wife(Y) for married couples living together in a sample survey. Calculate the coefficient of correlation between age of husband and that of his wife. Test the significance of the calculated r.

$N = 72, \Sigma fX = 3560, \Sigma fX^2 = 196800, \Sigma fY = 3260, \Sigma fY^2 = 168400, \Sigma fXY = 172000$

[Ans: 0.52, Significant]

12. The ranking of Ten Competitors in voice test given by two experts A and B are as follows:

A	3	5	8	4	7	10	2	1	6	9
B	6	4	9	8	1	2	3	10	5	7

Calculate correlation coefficient between the rank of A and B. [Ans: −0.296]

13. From the following data calculate Spearman's rank correlation.

Variable (X)	10	15	18	14	30	27
Variable (Y)	100	120	118	119	130	105

[Ans: 0.48]

14. Find rank correlation coefficient between age of husband and age of wife from following data:

Age of husband (yrs)	23	27	28	29	30	31	33	35
Age of wife (yrs)	21	22	23	24	25	26	28	29

[Ans: 1]

15. Calculate Spearman's rank correlation coefficient from the following

X	20	25	60	45	80	25	55	65	25	75
Y	52	50	55	50	60	70	72	78	80	63

[Ans: 0.187]

16. An examination of 10 applicants was taken by a company on skill and ability of candidates. From the marks obtained by the applicant in Skill and Ability Papers. Calculate the rank correlation coefficient.

Applicant	A	B	C	D	E	F	G	H	I	J
Marks in Skill	38	41	68	41	38	55	85	81	28	41
Marks in Ability	48	39	38	36	58	61	72	83	61	82

$m_1 = 2, m_2 = 3, m_3 = 2$ [Ans: 0.278]

17. Ten competitors in a competition are ranked by three experts in the following order.

1st expert	1	5	4	8	9	6	10	7	3	2
2nd expert	4	8	7	6	5	9	10	3	2	1
3rd expert	6	7	8	1	5	10	9	2	3	4

Use the rank correlation coefficient to discuss which pair of experts has the nearest approach to competitors. [Ans: 2nd and 3rd]

18. Following data represent the preference of 10 students studying B.Sc. CSIT towards two brands of computers namely DELL and HP.

Computer	Student preference									
DELL	5	2	9	8	1	10	3	4	6	7
HP	10	5	1	3	8	6	2	7	9	4

Apply appropriate statistical tool to measure whether the brand preference is correlated. Also interpret your result. [Ans: -0.309]

19. The coefficient of rank correlation of the marks obtained by 10 students in mathematics and statistics was found to be 0.8. It was discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct rank correlation coefficient. [Ans: 0.78]

20. Coefficient of rank correlation between debenture prices and share prices is found to be 0.143. If the sum of the squares of differences in ranks is given to be 48, find the value of n. [Ans: 7]

21. A large company wants to measure the effectiveness of radio advertising media(s) on the sale promotion (y) of its products. A sample of 22 cities with approximately equal populations is selected for study. The sales of the product in thousand rupees and the level of radio advertising expenditure in thousand rupees are recorded for each of 22 cities. The sum, sum of square and sum of product of x and y are summarized below

$\sum y = 26953$, $\sum x = 950$, $\sum y^2 = 35528893$, $\sum x^2 = 49250$, $\sum xy = 1263940$

a) Fit a simple linear regression model of y on x using the least square method. Interpret the estimated slope coefficient

b) Compute R^2 and interpret. [Ans: y=699.957+12.162x, 0.4852]

22. The following measurements show the respective height in inches of 10 fathers and their eldest sons

Father	67	63	66	71	69	65	62	70	61	72
Son	68	66	65	70	67	67	64	71	62	63

Find the regression line of son's height on father's height and estimate the height of son for the given height of father as 70 inches. Also determine coefficient of determination and interpret. [Ans: y=40.43+0.388x, 67.62, 0.266]

23. The following data gives the experience of machine operators in years and their performance as given by the number of good parts turned out per 100 pieces.

Operator	I	II	III	IV	V	VI	VII	VIII
Experience	16	12	18	4	3	10	5	12
Performance	87	88	89	68	78	80	75	83

Calculate the regression equation of performance on experience and hence estimate the probable performance if an operator has 8 years experiences. Interpret the regression coefficient. [Ans: y=69.669+1.133x, 78.73]

24. A city council has gathered data on number of minor traffic accidents and the number of youth football games that occurred in town over the weekends.

X(football games)	20	30	10	12	15	25	34
Y(minor accidents)	6	9	4	5	7	8	9

(i) Develop the regression equation to predict minor accidents from football games.

(ii) Predict the number of minor traffic accidents that will occur at weekends during which X=30.

(iii) Calculate the value of coefficient of determination

(iv) Calculate the value of standard error of estimate. [Ans: y=2.732+0.198x, 9, 0.87, 0.77]

25. A chemical company wishing to study the effect of extraction time on the efficiency of an extraction operation obtained the data as follows

Extraction time in minute(X)	27	45	41	19	35	39	19
Extraction efficiency in %(Y)	57	64	80	46	62	72	52

a) Fit a straight line to the given data by the method of least square and use it to predict the extraction efficiency one can expect when the extraction time is 35 minutes.

b) Determine the coefficient of determination and interpret its meaning.

[Ans: y=32.096+0.926x, 64.5, 0.843]

26. For the data given below i)Fit linear regression Y = a+bX by the method of least square and interpret regression coefficient ii) determine coefficient of determination and interpret.

X	0	5	10	15	20	25
Y	12	15	17	22	24	30

[Ans: y=11.29+0.697x, 0.97]

27. National Planning Commission(NPC) is performing preliminary study to determine the relationship between certain economic indicator and annual percentage change in Gross National Product(GNP). The concern is to estimate the percentage change in GNP. One of such indicator being examined is government's deficit. Data on 6 years are given below;

Percentage change in GNP	3	1	4	1	2	3
Government deficit in lakh Rs	50	200	70	100	90	40

a) Develop the estimating equation to predict percentage change in GNP from government deficit.

b) Interpret the estimated regression coefficient.

c) What percentage change in GNP would be expected in a year in which government deficit was Rs 110 lakh.

d) Compute the coefficient of determination and interpret.

[Ans: y=3.725 – 0.015x,2.055, 0.524]

28. The annual advertising expenditure (in lakh rs) and the corresponding annual sales (in crore rs) for the past 10 years of a company are presented in the following table.

Year	Annual advertising expenditure	Annual sales revenue
1	10	20
2	12	30
3	14	37
4	16	50
5	18	56
6	20	78
7	22	89
8	24	100
9	26	120
10	28	110

a. Find the correlation coefficient between annual advertising expenditure and annual sales revenue and comment the result

b. Develop the regression model of sales as a function of advertising expenditures.

c. Predict the value of annual sales while advertising expenditure was 27 lakh rupees.

[Ans: 0.985,y=-40.048+5.739x,114.915]

29. Career airline pilots face the risk of progressive hearing loss due to the noisy cockpits of most jet aircrafts. Much of the noise comes not from engines but from air roar which increases at

high speeds. To assess this workplace hazard a pilot measured cockpit noise level(in decibels) and airspeed (knots indicated air speed). The data are shown in the given table

Speed	250	340	320	330	346	260	280	395	380	400
Noise level	83	89	88	89	92	85	84	92	93	96

a. Determine association between noise level and air roar which is increased due to high speed. Comment on strength of association

b. Develop a least square regression model to estimate the noise level with the help of speed of aircraft. Also interpret the regression coefficient.

[Ans: 0.957;y=64.191+0.075x]

30. A computer manager interested to know how efficiency of his/ her new computer program which depends on the size of incoming data. Efficiency will be measured by the number of processed requests per hour. In general, larger data sets require more computer time, and therefore, fewer requests are processed within 1 hour. Applying the program to data sets of different sizes, the following data were gathered.

Data size(gigabytes)	6	7	7	8	10	10	15
Processed requests	40	55	50	41	17	26	16

a. Identify which one response variable and fit a simple regression line assuming that the relationship is linear

b. Interpret the regression coefficient with reference to your problem

c. Obtain the coefficient of determination and interpret this

d. Based on the fitted model predict the efficiency of new computer for data size 12(gigabites). Does it possible to predict efficiency for data size of 30 (gigabites)? Discuss. [Ans: $y = 72.278 - 4.142x$, $R^2 = 0.661$, 22.57, *No*]

Date.
Page No.

Q.No.1 Soln

$r = 0.8$

$\Sigma xy = 600$

S.D of $Y = 6$

and $\Sigma X^2 = 90.9$

or, $\Sigma (X-\bar{X})^2 = 9$.

We know,

$$r = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

or, $$r = \frac{\frac{1}{n}\Sigma (x-\bar{x})(Y-\bar{Y})}{\sqrt{\frac{1}{n}\Sigma (x-\bar{x})^2} \cdot \sigma_y}$$

or, $$r = \frac{\Sigma (x-\bar{x})(Y-\bar{Y})}{\sqrt{n}\sqrt{\Sigma (x-\bar{x})^2} \cdot \sigma_y}$$

or, $$0.8 = \frac{600}{\sqrt{n} \cdot \sqrt{9} \cdot 6}$$

or, $$\sqrt{n} = \frac{600}{0.8 \times 3 \times 6}$$

or, $n = (14.67)^2$

$= 173611$

≈ 1736 //

Date.
Page No.

Q.No.5 Soln

X	Y	X^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135
$\Sigma X = 45$	$\Sigma Y = 108$	$\Sigma X^2 = 285$	$\Sigma Y^2 = 1356$	$\Sigma XY = 547$

Here $n = 9$.

Now, Correlation coefficient $r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$

$= \frac{9 \times 547 - 45 \times 108}{\sqrt{9 \times 285 - (45)^2}\sqrt{9 \times 1356 - (108)^2}}$

$= 0.95$.

Q.No.7 Soln

X	Y	$U = X - A = X - 25$	$V = Y - B = Y - 17$	U^2	V^2	UV
25	17	0	0	0	0	0
26	18	1	1	1	1	1
27	19	2	2	4	4	4
25	17	0	0	0	0	0
26	19	1	2	1	4	2
28	20	3	3	9	9	9
25	17	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
		$\Sigma U =$	$\Sigma V =$	$\Sigma U^2 =$	$\Sigma V^2 =$	$\Sigma UV =$

Date. _____
Page No. _____

$r = 0.654$

$n = 19$

For testing significance correlation coefficient.

$$P.E = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - (0.654)^2}{\sqrt{19}}$$

$$= 0.08$$

Here, $|r| = 0.654$

and $6 \times P.E = 6 \times 0.08 = 0.48$.

Since $|r| = 0.654 > P.E = 0.48$ So, correlation coeff. of 'r' is significant.

Q.No.6 Soln.

$\frac{150}{200} \times 100 = 75$

Size of Group	Mid value (X)	$U = \frac{X-A}{h}$	Y	$V = Y - B$	U^2	V^2	$U'V'$
15-20	17.5	-2	75	5	4	25	-10
20-25	22.5	-1	60	2	1	4	-2
25-30	27.5	0	50	0	0	0	0
30-35	32.5	1	50	0	1	0	0
35-40	37.5	2	45	-1	4	1	-2
40-45	42.5	3	40	-2	9	4	-6
		$\Sigma U' = 3$		$V' = 4$	$\Sigma U'^2 = 19$	$\Sigma V'^2 = 34$	$\Sigma U'V' = -20$

Now, $$r = \frac{n\Sigma U'V' - \Sigma U' \Sigma V'}{\sqrt{n\Sigma U'^2 - (\Sigma U')^2} \cdot \sqrt{n\Sigma V'^2 - (\Sigma V')^2}}$$

$$= \frac{6 \times (-20) - 3 \times 4}{\sqrt{6 \times 19 - 9} \sqrt{6 \times 34 - 16}}$$

$$= -$$

Date.
Page No.

~~V. Imp~~

Q.No.8 Sol^n

Given,

no. of pairs $(n) = 20$.

$\Sigma X = 127$, $\Sigma Y = 100$.

$\Sigma X^2 = 860$

$\Sigma Y^2 = 549$.

$\Sigma XY = 674$

Correct values of X are 8, 6

Incorrect values of X are 10, 8.

Correct values of Y are 12, 8

and Incorrect values of Y are 14, 6.

For the calculation of correct correlation coefficient.

$(\Sigma X)_{correct} = 127 - 10 - 8 + 8 + 6$

(Subtract incorrect values; add correct values)

$= 123$.

$(\Sigma X)_{incorrect} = 100 - 14 - 6 + 12 + 8$

$= 100$

$(\Sigma Y^2)_{correct} = 549 - 14^2 - 6^2 + 12^2 + 8^2$

$= 525$

$(\Sigma X^2)_{correct} = 860 - 10^2 - 8^2 + 8^2 + 6^2$

$= 796$

$(\Sigma XY)_{correct} = 674 - 14 \times 10 - 6 \times 8 + 12 \times 8 + 6 \times 8$

$= 630$

$\therefore$ Correct correlation coeffiient.

$$(r)_{correct} = \frac{n(\Sigma XY)_{correct} - (\Sigma X)_{correct}(\Sigma Y)_{correct}}{\sqrt{n(\Sigma X^2)_{correct} - (\Sigma X)^2_{correct}}\ \sqrt{n(\Sigma Y^2)_{correct} - (\Sigma Y^2)_{correct}}}$$

(Correct value of r)

$$= \frac{20 \times 630 - 123 \times 100}{\sqrt{20 \times 796 - (125)^2}\ \sqrt{20 \times 525 - 100}}$$

$= 0.477$ //

Date.
Page No.

Spearman's Rank Correlation: (Rank Correlation)

Rank correlation measures the degree of relationship between two attributes (or qualitative variable) It is denoted by r_s. Rank Correlation coefficient is also known as spearman's Rank Correlation. It is used to find the correlation coefficient between qualitative variables such as beauty, knowledge, honesty etc. which can not be measured quantitively directly and the Rank correlation can be calculated in the following conditions:-

(a) If the ranks are given.
(b) If the ranks are not given and not repeated.
(c) If the ranks are not given and repeated.

If ranks are given

then,

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

where, $d = R_1 - R_2$ (difference between the two ranks).
n = no. of pairs of observation.
& ($\Sigma d = 0$ always).

If ranks are not given and not repeated

For the calculation of r_s in this case first we need to assign rank to all values in different series then,

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$ (same formula rank)

Date.
Page No.

If the ranks are not given and repeated.

Then,

$$r_s = 1 - \frac{6\left[\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \dots\right]}{n(n^2-1)}$$

where, $m_1, m_2, \dots$ are the no of repeatation of values.

Numerical Questions:

Q.No.12 Soln

For the calculation of Rank Correlation coefficient

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

Rank of A (R_1)	Rank of B (R_2)	$d = R_1 - R_2$	d^2
3	6	-3	9
5	4	1	1
8	9	-1	1
4	8	-4	16
7	1	6	36
10	2	8	64
2	3	-1	1
1	10	-9	81
6	5	1	1
9	7	2	4
		$\Sigma d = 0$.	$\Sigma d^2 = 214$

Now,

$$r_s = 1 - \frac{6 \times 214}{10(10^2-1)}$$

$$= \frac{1 - 1284}{10 \times 99}$$

$$= -0.296 //$$

Date.
Page No.

Q.No.13 Sol^{n}

For the calculation of Rank Correlation coefficient.

X	Y	R_1	R_2	$d = R_1 - R_2$	d^2
10	100	1	1	0	0
15	120	3	5	−2	4
18	118	4	3	1	1
14	119	2	4	−2	4
30	130	6	6	0	0
27	105	5	2	3	9
				$\Sigma d = 0$	$\Sigma d^2 = 18$

Now,

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 18}{6 \times 35}$$

$$= 0.48 //$$

Date.
Page No.

Now,

$$r_s = 1 - 6\left[\frac{1315 + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}}{10\times 99}\right]$$

$$= 1 - 0.813$$

$$= 0.187.$$

Ans

Q.No.17 Sol^n. For the calculation of rank correlation coefficient between 1st and 2nd, 1st and 3rd & 2nd and 3rd pair of experts.

We have,

$$r_{12} = r_3 = 1 - \frac{6\Sigma d_1^2}{n(n^2-1)}$$

where, $\Sigma d_1 = 0$.

Similarly, $r_{13} = 1 - \frac{6\Sigma d_2^2}{n(n^2-1)}$ where, $\Sigma d_2 = 0$.

& $r_{23} = 1 - \frac{6\Sigma d_3^2}{n(n^2-1)}$ where, $\Sigma d_3 = 0$.

R_1	R_2	R_3	$d_1 = R_1 - R_2$	$d_2 = R_1 - R_3$	$d_3 = R_2 - R_3$	d_1^2	d_2^2	d_3^2
1	4	6	−3	−5	−2	9	25	4
5	8	7	−3	−2	1	9	4	1
4	7	8	−3	−4	−1	9	16	1
8	6	1	2	7	5	4	49	25
9	5	5	4	4	0	16	16	0
6	9	10	−3	−4	−1	9	16	1
10	10	9	0	1	1	0	1	1
7	3	2	4	5	1	16	25	1
3	2	3	1	0	−1	1	0	1
2	1	4	1	−2	−3	1	4	9
			$\Sigma d_1 = 0$	$\Sigma d_2 = 0$	$\Sigma d_3 = 0$	$\Sigma d_1^2 = 74$	$\Sigma d_2^2 = 156$	$\Sigma d_3^2 = 44$

Date. _____
Page No. _____

Now,

$$r_{12} = 1 - \frac{6 \times 74}{10 \times 99} = 0.55$$

$$r_{13} = 1 - \frac{6 \times 156}{990} = 0.05$$

and $$r_{23} = 1 - \frac{6 \times 44}{990} = 0.733$$ //

Since $r_{23} > r_{12}$ and r_{13}. So, 2nd and 3rd experts have nearest approach to competitors.

Q.No.19 Soln

Given, no. of pair of observation (n) = 10.

rank corr. coeff (r_s) = 0.8.

Correct d = 9.

Incorrect d = 7.

We have,

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

or, $$0.8 = 1 - \frac{6 \times \Sigma d^2}{10 \times 99}$$

or, $\Sigma d^2 = 33$

Now, the $(\Sigma d^2)_{correct} = 33 - 7^2 + 9^2$ (subtract incorrect, add correct.)

$= 65$ //

Now, the correct rank correlation coefficient

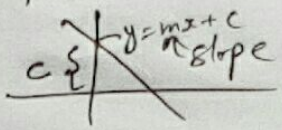
$$(r_s)_{correct} = 1 - \frac{6(\Sigma d^2)_{correct}}{n(n^2-1)}$$

$$= 1 - \frac{65}{10 \times 99}$$

$$= 0.606$$ //.

$\frac{dx^n}{dx} = nx^{n-1}$

$Y = a + bx$ similar to $y = mx + c$



Date.
Page No.

⊛ Regression:

The correlation measures the degree of relationship between two or more than two variable but the regression shows the functional relationship between two or mor than two variable. If we study about only two variable then the regression is called linear or simple regression. By the help of regression line or regression equation we can able to estimate or predict the value of dependent variable with the help of independent variable.

Let a regression line of Y on X (Y depends on X) is $Y = a + bX$ where Y is dependent and X is independent variable & b is the slope or the regression coefficient of Y on X i.e, b_{yx} and a is the constant or y-intercept.

Similarly $X = a' + b'y$ where

b' = reg. coeff. of X on Y i.e, b_{xy}.

Important formulas

i) $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$.

ii) $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$

iii) $b_{xy} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma Y^2 - (\Sigma Y)^2}$

Date. _____
Page No. _____

⊛ Measures of Variation:

Total sum of square (TSS) = Sum of square due to regression (SSR) + Sum of square due to error (SSE)

$$\boxed{\text{i.e. } TSS = SSR + SSE}$$

where,

$$TSS \text{ or } SST = \Sigma(Y-\bar{Y})^2$$
$$= \Sigma Y^2 - n\bar{Y}^2$$

$$SSE = \Sigma(Y-\hat{Y})^2$$
$$= \Sigma Y^2 - a\Sigma Y - b\Sigma XY.$$

Then $\boxed{SSR = TSS - SSE}$

⊛ Coefficient of Determination

Coefficient of determination $(R^2) = \frac{SSR}{SST}$

or, $R^2 = \frac{SST - SSE}{SST}$

or, $\boxed{R^2 = 1 - \frac{SSE}{SST}}$

Interpretation: Let coefficient of determination R^2 is 0.81 this means total variation on dependent variable Y that is explained by independent variable X is 81% and remaining 19% variation on dependent variable Y is due to the effect of other independent variable.

⊛ Standard Error of the Estimate:

Standard error of estimate $(S.E) = \sqrt{\frac{SSE}{n-k-1}}$

where k is no. of independent variable. (here in simple or linear form of regression $k=1$)

Date.
Page No.

Numerical Questions:

Q.No.21 Soln

Let a regression of Y on X is $Y = a + bX$ — (I)

where Y → sale promotion of product

X → advertising media.

For finding the value of 'a' and 'b' we need to solve the following normal equation.

$$\Sigma Y = na + b\Sigma X \quad \text{— (II)}$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 \quad \text{— (III)}$$

Putting the values of ΣX, ΣY, ΣX^2, ΣXY and n in eqn (II) and (III)

$$22a + 950b = 26953 \quad \text{— (IV)}$$

$$950a + 49250b = 1263940 \quad \text{— (V)}$$

Solving (IV) and (V) by multiplying eqn (IV) by 950 and eqn (V) by 22.

$$950 \times 22a + 950 \times 950b = 26953 \times 950$$

$$950 \times 22a + 49250 \times 22b = 1263940 \times 22$$

$$182000b = 2201330$$

$$b = \frac{2201330}{182000}$$

$$= 12.16$$

Putting value of b in eqn (IV) we get,

$$a = 699.957$$

(a) From eqn (I)

$$\boxed{Y = 699.957 + 12.16X}$$

Here slope coefficient $(b_{yx}) = b = 12.16$

(b) Here,

$$R^2 = \frac{SSR}{TSS \text{ (or } SST)}$$

$$= 1 - \frac{SSE}{SST}$$

$\frac{\Sigma Y}{n} = \bar{Y}$

Date.
Page No.

where, $SST = \Sigma(Y-\bar{Y})^2 = \Sigma Y^2 - n\bar{Y}^2$

$= 35528893 - 22\left(\frac{25953}{22}\right)^2$

$= 2507792.59$

and $SSE = \Sigma y^2 - a\Sigma Y - b\Sigma XY$

$= 35528893 - 699.957 \times 26953 - 1216 \times 263940$

$= 1293441.57$

Now,

$R^2 = 1 - \frac{1293441.57}{2507792.59}$

$= 1 - 0.5157$

$= 0.4843$

$= 48.43\%$

This means total variation on dependent variable is 48.43 which is explained by independent variable X and remaining variation is due to is dependent upon Y is due to the effect of other independent variable. 51.52

Q.No.22 Soln

Given no. of pairs of observation $(n) = 10$.

Let $X \rightarrow$ height of father.

and $Y \rightarrow$ height of son.

Now, regression line of hight of son (Y) on hight of father (X) is $Y = a + bX$ — (I).

Let deviation, $U = X - A$

and $V = Y - B$

then, $V = a + bU$ — (II) for finding the value of a and b we need to solve the following normal equation.

$$\Sigma V = na + b\Sigma U \quad \text{(iii)}$$

$$\Sigma UV = a\Sigma U + b\Sigma U^2 \quad \text{(iv)}$$

For the calculation of $\Sigma U, \Sigma V, \Sigma UV, \Sigma U^2, \Sigma V^2$.

X	Y	$U = X - A$ $= X - 69$	$V = Y - B$ $= Y -$	U^2	V^2	UV
67	68	−2	1	4	1	−2
63	66	−6	−1	36	1	6
66	65	−3	−2	9	4	6
71	70	2	3	4	9	6
69	67	0	0	0	0	0
65	67	−4	0	16	0	0
62	64	−7	−3	49	9	21
70	71	1	4	1	16	4
61	62	−8	−5	64	25	40
72	63	3	−4	9	16	−12
		$\Sigma U = -24$	$\Sigma V = -7$	$\Sigma U^2 = 192$	$\Sigma V^2 = 81$	$\Sigma UV = 69$

Now, putting value of $\Sigma U, \Sigma V, \Sigma U^2$ in eqn (iii) and (iv).

$$10a - 24b = -7 \quad \text{(v)}$$

$$-24a + 192b = 69 \quad \text{(vi)}$$

Putting. Solving eqn (v) and (vi) by multiplying eqn (v) by 8.

$$80a - 192b = -56$$

$$-24a + 192b = 69$$

$$56a = 13$$

$$\therefore a = \frac{56}{13}$$

$$= 0.232$$

Putting value of (a) in eqn (i)

$$b = 0.388$$

Date.
Page No.

Now,

$V = a + bU$

or, $Y - B = a + b(X - A)$

or, $Y - 67 = 0.232 + 0.388(X - 69)$

or, $Y - 67 = 0.232 + 0.388X - 69 \times 0.388$

or, $\boxed{Y = 40.46 + 0.388X}$

When height of father $(X) = 70$ then the height of son $Y = 40.46 + 0.388 \times 70$

$= 676.2$

We know, Correlation coefficient $= \frac{n\Sigma UV - \Sigma U \Sigma V}{\sqrt{n\Sigma U^2 - (\Sigma U)^2}\sqrt{n\Sigma V^2 - (\Sigma V)^2}}$

..... this is r.

and we find coeff. of determination by finding r^2 directly not using SST and SSE.

Q.N.29 Soln

For the calculation of degree of association between noise level (Y) and speed (X) then correlation coefficient $(r) = \frac{n\Sigma UV - \Sigma U \Sigma V}{\sqrt{n\Sigma U^2 - (\Sigma U)^2}\sqrt{n\Sigma V^2 - (\Sigma V)^2}}$

where, $U = X - A$

$V = Y - B$

and regression model, $V = a + bU$ — (i)

for finding the value of 'a' and 'b' we need to solve the following equations

$\Sigma V = na + b\Sigma U$ — (ii)

$\Sigma UV = a\Sigma U + b\Sigma U^2$ — (iii)

28409

Speed (X)	Noise Level (Y)	$U = X - A$, $A = 346$	$V = Y - B$, $B = 85$	U^2	V^2	UV
250	83	-96	-2	9216	4	
340	89	-6	4	36	16	
320	88	-26	3	676	9	
330	89	-16	4	256	16	
346	92	0	7	0	49	
260	85	-80	0	6400	0	
280	84	-66	-1	4356	1	
395	92	49	7	2401	49	
380	93	34	8	1156	64	
400	96	54	11	2916	121	
		$\Sigma U = -159$	$\Sigma V = 41$	$\Sigma U^2 = 28409$	$\Sigma V^2 = 329$	$\Sigma UV = 1302$

Now,

$$r = \frac{10 \times 1302 - (-159) \times (41)}{\sqrt{10 \times 28409 - (-159)^2}\sqrt{10 \times 329 - (41)^2}}$$

$$= 0.957$$

Putting the value of $\Sigma U, \Sigma V, \Sigma UV, \Sigma U^2$ and n in (iii) and (iv)

$$10a - 159b = 41 \quad \text{(iv)}$$

$$-159a + 28409b = 1302 \quad \text{(v)}$$

Solving eqn (iv) and (v) by multiplying eqn (iv) by 159 and (v) by 10.

$$1590a - 159 \times 159b = 41 \times 159$$

$$-1590a + 284090b = 13020$$

$$258809b = 19539$$

$$\text{or, } b = \frac{19539}{258809}$$

$$\text{or, } b = 0.075$$

Date. _____
Page No. _____

Putting the value of b in eqn (iv)

$10a - 159 \times 0.075 = 41$

or, $10a = 41 + (159 \times 0.075)$

or, $a = \frac{41 + (159 \times 0.075)}{10}$

or, $a = 5.29$

Now From eqn (i)

$V = 5.29 + b.0.075.U$

$Y - 85 = 5.29 + 0.075(X - 346)$

or, $Y - 85 = 5.29 + 0.075X - 25.95$

or, $\boxed{Y = 64.34 + 0.075X}$

(a) Since $r = 0.957 = 95.7\%$ so, there is high degree of association between noise level and speed.

(b) Since regression coefficient $b = 0.075$, It shows that there is increase in noise level by 0.075 as per unit change in the speed.