Date ____
Page ____

Module - 04
Correlation Analysis.

Syllabus:-
Correlation Analysis

* Meaning and types of correlation.
* Karl pearsons, co-efficient of correlation and calculation of probable error.

CORRELATION MEANING:-

A statistical relationship between two variable such that a systematic change in one variable is accompanied by systematic change in another.

Correlation refers to the technique used in measuring the closeness or degree of the relationship between two quantitative variables. sometimes it is termed as co-variations.

DEFINITIONS:-

* According to A. M. Tuttle, "correlation is the analysis of the co-variations between two or more variables".

* According to Ya-lun-chou, "correlation analysis attempts to determine the degree of relationship between variables".

* According to croxton and cowden, "when the relationship between two variables is of a quantitative nature, the appropri-ate the statistical tool for discoveri-ng and measuring the relationship and expressing in brief formula is known as correlation".

USES:-

* It helps in measuring the relationship between the variables.
* It helps in analyzing the economic behaviour such as salary and inflation, price and demand, output & input etc.
* It facilitates comparison and its relative measure.
* correlation is used for the prediction of future.
* The predication based on correlation analysis will be more reliable.

TYPES OF CORRELATION:-

(i) Positive and Negative correlation.
(ii) Simple and multiple correlation.
(iii) Partial and Total correlation.
(iv) Linear and Non-linear correlation.
(v) logical and illogical correlation.

(i) POSITIVE AND NEGATIVE CORRELATION:-

On the basis of direction in which the variables move, the correlation maybe called as positive or negative.

"If both the variable is moving in the same direction, correlation is said to be positive or direct". If one of the two variables is increasing and the other on an average, is also increasing (or) If one of the two variables is decreasing and the other on an average, is also decreasing. Then the relationship said to be positively correlated.

Ex:-

No. of days of work	1	2	3	4	5	6
wages earned	100	200	300	400	500	600

"If both the variable is moving in the opposite direction, correlation is said to be negative or inverse". If one of the two variables increasing and the other variable is decreasing (or) If one of the two variables is decreasing and the other variable is increasing; then the relationship is said to be negatively correlated.

Example:-

Price	10	20	30	40	50	60
Demand	100	90	80	70	60	50

(ii) SIMPLE AND MULTIPLE CORRELATIONS:-

when we have 'n' number of variables, from that only two variables are studied is called as simple correlation.

Ex:- Study of 1^{st} sem B.Com results (only Pass and fail are the variables).

when the study is done wholly or more than two variables are studied for correlating, such a case it is called as multiple correlation.

Ex:- Study of marks obtained in FA, IFIM, BDE, CSA and IC at a time with pass and fail.

(iii) PARTIAL AND TOTAL CORRELATIONS:-

when there are more than two variables are given, in that the relationship between any two variable in studied is called as partial correlation.

The correlation between all the variable under the study or taken together for measuring co-efficient of correlation, such correlation is known as total correlation.

Ex:- Agricultural production needs land, fertilizers, Pesticides, Irrigation etc. If every variables is taken into account then it is total correlation and if only two variables is taken into account or study then it is a partial correlation.

(iv) LINEAR AND NON-LINEAR CORRELATION:-

If the ratio of change in one variable tends to bear a constant ratio of change in another variable then the correlation is said to be linear correlation.

X	1	2	3	4	5
Y	50	100	150	200	250

If the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable is said to be Non-linear correlation.

X	1	3	4	6	10
Y	100	185	254	362	472

(v) LOGICAL AND ILLOGICAL CORRELATION:-

when correlation between two variables is not only mathematically defined, but also logically sound too it is called as logical correlation.

Ex:- Correlation between the price and demand i.e, price of petrol and sales of vehicles.

when correlation between two unrelated variables which can be determined mathematically, yet or not based on any logic is called as illogic correlation.

Ex:- correlation between the price and demand i.e., price of petrol and sales of mobiles.

METHODS OF STUDY OF CORRELATION:-

* Scatter diagram method
* Graphical Method
* Karl pearson's coefficient of correlation.
* Spearmans Rank correlation.
* Concurrent deviation method.

SCATTER DIAGRAM METHOD:-

It is also known as dot chart, dot gram and scatter gram. It is also a graphical method, the term scatter means dispersion or spread of dots on the graph.

ADVANTAGES:-

* It is easy to allocate dot on graph.
* It is best fit as free hand method.
* It is used to form a rough idea abort the nature of the relationship between two variables.

DISADVANTAGES:-

* It is not suitable for the number of observation is large.
* It does not provide exact measure of the extent of the relationship between the variables.

GRAPHICAL METHOD:-

It is also known as correlogram or simple graph method. It is the simplest method to determine the presence of correlation between two variables. It is easy to determine the correlation negati-ve or positive.

Advantages:- we can determine the nature and extent of correlation from a graph.

Disadvantages:-

we cannot know the exact degree of correlation from this method.

KARL PEARSON'S CO-EFFICIENT OF CORRELATION:-

Karl pearson (1867 - 1936) was a reputed biometrician and statistication. In 1890, he developed a formula on co-efficient of correlation. After his name the method is called as karl pearson's co-efficient of correlation and denoted by 'r'.

It is widely used as mathematical method where in the numerical expression is used to calculate the degree and directions of the relationship between linear related variable.

It is defined as the ratio of the co-variance between x and y variable to the product of standard deviation of x and y.

CHARACTERISTICS:-

* Based on arithmetic mean and standard deviation.
* Determine the directions of the relationship.
* Establishes the size(degree) of the relationship.
* Ideal Measure (Prefect measurement).

FORMULA:-

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

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

$$r = \frac{\Sigma xy}{N \sigma_x \sigma_y}$$

Date
Page

STEPS TO BE FOLLOWED:-

1. Calculate actual means of the given two series of variables.
2. Take the deviations of each variables of x series from actual mean and denote them as 'x'.
3. Take the deviations of each variables of Y series from actual mean and denote them as 'y'.
4. Square those deviations.
5. Then find the product of each deviation ($x \times y = xy$)
6. Sum up $x^2 y^2$ and xy. That means find $\Sigma x^2 \Sigma y^2$ and Σxy.

ASSUMPTION:-

* Each variable produces a normal distribution.
* The relationship is linear.

DISADVANTAGES:-

* It is based on the assumption that there is linear relationship between the variables without verifying whether it is correct or not.
* More time consuming.
* It is affected by extreme points.

Kavitha

PROBABLE ERROR:-

Probable error of a co-efficient of correlation is an amount which if added to or subtracted from the value given two limits within which the correlation co-efficient obtained will fall. It is 0.6745 times of the standard error of 'r'.

It is an instrument which measures the reliability and dependability of the value 'r'. This method is used in interpreting whether 'r' is significant or not.

a) If the value of r is more than six times the probable error, the existence of correlation is certain = the value of 'r' is significant.

b) If the value of r is less than six times the probable error, there is no evidence of correlation = the value of 'r' is not at all significant.

INTERPRETATION OF CORRELATION CO-EFFICIENT :-

value of correlation co-efficient	Interpretation.
a) when $r = +1$	Perfect positive correlation
b) when $r = -1$	perfect negative correlation.
c) when r is in between $+0.75$ to $+1$	High degree positive correlation.
d) when r is in between -0.75 and -1	High degree negative correlation.
e) when r is in between $+0.25$ and $+0.75$	Moderate degree positive correlation.
f) when r is in between -0.25 and -0.75	Moderate degree negative correlation.
g) when r is in between 0 and $+0.25$	low degree positive correlation.
h) when r is in between 0 and -0.25	low degree negative correlation.
i) when $r = 0$	No correlation.

PROBLEMS:-

1) Find the correlation co-efficient between sales and advertising expenditure from the following data.

Sales (₹ in lakh)	Advertising exp's (in 000's)	x (X − X̄)	x^2	y (Y − Ȳ)	y^2	xy
65	66	−4	16	−2	4	8
66	67	−3	9	−1	1	3
67	64	−2	4	−4	16	8
68	67	−1	1	−1	1	1
69	71	0	0	3	9	0
70	69	1	1	1	1	1
71	70	2	4	2	4	4
72	68	3	9	0	0	0
73	70	4	16	2	4	8
Σx = 621	Σy = 612	✓	60		40	33

$$\bar{x} = \frac{\Sigma x}{N} = \frac{621}{9} = 69$$

$$\bar{y} = \frac{\Sigma y}{N} = \frac{612}{9} = 68$$

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

$$= \frac{33}{\sqrt{60 \times 40}}$$

$$= \frac{33}{\sqrt{2400}}$$

$$= \frac{33}{48.98} = 0.67$$

∴ Moderate positive correlation.

Their exist a moderate degree positive correlation between sales and advertising expenditure.

4) calculate karl pearsons correlation co-efficient from the following

X	Y	x	x^2	y	y^2	xy
5	9	-12	144	-9	81	108
7	11	-10	100	-7	49	70
11	14	-6	36	-4	16	24
14	14	3	9	-4	16	12
17	17	0	0	-1	1	0
19	24	2	4	6	36	12
23	21	6	36	3	9	18
27	25	10	100	7	49	70
30	27	13	169	9	81	117
153	162		598		338	431

$$\bar{X} = \frac{\Sigma x}{N} = \frac{153}{9} = 17$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{162}{9} = 18$$

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

$$= \frac{431}{\sqrt{598 \times 338}} = \frac{431}{\sqrt{202124}} = \frac{431}{449.58}$$

Kavitha

$= 0.9586 / 0.96$

$\therefore$ Their exist high degree positive correlation.

34) Calculate karl pearson's correlation co-efficient from the following.

Price	demand	x	x^2	y	y^2	xy
21	20	-4	16	3	9	-12
22	19	-3	9	2	4	-6
23	19	-2	4	2	4	-4
24	17	-1	1	0	0	0
25	17	0	0	0	0	0
26	16	1	1	-1	1	-1
27	16	2	4	-1	1	-2
28	15	3	9	-2	4	-6
29	14	4	16	-3	9	-12
225	153		60		32	-43

$\bar{X} = \frac{\Sigma x}{N} = \frac{225}{9} = 25$

$\bar{Y} = \frac{\Sigma y}{N} = \frac{153}{9} = 17$

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

$= \frac{-43}{\sqrt{60 \times 32}}$

$= \frac{-43}{\sqrt{1920}}$

$= \frac{-43}{43.82}$

$= -0.98$

∴ Their exist High degree negative correlation.

CHANGE OF ORIGIN AND SCALE :-

while calculating correlation co-efficient, variable of large value can be reduced to smaller value by dividing each of them by a common factor. Since 'r' is a pure number, shifting the origin and changing the scale of the given series do not affect its value.

Ab calculate the karl pearsons correlation coefficient and comment:-

X	Y
5000	1000
10000	2000
15000	3000
20000	4000
25000	5000
30000	6000
35000	7000

Kavitha
Date
Page

solⁿ:

NOTE:- Here, the variables are large values. therefore, the variables will be reduced to its simplest value by dividing variable X by 5000 and variable Y by 1000 as a common factor to find out correlation co-efficient.

X	Y	x	x^2	y	y^2	xy
1	1	-3	9	-3	9	9
2	2	-2	4	-2	4	4
3	3	-1	1	-1	1	1
4	4	0	0	0	0	0
5	5	1	1	1	1	1
6	6	2	4	2	4	4
7	7	3	9	3	9	9
28	28		28		28	28

$$\bar{X} = \frac{\Sigma X}{N} = \frac{28}{7} = 4$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{28}{7} = 4$$

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

$$= \frac{28}{\sqrt{784}} = \frac{28}{28} = 1$$

∴ Their exist a perfect positive correlation.

Kavitha

5) calculate the correlation of co-efficient and comment

x	Y
3750	5000
7500	7000
2500	9000
1250	5000
6250	8000
8750	8000

NOTE:- Here, the variable are large value. Therefore, the variables will be reduced to its simplest value by dividing variable x by 1250 and variable Y by 1000 as a common factor to findout correlation co-efficient.

x	Y	x	x^2	y	y^2	xy
3	5	-1	1	-2	4	2
6	7	2	4	0	0	0
2	9	-2	4	2	4	-4
1	5	-3	9	-2	4	6
5	8	1	1	1	1	1
7	8	3	9	1	1	3
24	42		28		14	8

$$\bar{X} = \frac{\Sigma x}{N} = \frac{24}{6} = 4$$

$\bar{y} = \frac{\Sigma y}{n} = \frac{42}{6} = 7$

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

$= \frac{8}{\sqrt{28 \times 14}}$

$= \frac{8}{\sqrt{392}} = \frac{8}{19.79} = 0.4042$

∴ Their exist moderate degree positive correlation

PRODUCT MOMENT METHOD:-

16 calculate the correlation coefficient.

x	y	x^2	y^2	xy
1	2	1	4	2
3	3	9	6	9
5	4	25	16	20
7	3	49	9	21
9	2	81	4	18
11	4	121	16	44
36	18	286	58	114

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

$$= \frac{6(114) - (36 \times 18)}{\sqrt{6(286) - (36)^2}\sqrt{6(58) - (18)^2}}$$

$$= \frac{684 - 648}{\sqrt{1716 - 1296}\sqrt{348 - 324}}$$

$$= \frac{36}{\sqrt{420}\sqrt{24}}$$

$$= \frac{36}{20.49 \times 4.9}$$

$$= \frac{36}{100.4}$$

$$= 0.3585$$

∴ Their exist moderate degree positive correlation.

ASSUMED MEAN METHOD:-

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

N = Total no of observation
A = Assumed mean
dx = x - A = deviations of variables x from their assumed mean.
dy = Y - A = deviations of variables y from their assumed mean.
Σdxdy = Sum of product of dx and dy.
Σdx = Sum of deviations of dx variables.
Σdy = Sum of deviations of dy variables.
Σdx^2 = Sum of squares of dx^2.
Σdy^2 = Sum of squares of dy^2.

PROBLEMS:-

1) From the following data calculate co-efficient of correlation and interpret the value

x	y	dx	dx^2	dy	dy^2	dxdy
38	10	0	0	0	0	0
40	21	2	4	11	121	22
42	15	4	16	5	25	20
44	12	6	36	2	4	12
44	14	6	36	4	16	24
45	13	7	49	3	9	21
46	15	8	64	5	25	40
48	12	10	100	2	4	20
50	20	12	144	10	100	120
397	132	55	449	42	304	279

$$r = \frac{N\Sigma dxdy - \Sigma dx \cdot \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\ \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{9(279) - (55)(42)}{\sqrt{9(449) - (55)^2}\ \sqrt{9(304) - (42)^2}}$$

$$= \frac{2511 - 2310}{\sqrt{4041 - 3025}\ \sqrt{2736 - 1764}}$$

$$= \frac{2511 - 2310}{\sqrt{1016}\ \sqrt{972}}$$

$$= \frac{201}{31.87 \times 31.17}$$

$$= \frac{201}{993.38}$$

$$= 0.2023$$

∴ There exist low degree positive correlation.

PROBLEMS ON GROOPED DATA:-

Q) From the data given in the following data, Find the karl pearson's correlation co-efficient between Age and playing habits of students.

Age (in yrs)	No. of students	regular player	Playing habit(y)
15	250	200	80%
16	200	150	75%
17	150	90	60%
18	120	48	40%
19	100	30	30%
20	80	12	15%

(A=15) A=15

x	y	dx	dx^2	dy	dy^2	dxdy
15	80	0	0	65	4225	0
16	75	1	1	60	3600	60
17	60	2	4	45	2025	90
18	40	3	9	25	625	75
19	30	4	16	15	225	60
20	15	5	25 ~~20~~	0	0	0
105	300	15	55	210	10700	285

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{6(285) - 15(210)}{\sqrt{6(55) - (15)^2}\sqrt{6(10700) - (210)^2}}$$

$$= \frac{1710 - 3150}{\sqrt{\cancel{300}\ 330 - 225}\sqrt{64200 - 44100}}$$

$$= \frac{-1440}{\sqrt{75}\ \sqrt{20100}}$$

$\sqrt{105}$

$$= \frac{-1440}{\cancel{8.66} \times 141.77}$$

10.25

$$= \frac{-1440}{\cancel{1227.72}}$$

1453.1425

$$= \cancel{-1.17}$$

$$-0.9909$$

3) The following data relates to the percentages of failure in the 10th std examination. 0.67

Age of candidate	Percentage of failure
13	39
14	40
15	43
16	34
17	36
18	39
19	48
20	47
21	52

Soln :- Actual mean.

x	y	x	x^2	y	y^2	xy
13	39	-4	16	-3	9	12
14	40	-3	9	-2	4	6
15	43	-2	4	1	1	-2
16	34	-1	1	-8	64	8
17	36	0	0	-6	36	0
18	39	1	1	-3	9	-3
19	48	2	4	6	36	12
20	47	3	9	5	25	15
21	52	4	16	10	100	40
153	378		60		284	88

$$\bar{x} = \frac{\Sigma x}{N} = \frac{153}{9} = 17$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{378}{9} = 42$$

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

$$= \frac{88}{\sqrt{60 \times 284}}$$

$$= \frac{88}{\sqrt{17040}} = \frac{88}{130.54} = 0.6741$$

∴ their exist moderate degree positive correlation.

⇒ Assumed mean.

x	y	dx	dx^2	dy	dy^2	dxdy
13	39	0	0	0	0	0
14	40	1	1	1	1	1
15	43	2	4	4	16	8
16	34	3	9	-5	25	-15
17	36	4	16	-3	9	-12
18	39	5	25	0	0	0
19	48	6	36	9	81	54
20	47	7	49	8	64	56
21	52	8	64	13	169	104
153	378	36	204	27	365	196

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\ \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{9(196) - (36)(27)}{\sqrt{9(204) - (36)^2}\ \sqrt{9(365) - (27)^2}}$$

$$= \frac{1764 - 972}{\sqrt{1836 - 1296}\ \sqrt{3285 - 729}}$$

$$= \frac{792}{\sqrt{540}\ \sqrt{2556}}$$

$$= \frac{792}{23.24 \times 50.56} = \frac{792}{1175.01} = 0.6740$$

∴ Their exist moderate degree positive correlation.

Kavitha

PROBABLE ERROR:-

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$SE = \frac{1 - r^2}{\sqrt{N}}$$

1. $r = 0.67 \quad N = 9$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.67)^2}{\sqrt{9}}$$

$$= 0.6745 \times \frac{1 - 0.4489}{3}$$

$$= 0.6745 \times \frac{0.5511}{3}$$

$$= 0.6745 \times 0.1837$$

$$PE = 0.1239$$

$$r \pm PE$$

$r + PE$ | $r - PE$
$0.67 + 0.1239 =$ | $0.6 - 0.1239$
0.7939 | 0.67 | 0.4761

Interpretation of 'r'

PE = 0.1239 × 6 = 0.7434

$\frac{r}{PE} = \frac{0.67}{0.1239} = 5.41$ times

∴ The r value may fall between the 0.54 and 0.79. Since, r is lesser than the six times of PE - r is insignificant.

vi) $r = 0.96$; $N = 9$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.96)^2}{\sqrt{9}}$$

$$= 0.6745 \times \frac{1 - 0.9216}{3}$$

$$= 0.6745 \times \frac{0.0784}{3}$$

$$= 0.6745 \times 0.0261$$

$$= 0.0176$$

Kavitha

$r \pm PE$

$r + PE$ $\quad$ $r - PE$

$0.96 + 0.0176$ $\quad$ $0.9 - 0.0176$

0.9776 — 0.96 — 0.9424

⇒ Interpretation of 'r'

$PE = 0.0176 \times 6 = 0.1056$

$\frac{r}{PE} = \frac{0.96}{0.0176} = 54.54$ times

∴ The value of r may fall between 0.94 and 0.97. Since, r is greater than six time of PE - r is significant.

34 $r = -0.98$ $N = 9$

$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$

$= 0.6745 \times \frac{1 - (-0.98)^2}{\sqrt{9}}$

$= 0.6745 \times \frac{1 - 0.9604}{3}$

$= 0.6745 \times \frac{0.0396}{3}$

$= 0.6745 \times 0.0132 = 0.0089$

$r \pm PE$

$r + PE$ $\quad$ $r - PE$

$-0.98 + 0.0089$, $-0.98 - 0.0089$

-0.9711 — -0.98 — -0.9889

Interpretation of 'r'

$PE = 0.0089 \times 6 = 0.0534$

$$\frac{r}{PE} = \frac{-0.98}{0.0089} = -110.1123 \text{ times}$$

∴ The value of r may fall between -0.98 and -0.97, Since, r is lesser than 6 times of PE - r is insignificant.

$r = 1$ $\quad$ $N = 7$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (1)^2}{\sqrt{7}}$$

$$= 0.6745 \times \frac{1 - 1}{2.6457}$$

$$= 0.6745 \times 0$$

$PE = 0$

∴ The value of r is +1, perfect positive correlation hence, there is no probable error. Since r = 1, it is highly significant.

5) $r = 0.4042 \quad N = 6$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.40)^2}{\sqrt{6}}$$

$$= 0.6745 \times \frac{1 - 0.16}{2.4494}$$

$$= 0.6745 \times \frac{0.84}{2.4494}$$

$$= 0.6745 \times 0.3429$$

$$= 0.2313$$

$r \pm PE$

r + PE		r − PE
0.40 + 0.2313		0.40 − 0.2313
0.6313 —	0.40	— 0.1687

=> Interpretation of 'r'

$PE = 0.2313 \times 6 = 1.3878$

Kavitha

$$\frac{r}{PE} = \frac{0.40}{0.2313} = 1.7293 \text{ times}$$

∴ The value of r may fall between 0.16 and 0.63. Since, r is lesser than 6 times of PE - r is insignificant.

6) $r = 0.35$ $N = 6$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.35)^2}{\sqrt{6}}$$

$$= 0.6745 \times \frac{1 - 0.1225}{2.4494}$$

$$= 0.6745 \times \frac{0.8775}{2.4494}$$

$$= 0.6745 \times 0.3582$$

$$= 0.2416$$

$r \pm PE$

$r + PE$		$r - PE$
$0.35 + 0.2416$		$0.35 - 0.2416$
0.5916 ———	0.35	——— 0.1084

Interpretation of r^2

$PE = 0.2416 \times 6 = 1.4486$

$\frac{r}{PE} = \frac{0.35}{0.2416} = 1.4486$ times

∴ The value of r may fall between 0.10 and 0.59. Since r is lesser than 6 times of PE - r is insignificant.

Pb $r = 0.20$ $N = 9$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.20)^2}{\sqrt{9}}$$

$$= 0.6745 \times \frac{1 - 0.04}{3}$$

$$= 0.6745 \times \frac{0.96}{3}$$

$$= 0.6745 \times 0.32$$

$$= 0.2158$$

Kavitha

$r \pm PE$

$r + PE$ | $r - PE$

$0.20 + 0.2158$ | $0.20 - 0.2158$

0.4158 —— 0.20 —— -0.0158

Interpretation of 'r'

$PE = 0.2158 \times 6 = 1.2948$

$$\frac{r}{PE} = \frac{0.20}{0.2158} = 0.9267 \text{ times}$$

$\therefore$ The value of r may fall between -0.0158 and 0.4158. Since r is lesser than 6 times of PE – r is insignificant.

$r = -0.9919$ $\quad N = 6$

$$PE = 0.6745 \times \frac{1-r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1-(-0.9919)^2}{\sqrt{6}}$$

$$= 0.6745 \times \frac{1 - 0.9838}{2.4494}$$

$$= 0.6745 \times \frac{0.0162}{2.4494}$$

$$= 0.6745 \times 0.0066 = 0.0044$$

$r \pm PE$

$r + PE$	$r - PE$
$-0.9919 + 0.0044$	$-0.9919 + 0.0044$

$-0.9875 \longleftarrow -0.9919 \longrightarrow -0.9963$

=> Interpretation of 'r'

$PE = 0.0044 \times 6 = 0.0264$

$\frac{r}{PE} = \frac{-0.9919}{0.0044} = -225.4318$ times

∴ The value of 'r' may fall between -0.9963 and -0.9875. Since r is lesser than 6 times PE - r is insignificant.

10) Following table is distribution of density of populations and death rates. Find out if there is any relationship between density of population and death rates.

Districts	kilometers	population	No. of deaths
A	120	24000	288
B	150	75000	1125
C	80	48000	768
D	50	40000	720
E	200	50000	650

=> Density of $Popl^n = \frac{Popl^n}{Kilometers}$

=> Death rate $= \frac{No.\ of\ death}{Popl^n} \times 100$

Kavitha

Soln:-

District	km	Popl^n	No. of death	Density of Popl^n (X)	Death rate (Y)
A	120	24000	288	200	1.2
B	150	75000	1125	500	1.5
C	80	48000	768	600	1.6
D	50	40000	720	800	1.8
E	200	50000	650	250	1.3

=> Assumed mean

A = 200 A = 1.2

x	Y	dx	dx^2	dy	dy^2	dxdy
200	1.2	0	0	0	0	0
500	1.5	300	90000	0.3	0.09	90
600	1.6	400	160000	0.4	0.16	160
800	1.8	600	360000	0.6	0.36	360
250	1.3	50	2500	0.1	0.01	5
2350	7.4	1350	612500	1.4	0.62	615

$$r = \frac{N\Sigma dxdy - \Sigma dx\Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{5(615) - (1350)(1.4)}{\sqrt{5(612500) - (1350)^2}\sqrt{5(0.62) - (1.4)^2}}$$

$$= \frac{3075 - 1890}{\sqrt{3062500 - 1822500}\sqrt{3.1 - 1.96}}$$

$$= \frac{1185}{\sqrt{1240000}\sqrt{1.14}}$$

$$= \frac{1185}{(1113.55)(1.0677)}$$

Kavitha

$= \frac{1185}{1188.97}$

$= 0.9966$

∴ Their exist high degree positive correlation.

⇒ Probable error

$r = 0.9966 \quad N = 5$

$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$

$= 0.6745 \times \frac{1 - (0.9966)^2}{\sqrt{5}}$

$= 0.6745 \times \frac{0.0068}{2.2360}$

$= 0.6745 \times 0.0030$

$= 0.0020$

$r \pm PE$

$r + PE \qquad r - PE$

$0.9966 + 0.0020 \qquad 0.9966 - 0.0020$

0.9986 —— 0.9966 —— 0.9946

Kavitha
Date ______
Page ______

* Interpretation of 'r'

$PE = 0.0020 \times 6 = 0.012$

$\frac{r}{PE} = \frac{0.9966}{0.0020} = 498.3$ times

∴ The value of r may fall between 0.9946 and 0.9986, since r is greater than 6 times PE - r is significant.

11) The following tables gives the distribution of the total population and those who are totally or partially blind among them. Find out if there is any relation between age and blindness.

Age (in years)	population (in '000s)	blind
0-10	100	55
10-20	60	40
20-30	40	40
30-40	36	40
40-50	24	36
50-60	11	22
60-70	6	18
70-80	3	15

NOTE:- whenever the variables are given in class intervals we have to take midpoints as variables

Kavitha

⇒ Actual mean.

Solut

Age	midpoint x	Y	x	x^2	y	y^2
0-10	5	0.055	-35	1225	-0.1303	0.0169
10-20	15	0.0666	-25	625	-0.1187	0.0140
20-30	25	0.1	-15	225	-0.0853	0.0072
30-40	35	0.1111	-5	25	-0.0742	0.0055
40-50	45	0.15	5	25	-0.0353	0.0012
50-60	55	0.2	15	225	0.0147	0.0002
60-70	65	0.3	25	625	0.1147	0.0131
70-80	75	0.5	35	1225	0.3147	0.0990
	320	1.4827		4200		0.1571

$$\bar{X} = \frac{\Sigma x}{N} = \frac{320}{8} = 40$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{1.4827}{8} = 0.1853$$

xy
4.5605
2.9675
1.2795
0.371
-0.1765
0.2205
2.8675
11.0145
$\Sigma xy =$ 23.1045

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

$$= \frac{23.1405}{\sqrt{4200 \times 0.1571}}$$

$$= \frac{23.1405}{\sqrt{659.82}}$$

$$= \frac{23.1405}{25.6869} = 0.8994$$

Kavitha

Date ______
Page ______

=> Probable error

$r = 0.8994 \quad N = 8$

$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$

$= 0.6745 \times \frac{1 - (0.8994)^2}{\sqrt{8}}$

$= 0.6745 \times \frac{0.8994}{2.8284}$

$= 0.6745 \times 0.3179$

$= 0.2144$

$r \pm PE$

$r + PE$ $\qquad\qquad$ $r - PE$

$0.8994 + 0.2144$ $\qquad\qquad$ $0.8994 - 0.2144$

1.1138 ———— 0.8994 ———— 0.685

Interpretation of 'r'

$PE = 0.2144 \times 6 = 1.2864$

$\frac{r}{PE} = \frac{0.8994}{0.2144} = 4.1949$ times

∴ The value of r may fall between 0.685 and 1.1138. Since r is lesser than 6 times PE - r is insignificant.

Kavitha

Kavitha
Date
Page

10) Calculate Karl pearson's correlation co-efficient between advertising expenditure and total profit

Advertising expenditure	profit per unit	units sold.
0 - 10	10	10
10 - 20	12	15
20 - 30	8	20
30 - 40	14	25
40 - 50	10	50
50 - 60	16	40
60 - 70	20	30

soln → Assumed Mean Method

x	y	dx	dx^2	dy	dy^2	dxdy
5	100	0	0	0	0	0
15	180	10	100	80	6400	800
25	160	20	400	60	3600	1200
35	350	30	900	250	62500	7500
45	500	40	1600	400	160000	16000
55	640	50	2500	540	291600	27000
65	600	60	3600	500	250000	30000
245	2530	210	9100	1830	774100	82500

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

Kavitha

$$= \frac{7(82500) - (210)(1830)}{\sqrt{7(9100) - (210)^2}\ \sqrt{7(774100) - (1830)^2}}$$

$$= \frac{577500 - 384300}{\sqrt{63700 - 44100}\ \sqrt{5418700 - 3348900}}$$

$$= \frac{193200}{\sqrt{19800}\ \sqrt{2069800}}$$

$$= \frac{193200}{(140)(1438.6799)}$$

$$= \frac{193200}{201415.186}$$

190307.26

$$= 0.9592$$

⇒ Probable error.

$r = 0.9592 \qquad N = 7$

$$PE = 0.6745 \times \frac{1 - r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1 - (0.9592)^2}{\sqrt{7}}$$

$$= 0.6745 \times \frac{0.08}{2.6457}$$

$$= 0.6745 \times 0.1238 = 0.0835$$

$r \pm PE$

$r+PE$		$r-PE$
$0.9592+0.0835$		$0.9592-0.0835$
1.0427	0.9592	0.8757

Interpretation of 'r'

$PE = 0.0835 \times 6 = 0.501$.

$$\frac{r}{PE} = \frac{0.9592}{0.0835} = 11.4874$$

∴ The value of r may fall between 0.8757 and 1.0427, sin r is greater than 6 times PE - r is significant

(8) Find the correlation co-efficient between age and cinema goers

Age (in years)	locality	population	Cinema goers
10 - 20	Jayanagar	5000	2000
20 - 30	Vijayanagar	8000	40
30 - 40	J.P Nagar	9000	45
40 - 50	Sadashivnagar	2000	800
50 - 60	Shivnagar	10000	2500
60 - 70	R.R Nagar	4000	1000

Date
Page

Soln

x	y	dx	dx^2	dy	dy^2	dxdy
15	40	0	0	15	225	0
25	50	10	100	25	625	250
35	50	20	400	25	625	500
45	40	30	900	15	225	450
55	25	40	1600	0	0	0
65	25	50	2500	0	0	0
240	230	150	5500	80	1700	1200

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{6(1200) - 150 \times 80}{\sqrt{6(5500) - (150)^2}\sqrt{6(1700) - (80)^2}}$$

$$= \frac{7200 - 12000}{\sqrt{33000 - 22500}\sqrt{10200 - 6400}}$$

$$= \frac{-4800}{\sqrt{10500}\sqrt{3800}}$$

$$= \frac{-4800}{102.47 \times 61.64}$$

$$= \frac{-4800}{6316.2508}$$

$$= -0.7599$$

Kavitha

→ Probable error

$r = -0.7599 \quad n = 6$

$$PE = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1-(-0.7599)^2}{\sqrt{6}}$$

$$= 0.6745 \times \frac{1-0.5774}{2.4494}$$

$$= 0.6745 \times \frac{0.4226}{2.4494}$$

$$= 0.6745 \times 0.1725$$

$$= 0.1163$$

$r \pm PE$

$r + PE$ | $r - PE$

$-0.7599 + 0.1163$ | $-0.7599 - 0.1163$

-0.6436 —— 0.1163 —— -0.8762.

Interpretation of 'r'

$PE = 0.1163 \times 6 = 0.6978$

$$\frac{r}{PE} = \frac{-0.7599}{0.1163} = -6.5339 \text{ times.}$$

∴ The value of r may fall between -0.8762 and -0.6436. Since r is lesser than 6 times of PE - r is insignificant.

MISCELLANEOUS PROBLEMS:

1) $cov(x,y) = 10$ $\sigma x = 4$, $\sigma y = 3$. Find out correlation coefficient.

soln

$$r = \frac{cov(x,y)}{\sigma x \, \sigma y}$$

$$= \frac{10}{4 \times 3} = \frac{10}{12} = 0.8333$$

∴ There exists a high degree positive correlation.

2) From the following data. Find the correlation co-efficient between x and y series

Particulars	x	y
N = Total no of observations	15	15
$\bar{X}$ & $\bar{Y}$ Arithmetic mean	25	18
σx & σy standard deviation	3.01	3.03
Σx^2 & Σy^2 Sum of squares of deviation from mean	135	138
Sum of product of deviation	$\Sigma xy = 122$	

sol^n

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

$$= \frac{122}{\sqrt{135 \times 138}}$$

$$= \frac{122}{\sqrt{18630}}$$

$$= \frac{122}{136.4}$$

$$= 0.894$$

OR

$$r = \frac{\Sigma xy}{N \sigma x \sigma y}$$

$$= \frac{122}{15 \times 3.01 \times 3.03}$$

$$= \frac{122}{136.8}$$

$$= 0.891$$

$\therefore$ their exist a high degree positive correlation.

3) From the following data find the correlation co-efficient between x & y.

Particulars	x	y
Total no of observation (N)	10	10
Arithmetic mean ($\bar{x}$, $\bar{y}$)	14.6	12.7
Sum of squares of deviation from mean (Σx^2, Σy^2)	115.96	59.04
Sum of product of observation	$\Sigma xy = 53.95$	

Soln

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

$$= \frac{53.95}{\sqrt{115.96 \times 59.04}}$$

$$= \frac{53.95}{\sqrt{6846.27}}$$

$$= \frac{53.95}{82.7421}$$

$$= 0.6520$$

∴ Their exist a moderate degree positive correlation.

53. From the following data find the correlation co-efficient between x & y.

Particulars	x	y
Total no of observation (N)	15	15
standard deviation (σx, σy)	10	12
Sum of product of deviation	$\Sigma xy = 122$	

Soln

$$r = \frac{\Sigma xy}{N \sigma_x \sigma_y}$$

$$= \frac{122}{15(10)(12)}$$

$$= \frac{122}{1800}$$

$$= 0.0677$$

∴ Their exist a low degree positive correlation.

54. From the following data. Find the correlation co-efficient between x & y series.

Particulars	x	y
Total no of observations (N)	7	7
standard deviations (σx, σy)	2.76	2.05
Sum of product of deviations	$\Sigma xy = 38$	

solu

$$r = \frac{\Sigma xy}{N \sigma_x \sigma_y}$$

$$= \frac{38}{7(2.76)(2.05)}$$

$$= \frac{38}{39.606}$$

$$= 0.9594 / 0.96$$

∴ Their exist a high degree positive correlation.

⇒ ## WHEN ACTUAL AND ASSUMED MEAN ARE GIVEN :-

$$r = \frac{\Sigma xy - N(\bar{x} - A_x)(\bar{y} - A_y)}{N \sigma_x \sigma_y}$$

6) From the following data, find the correlation co-efficient between x & y series.

Particulars	x	y
Total no of observation (N)	9	9
Standard deviation (σ_x σ_y)	9.07	11.85
Arithmetic mean ($\bar{x}$, $\bar{y}$)	70.5	121.5
Assumed mean (A_x, A_y)	65	108
Sum of product of deviation	$\Sigma xy = 1451$	

solut

$$r = \frac{\Sigma xy - N(\bar{x} - A_x)(\bar{y} - A_y)}{N \sigma_x \sigma_y}$$

$$= \frac{1451 - 9(70.5 - 65)(121.5 - 108)}{9(9.07)(11.85)}$$

$$= \frac{1451 - 9(5.5)(13.5)}{967.3155}$$

$$= \frac{1451 - 668.25}{967.3155}$$

$$= \frac{782.75}{967.3155}$$

$= 0.8091 / 0.8092$

∴ Their exists a high degree positive correlation.

74 From the following data. find correlation co-efficient between x & y series.

Particulars	x	y
Total no of observation (N)	10	10
Standard deviation (σ_x, σ_y)	13.17	15.75
Arithmetic mean ($\bar{x}, \bar{y}$)	75.5	126.5
Assumed mean (A_x, A_y)	69	113
Sum of product of deviation	$\Sigma xy = 2176$	

solu

$$r = \frac{\Sigma xy - N(\bar{x} - A_x)(\bar{y} - A_y)}{N \sigma_x \sigma_y}$$

$$= \frac{2176 - 10(75.5 - 69)(126.5 - 113)}{10(13.17)(15.75)}$$

$$= \frac{2176 - 10(6.5)(13.5)}{2074.275}$$

$$= \frac{2176 - 877.5}{2074.275}$$

$$= \frac{1298.5}{2074.275}$$

$$= 0.6260$$

∴ Their exist a moderate degree positive correlation.

8) From the following data, Find correlation co-efficient between x & y series.

Particulars	x	y
Total no of observations (n)	8	8
Standard deviation (σ_x, σ_y)	13.07	16
Arithmetic mean ($\bar{x}, \bar{y}$)	16	20.6
Assumed mean (A_x, A_y)	18	20
Sum of product of deviation	$\Sigma xy = 220$	

Soln

$$r = \frac{\Sigma xy - N(\bar{x} - A_x)(\bar{y} - A_y)}{N \sigma_x \sigma_y}$$

$$= \frac{220 - 8(16 - 18)(20.6 - 20)}{8(13.07)(16)}$$

$$= \frac{220 - 8(-2)(0.6)}{8(13.07)(16)}$$

$$= \frac{220 + 9.6}{1672.96}$$

$$= \frac{229.6}{1672.96}$$

$$= 0.1372$$

∴ Their exist a low degree positive correlation.

=t 94 SDP Calculate the correlation coefficient for the age of husband wives.

Age of husband (in years)	Age of wife (in years)
23	18
27	22
28	23
29	24
30	25

Date
Page

31	26
33	28
35	29
36	30
39	32

solut =) Actual mean method.

X	Y	x	x^2	y	y^2	xy
23	18	-8.1	65.41	-7.7	59.29	62.37
27	22	-4.1	16.81	-3.7	13.69	15.17
28	23	-3.1	9.61	-2.7	7.29	8.37
29	24	-2.1	4.41	-1.7	2.89	3.57
30	25	-1.1	1.21	-0.7	0.49	0.77
31	26	-0.1	0.01	0.3	0.09	-0.03
33	28	1.9	3.61	2.3	5.29	4.37
35	29	3.9	15.21	3.3	10.89	12.87
36	30	4.9	24.01	4.3	18.49	21.07
39	32	7.9	62.41	6.3	39.69	49.77
311	257		202.9		158.1	178.3

$\bar{X} = \frac{\Sigma xy}{N} = \frac{311}{10} = 31.1$

$\bar{Y} = \frac{\Sigma xy}{N} = \frac{257}{10} = 25.7$

$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \frac{178.3}{\sqrt{(202.9)(158.1)}}$

Kavitha

$$= \frac{178.3}{\sqrt{32078.49}}$$

$$= \frac{178.3}{179.1046}$$

$$= 0.9955$$

∴ Their exist a high degree positive correlation.

→ Assumed mean method.

x	y	dx	dx^2	dy	dy^2	dxdy
23	18	0	0	0	0	0
27	22	4	16	4	16	16
28	23	5	25	5	25	25
29	24	6	36	6	36	36
30	25	7	49	7	49	49
31	26	8	64	8	64	64
33	28	10	100	10	100	100
35	29	12	144	11	121	132
36	30	13	169	12	144	156
39	32	16	256	14	196	224
311	257	81	859	77	751	802.

$$r = \frac{N\Sigma dxdy - \Sigma dx \Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{10(802) - (81)(77)}{\sqrt{10(859) - (81)^2}\sqrt{10(751) - (77)^2}}$$

$$= \frac{8020 - 6237}{\sqrt{8590 - 6561}\sqrt{7510 - 5929}}$$

$$= \frac{1783}{\sqrt{2069}\sqrt{1581}}$$

$$= \frac{1783}{(45.4863)(39.7617)}$$

$$= \frac{1783}{1808.6126}$$

$$= 0.9858 / 0.99$$

∴ Their exist a high degree positive correlation.