Statistics for Economics

18K4ECAS2

Unit - I

Association of Attributes (two attributes), Nine square table, type of association, methods of studying association - Yule's co-efficient of association - Definition and problems.

Unit - II

Simple correlation - Definition and types of correlation - methods of studying correlation - scatter diagram, Karl Pearson's coefficient of correlation, Spearman's rank correlation coefficient and simple linear Regression analysis (problems).

Unit - III

Time series - concept and definition, components of Time series - Secular trend, seasonal variation, cyclical variation and Irregular variation, measurement of Trend only by the method of moving average and method of least squares.

Unit - IV

Testing of hypothesis - Definition of hypothesis - null hypothesis and alternative hypothesis, standard error, level of significance, critical region, parameters and statistic. Type I and Type II errors, one tailed and two tailed tests, Test procedure.

Unit - V

Large sample tests - Test for single mean and difference between two means. Test for single proportion and difference between proportions - procedure and problems small sample tests - chi-square test for independence of attributes (Two attributes only), 't' test for single mean, difference between two means and paired 't' test (procedure and problems).

Books of Study:-

1. Statistics Theory and Practice - R.S.N. Pillai & V. Bagavathi (VII Edition) (Reprint - 2013).
2. Comprehensive Statistical methods - P.N. Arora, Sumeet Arora, S. Arora (IV Edition) (Reprint - 2013).

Unit - I
Association of Attributes

meaning:

The characteristics which are not capable of quantitative measurement are termed as statistics of attributes. The relationship between such attributes is known as association of attributes. Eg: literacy, employment, blindness, deafness, beauty, colour etc...

Positive and negative classes: The attributes may be positive or negative. If the attribute is present, it is termed as positive class and it is denoted by capital letters A, B, C etc. If the attribute is not present, it is termed as negative class and it is denoted by small Greek letters α, β, γ.

Number of classes: The total number of classes comprising of the various attributes can be determined by 3^n, n representing the number of attributes. If one attribute is studied, then there will be $3=3$ classes. If two attributes are studied then there will be $3^2=9$ classes. They are (A), (α), (B), (β), (AB), (Aβ), (αB), ($\alpha\beta$), (N).

Class frequencies: The number of values which posses a particular attribute is called the class frequency. They are denoted by (A), (α), (B), (β), (AB), (Aβ), (αB), ($\alpha\beta$), N.

Ultimate class frequency: Those classes which specify the attributes of the highest order are known as the ultimate classes and their frequencies are known as ultimate class frequencies. In the study of two attributes A and B. The B have 4 ultimate frequencies (AB), (αB), (Aβ), ($\alpha\beta$). 2^n. The classes which represent the presence of an attribute or attributes are called positive N, A, B, AB classes. If the attribute is not present it is called negative α, β, $\alpha\beta$ classes. If one of the attribute is present and the other absent are called pairs of αB, Aβ of contraries.

Order of classes: The order of class depends upon the number of attributes under study. A class having one attribute is known as the class of first order; a class having two attributes as the class of the second order. N denotes the number of members.

N - zero order

A, B, α, β - first order

AB, Aβ, αB, $\alpha\beta$ - second order

Contigency table or Nine squaretable: The class frequencies in a study of two attributes A and B are represented by the following table which is called.

Contigency table.

	A	α	Total
B	(AB)	(αB)	(B)
β	(Aβ)	($\alpha\beta$)	(β)
Total	(A)	(α)	N

Since it has got nine squares which represent the class frequencies it is also called a nine square table.

Relationship:

Vertical totals	Horizontal totals
$N = (B) + (\beta)$	$N = (A) + (\alpha)$
$A = (AB) + (A\beta)$	$B = (AB) + (\alpha B)$
$\alpha = (\alpha B) + (\alpha\beta)$	$\beta = (A\beta) + (\alpha\beta)$

for two attributes A and B, we have (AB) = 35; (A) = 55; N = 100 and (B) = 65. Calculate the missing values, with the help of Nine square Table.

Solution:

	(A)	(α)	Total
(B)	35 (AB)	(αB)	65 (B)
(β)	(Aβ)	($\alpha\beta$)	(β)
Total	55 (A)	(α)	100 N

	(A)	(α)	
(B)	?	30 (65 - 35) (αB)	?
(β)	20 (55 - 35) ($A\beta$)	15 (35 - 20) ($\alpha\beta$)	35 (100 - 65) (β)
	?	45 (100 - 55) (α)	?

$(\alpha) = N - (A) = 100 - 55 = 45$

$(A\beta) = (A) - (AB) = 55 - 35 = 20$

$(\alpha B) = (B) - (AB) = 65 - 35 = 30$

$(\beta) = N - (B) = 100 - 65 = 35$

$(\alpha\beta) = (\beta) - (A\beta) = 35 - 20 = 15$

consistency of data: statistics of attributes is obtained by counting & and as such no class frequency can be negative. The class frequency, can be positive or zero, but cannot be negative. Data observed may be described as consistent. if they do not conflict with one another. In case any class frequency is negative then the given data are inconsistent. It is simple test to be applied and verified whether the frequency or frequencies of classes are negative or not. If no conflict is there, no frequencies are negative. it is concluded that the given data are consistent.

Types of Association; there are three types of association:

(i) positive association

(ii) Negative association

(iii) Independent association.

positive association: when two attributes are present or absent together in the data. and actual frequency is more than the expected frequency it is called positive association Eg: smoking and cancer, literacy and employment, i.e., $(AB) > \frac{(A) \times (B)}{N}$

(Actual) > (Expected)

Negative Association: (Diassociation)

when the existence of one attribute causes absence of another attribute and actual frequency is less than the expected frequency, it is called Negative association or disassociation). Eg: cleanliness and ill-health. ie., $(AB) < \frac{(A) \times (B)}{N}$

(Actual) < (Expected)

Independent association:

when there exists no association between two attributes or when they have no tendency to be present together or the presence of one attribute does not affect the other attribute the two attributes are said to be independent. Actual frequency is equal to the expected frequency ie $(AB) = \frac{(A) \times (B)}{N}$ (Actual) = (Expected)

method of determining association: Association can be studied by any one of the following methods.

1. Comparison of observed and expected frequencies.
2. Comparison of proportions.
3. Yule's coefficient of association.

1. Comparison of observed and expected frequencies: In this method the actual number of observation is compared with the expected frequencies. The probability is the expectation of $(AB) = \frac{(A) \times (B)}{N}$ and $(\alpha B) = \frac{(\alpha) \times (B)}{N}$ The expected frequency can be found by combination also. this will be clear from the following.

Attribute	Independent	Positive association	Negative association.
(A) and (B)	$(AB) = \frac{(A)(B)}{N}$	$(AB) > \frac{(A)(B)}{N}$	$(AB) < \frac{(A)(B)}{N}$
(A) and (B)	$(AB) = \frac{(A)(B)}{N}$	$(AB) > \frac{(A)(B)}{N}$	$(AB) < \frac{(A)(B)}{N}$
(α) and (B)	$(\alpha B) = \frac{(\alpha)(B)}{N}$	$(\alpha B) > \frac{(\alpha)(B)}{N}$	$(\alpha B) < \frac{(\alpha)(B)}{N}$
(α) and (B)	$(\alpha B) = \frac{(\alpha)(B)}{N}$	$(\alpha B) > \frac{(\alpha)(B)}{N}$	$(\alpha B) < \frac{(\alpha)(B)}{N}$

can vaccination be regards as a preventive measure for small pox from data given below?

(i) Of 2,000 persons in a locality exposed to small-pox, 450 in all were attacked.

(ii) 2000 persons, 365 had been vaccinated; of those only 50 were attracted.

Solution:-

Let (A): vaccinated, (α) Not vaccinated

(B): Exempted from small pox (β) Attack of small pox.

The missing values can be obtained from the following nine square table.

315 (AB)	1235 (αB)	1550 (B)
50 (Aβ)	4000 ($\alpha\beta$)	450 (β)
365 (A)	1635 (α)	2000 N

percentage of vaccinated, who were not attacked

$$= \frac{(AB)}{A} = \frac{315 \times 100}{365} = 86.3\%.$$

percentage of not vaccinated, but not attracted

$$= \frac{(\alpha B)}{(\alpha)} \cdot \frac{1235}{1635} \times 100 = 73.2\%.$$

thus

$\frac{(AB)}{A} > \frac{(\alpha B)}{(\alpha)}$ and therefore positively associated. Hence, vaccination is a good preventive measure for small pox.

Limitation: this method we can only determine the nature of association and not the degree of association.

2. Comparison of proportions: Under this method, ratios or proportions of the concerned variables are compared. Relationship is given below:

Association	proportion of B in A and α	proportion of A in B and β
Independent	$\frac{(AB)}{A} = \frac{(\alpha B)}{(\alpha)}$	$\frac{(AB)}{A} = \frac{(AB)}{B}$
Positive	$\frac{(AB)}{A} > \frac{(\alpha B)}{(\alpha)}$	$\frac{(AB)}{B} > \frac{(AB)}{B}$
Negative	$\frac{(AB)}{A} < \frac{(\alpha B)}{(\alpha)}$	$\frac{(AB)}{B} < \frac{(AB)}{B}$

Out of 700 literates in a particular district, the number of criminals was 5. Out of 93000 illiterates in the same district, the number of criminals was 150. On the basis of these figures do you find any association between illiteracy and criminality?

Solution:- Let A denote the attribute of illiteracy and B of criminality α will denote literacy and B non - criminality.

50 (AB)	5 (αB)	155 (B)
9150 (AB)	695 (αB)	9815 (B)
9300 (A)	700 (α)	10,000 N

proportion of illiterate criminals to illiterate $= \frac{(AB)}{(A)} = \frac{150}{9300} = 0.01607 \; 1.6\%$.

proportion of illiterate criminals to literates $= \frac{(\alpha B)}{(\alpha)} = \frac{5}{700} = 0.00710 \; 0.71\%$.

Hence, the proportion of criminals is more in illitrates than in the literates therefore, criminality and illieracy are positively association.

3. Yule's coefficient of association: The above mentioned methods will give us a rough idea about their associations but the degree of association cannot be found out. Prof. Yule has suggested a formula to mesure the association. (Q) falls between ± 1. If Q = 0, no association; if Q = +1. there is perfect positive association; and if Q = -1, there is perfect negative association.

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

In a co-educational institution, out of 200 students, 150 were boys, they took an examination and it was found that 120 passed. 10 girls failed. Is there any association between sex and success in the examination.

Solution :-

Let A denot boys and α denote girls.

Let B denote those who passed the examination and β denote those who failed.

We have given : $N = 200$, $(A) = 150$, $(AB) = 120$; $(\alpha\beta) = 10$.

Other frequencies can be obtained from nine-square table

	A	α	Total
B	120	40	160
β	30	10	40
Total	150	50	200

Applying yule's method :

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$Q = \frac{(120 \times 10) - (30 \times 40)}{(120 \times 10) + (30 \times 40)} = \frac{1200 - 1200}{1200 + 1200} = 0.$$

Therefore, there is no association between sex and success in the examination.

unit - II
CORRELATION.

DEFINITION

According to croxton and cowden," the relationship of quantiative nature, the appropride statistical tool for discovering and measuring the relationship and expressing it in brief formula is known as correlation.

According to A.M. Tuttle,' correlation is an analysis of the coveriation between two or more variables.

TYPES OF CORRELATION:

correlation is classified into many types, but the important are:

1. positive and negative.
2. simple and multiple
3. partial and total
4. linear and non-linear.

1. POSITIVE AND NEGATIVE CORRELATION: positive and negative correlation depend upon the direction of change of the variables. If two variables tend to move together in the same direction. ie, an increase in the value of one variables is accompanied by an increase in the value of the other variables; or a decrease in the value of the variables is accompanied by a decrease in the value of the other variable; or a decrease in the value of one variable is accompanied by a decrease in the value of the other variable, then the correlation is called positive or direct correlation. Height and weight, rainfall and yield of crops, price and supply are example of positive correlation.

If two variables, tend to move together in opposite directions so that an increase or decrease in the values of one variables is accom-pained by a decrease or increase in the value of the other variable, then

The correlation is called negative or inverse correlation. price and demand. Yield of crops and price, etc., are examples of negative correlation. Here, the increase in the values of the independent variables is associated with the decrease in the value of the dependent variable or vice versa.

2. SIMPLE AND MULTIPLE: when we study only two variables, the relationship is described as simple correlation; example, quantity of money and price level, demand and price, etc., But in a multiple correlation we study more than two variables simultaneously; example, the relationship of price, demand and supply of a commodity.

3. PARTIAL AND TOTAL: The study of two variables, the excluding some other variables is called partial correlation. for example, we study price and demand, eliminating the supply side. In Total correlation, all the facts are taken into account.

4. LINEAR AND NON-LINEAR: If the ratio of change between two variables is uniform, Then there will be linear correlation between them. consider the following

X	5	10	15	20
Y	4	8	12	16

The ratio of change between the variables is the same. If we plot these on the graph, we get a straight line.

In a curvilinear or non-linear correlation, The amount of change in one variable does not bear a constant ratio of the amount of change in the other variables. The graph of non-linear or curvilinear relationship will form a curve.

In majority of cases, we find curvilinear relationship, which is a complicated one, so we generally assume that the relationship between the variables under study is linear. In social sciences, linear correlation is rare, because the exactness is not so perfect as in natural sciences.

Methods of studying correlation: The different methods of finding out the relationship between two variables are:

Graphic method

1. Scatter diagram or scattergram

Mathematical method

2. Karl pearson's coefficient of correlation

3. Spearman's Rank coefficient of correlation

1. SCATTER DIAGRAM METHOD: This is the simplest method of finding out whether there is any relationship present between two variables by plotting the values on a chart, known as scatter diagram. In this method, the given data are plotted on a graph paper in the form of dots. X variables are plotted on the horizontal axis and Y variables on the vertical axes. Thus we have the dots and we can known the scatter or concentration of the various points. This will show the type of correlation.

perfect positive correlation

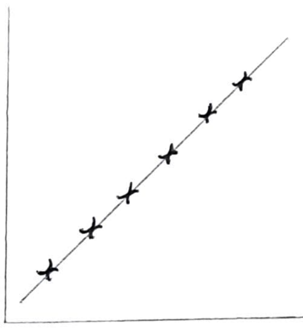

Diagram 1 ($r = +1$)

perfect Negative correlation

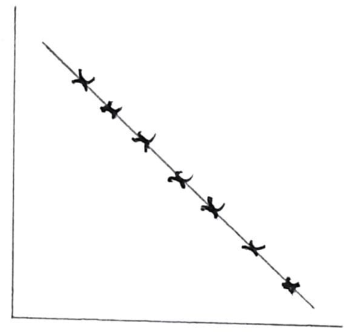

Diagram 2 ($r = -1$)

If the plotted points form a straight line runing from the lower left-hand corner to the upper right-hand corner, then there is a perfect positive correlation (i.e. $r = +1$, Diagram 1). On the other hand, if the points are in a straight line, having a falling trend from the upper left-hand corner to the lower right-hand corner, it reveals that there is a perfect negative or inverse correlation (i.e, $r = -1$, Diagram 2).

If the plotted points fall in a narrow band, and the points are rising from lower left-hand corner to the upper right-hand corner.

there will be a hight degree or positive correlation between the variables (Diagram 3). If the plotted points fall in a narrow band from the upper left-hand corner to the lower right-hand corner, there will be a hight degree or negative correlation (Diagram 4). If the plotted points lie scatter all over the diagram, there is no correlation between the two variables (Diagram 5).

High degree of positive correlation

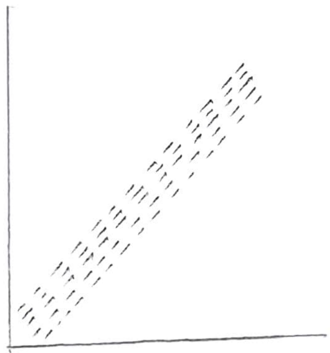

Diagram 3

High degree of negative correlation

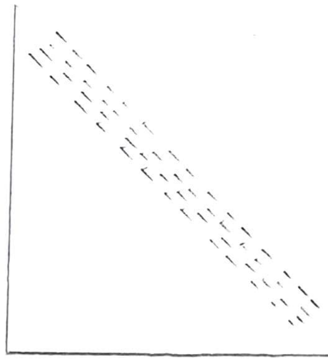

Diagram 4

No correlation ($r=0$).

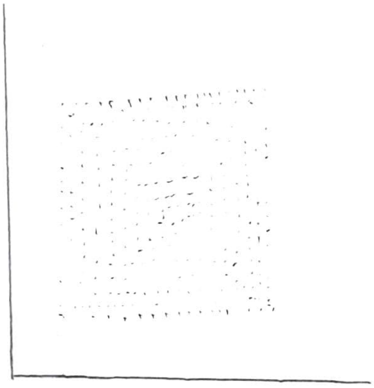

Diagram 5

Merits

1. Scatter diagram is a simple and attractive method of finding out the nature of correlation between two variables.

2. It is a non-mathematical method of studying correlation. It is easy to understand.

3. We can get a rough idea at a glance whether it is a positive or negative correlation.

4. It is not influencedy by extreme items.

5. It is a first step in finding out the relationship between two variables.

Demerits.

By this method we cannot get the exact degree or correlation between two variables. It gives only a rough idea.

Karl pearson's coefficient of correlation: Karl person, a great biometrician and statistician, suggested a mathematical method for measuring the magnitude of linear relationship between two variables. Karl person's method is the most widely used method in practice and is known as pearsonian coefficient of correlation. It is denoted by the symbol 'r'; the formula for calculating pearsonian r is:

(1) $r = \dfrac{\text{covariance of } xy}{\sigma x \times \sigma y}$, (2) $r = \dfrac{\Sigma xy}{N \sigma x \times \sigma y}$, (3) $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$x = (X - \bar{X})$ $y = (Y - \bar{Y})$

σx = standard deviation of series x

σy = standard deviation of series y

when the deviation of items are taken from the actual mean, we can apply any one of these methods but the simplest formula is the third one.

find karl pearson's coefficient of correlation from the following data.

wages: 100 101 102 102 100 99 97 98 96 95

cost of living: 98 99 99 97 95 92 95 94 90 91

Solution:-

computation of karl pearson's coefficient of correlation.

wages (X)	$d = X - \bar{X}$ $X - 99$	$d^2 (x)$	cost of living (Y)	$d = Y - \bar{Y}$ $Y - 95$	$d^2 (y)$	$d(x)\,d(y)$
100	+1	1	98	+3	9	+3
101	+2	4	99	+4	16	+8
102	+3	9	99	+4	16	+12
102	+3	9	97	+2	4	+6
100	+1	1	95	0	0	0
99	0	0	92	-3	9	0
97	-2	4	95	0	0	0
98	-1	1	94	-1	1	+1
96	-3	9	90	-5	25	+15
95	-4	16	91	-4	16	+16
$\Sigma X = 990$	$\Sigma x = 0$	$\Sigma d^2 = 54$	$\Sigma Y = 950$	$\Sigma d = 0$	$\Sigma d^2 = 96$	$\Sigma dx dy = 61$

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

$\Sigma xy = 61$; $\Sigma x^2 = 54$; $\Sigma y^2 = 96$

$$r = \frac{61}{\sqrt{54 \times 96}} = \frac{61}{\sqrt{5184}} = \frac{61}{72}$$

$$= 0.847$$

$$r = +0.847.$$

Rank correlation coefficient : In 1904, charles Edward spearman, a British psychologist found out the method of ascertaining the coefficient of correlation by ranks. this method is based on rank. this method is based useful in dealing with quatitive characteristics, such as intelligence, beauty, morality, character, etc. It cannot be measured quantitatively. as in the case of pearsons beauty coefficient of correlation; but it is based on the ranks given to the observations, It can be used when the data are irregular or extreme itemes are erratic or inaccurate, because rank correlation co-efficient is not based on the assumption of formality of data.

Rank correlation is applicable only to individual observations. The result we get from this method is only an approximate one, because under ranking method origional value are not taken into account. the formula for spearman's rank correlation which is denoted by p is;

$$p = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$p = 1 - \frac{6 \Sigma D^2}{N^3 - N}$$

p = Rank coefficient of correlation.

D^2: Sum of the squares of the differences of two ranks

N = Number of paired observations.

like the karl pearson's co-efficient of correlation, The value of p lies between

+1 and −1, If $\rho = +1$, then there is complete agreement in the order of ranks and the direction of the rank is also the same. when $\rho = -1$, then there is complete disagreement in the order of ranks and they are in opposite directions. we can find this in the following examples.

we may come across two types of problems:

a) where ranks are given
b) where ranks are not given

a) where ranks are given: when the actual ranks are given, the steps followed are;

1. compute the difference of the two ranks (R_1 and R_2) and denote by D.
2. square the D and get ΣD^2
3. substitute the figures in the formula.

following are the rank obtained by 10 students in two subjects statistics and mathematics. To what extent the knowledge of the students in the two subject is related?

statistics:	1	2	3	4	5	6	7	8	9	10
mathematics:	2	4	1	5	3	9	7	10	6	8

Solution:

Rank of statistics (X)	Rank of mathematics (Y)	$D = (X - Y)$	D^2
1	2	−1	1
2	4	−2	4
3	1	+2	4
4	5	−1	1
5	3	+2	4
6	9	−3	9
7	7	0	0
8	10	−2	4
9	6	+3	9
10	8	+2	4
			$\Sigma D^2 = 40$

$$P = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 40}{10(10^2-1)}$$

$$= 1 - \frac{240}{10(100-1)}$$

$$= 1 - \frac{240}{990} = 1 - 0.24$$

$$= +0.76.$$

where ranks are not given: when no rank is given, but actual data are given, then we most given ranks. we can give ranks by taking the highest as 1 or the lowest value as 1, next to the highest as 2 and follow the same procedure for both the variables.

A random sample of 5 college students is selected and their grades is mathematics and statistics are found to be:

	1	2	3	4	5
mathematics :	85	60	73	40	90
statistics :	93	75	65	50	80

calculate spearman's rank correlation co-efficient

solution:-

mark in mathematics x	Ranks x	marks in statistics Y	Rank Y	Rank difference $x-y$	D^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					$\Sigma D^2 = 4$

Spearman's Rank correlation $= 1 - \frac{6\Sigma D^2}{N(N^2-1)}$

$N = 5$, $D^2 = 4$

$$= 1 - \frac{6 \times 4}{5(5^2-1)}$$

$$= 1 - \frac{24}{5(25-1)} = 1 - \frac{24}{5 \times 24}$$

$$= 1 - \frac{1}{5} = 1 - 0.2$$

$$= +0.8.$$

Rank correlation is $+0.8$.

Equal or repeated ranks: when two or more items have equal values, it is difficult to give ranks to them. In that case the items are given the averages of the ranks they would have received, if they are not tied. For example, if two individuals are placed in the seventh place, they are each given the rank $\frac{7+8}{2} = 7.5$ which is common rank to be assigned; and the next will be 9; and if three are ranked equal at the seventh place, they are given the rank $\frac{7+8+9}{3} = 8$ which is the common rank to be assigned to each; and the next rank will be 10, in this case. A slightly different formula is used when there is more than one item having the same value. The formula is:

$$P = 1 - 6 \left\{ \frac{\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) \cdots}{N^3 - N} \right\}$$

m = The number of items whose ranks are common

From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

X	48	33	40	9	16	16	65	24	16	57

Y	13	13	24	6	15	4	20	9	6	19

Solution:-

first we have to assign ranks to the variables.

X	Rank (X)	Y	Rank (Y)	D(R(X) - R(Y))	D2
48	8	13	5.5	2.5	6.25
33	6	13	5.5	0.5	0.25
40	7	24	10	-3	9.00
9	1	6	2.5	-1.5	2.25
16	3	15	7	4	16.00
16	3	4	1	2	4.00
65	10	20	9	1	1.00
24	5	9	4	1	1.00
16	3	6	2.5	0.5	0.25
57	9	19	8	1	1.00
					$\Sigma D^2 = 41$

16 is repeated 3 times in x items hence m= 3. 13 and 6 are repeated twice in y items; hence m=2. Therefore the formula is;

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N^3 - N}$$

$$= 1 - \frac{6\left[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{10^3 - 10}$$

$$= 1 - \frac{6(41 + 2 + 0.5 + 0.5)}{990}$$

$$= 1 - \frac{264}{990} = 1 - 0.267 = +0.733.$$

merits of Rank correlation coefficient

1. It is simple to understand and easy to calculate.
2. It is useful in the case of data which are of qualitative nature, like intelligence, honesty, beauty, efficiency, etc.

3. No other method can be used when the ranks are given, except this.

4. when the actual data are given, this method can also be applied.

Demerits of Rank correlation coefficient.

1. It cannot be used in the case of bi-variate distribution.

2. If the number of items are greater than, say 30, the calculation becomes tedious and requires a lot of time. If we are given the ranks, then we can apply this method even though N exceeds 30.

Blair: " Regression is the measure of the average relationship between two or more variable in terms of the original units of the data.

Find two regression equation and estimate the value of y correspondin to x = 6.2

X	1	2	3	4	5	6	7	8	9	45
Y	9	8	10	12	11	13	14	16	15	108.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{45}{9} = 5$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{108}{9} = 12$$

$x = X - \bar{X}$	$y = Y - 9$	x^2	y^2	xy
-4	-3	16	9	12
-3	-4	9	16	12
-2	-2	4	4	4
-1	0	1	0	0
0	-1	0	1	0
1	1	1	1	1
2	2	4	4	4
3	4	9	16	12
4	3	16	9	12
		60	60	57

Regression equation of x on y :

$$(x-\bar{x}) = bxy\,(y-\bar{y})$$

$$\text{where } bxy \Rightarrow \frac{6x}{6y} = \frac{\Sigma xy}{\Sigma y^2}$$

$$= \frac{57}{60}$$

$$(x-5) = 0.95\,(y-12)$$

$$x-5 = 0.95y - 11.4$$

$$x = 0.95y - 11.4+5$$

$$x = 0.95y - 6.4$$

Regression equation of y on x :

$$(y-\bar{y}) = byx\,(x-\bar{x})$$

$$\text{where } byx \Rightarrow \frac{6y}{6x} = \frac{\Sigma xy}{\Sigma x^2}$$

$$= \frac{57}{60} = 0.95$$

$$(y-12) = 0.95\,(x-5)$$

$$y-12 = 0.95x - 4.75$$

$$y = 0.95x - 4.75 + 12$$

$$y = 0.95x + 7.25$$

when $x = 6.2$

$$y = 0.95 \times 6.2 + 7.25$$

$$= 5.89 + 7.25$$

$$y = 13.14.$$

Unit - III

Time series and forecasting.

Definition:-

A time series consists of data arranged chronological.

- Croxton and Cowden

A set of data depending on the time is called time series.

- Kenny.

Utility of a Time series:-

The following points indicate the utility of time series Analysis:-

1. Analysis:-

It helps in the analysis of past behaviour of a variable. Analysis of past data discloses the effect of various factors on the variables under study. with the help of such analysis the future behaviour of a variable under study can be predicted.

2. forecasting:-

It helps in forecasting. the analysis of past conditions in the basis of forecasting the future behaviour of the variable understudy, this helps in making future plans of actions. various five - year plans of our country are based on the analysis of past data.

3. education:-

It helps in the evaluations of current achievements. The review and evaluation of progress made on the basis of time series data. The progress of plans is judged by the yearly rates of growth in GNP (Gross National product).

4. Comparison:-

It helps in makings comparative studies, once the data is arranged chronologically the comparison between one time period and another is facilitated. It provides a scientific basis for making comparison by studying and isolating the effects of various components of a time series.

5. Approximation:-

It gives approximation indicators.

Components of a time series:-

There are four types of components of a time series, which are as follows.

1. Secular trend or Trend (T)
2. Seasonal variations (S)
3. Cyclic variations (C)
4. Irregular variation (I)

1. Secular Trend or Trend (T): Secular trend in the general tendency of the time series data to increase or decrease or stagnate during a long period of time series. An upward tendency is usually observed in time series relating to population, production and sales, price, income money in circulation while a downward tendency is noticed in data of death's and epidemics as a result of achievement in medical sciences, illiteracy, etc. Thus trend is either upward to downward. It should be clearly understood that trend is general, smooth, long-term average tendency. It is not necessary that the increase or decrease should be in the same direction throughout the given period. With the help of a long-term trend, it is possible to determine and present the direction of change.

2. Seasonal variations: (S) Seasonal variations refer to such movements in a time series which are due to forces which are rhythmic in nature and which repeat themselves periodically in every season. These variations repeat themselve in less than one year time. Seasonal variation are usually measured in an-interval. The seasonal variations may be attributes to those resulting from

natural forces and social customs and traditions. seasonal variations are the results of such factors which uniformly and regularly rise and fall in the magnitude.

3. Cyclic variations: (C) cyclic variations are the oscillatory movements in a time series with period of oscillation greater than one year. These variations in a time series are due to ups and downs recurring after a period greater than one year, these are not necessarily uniformly periodic, i.e., they may or may not follow exactly similarly patterns after equal intervals of time. The cyclic period normally lasts from 7 to 9 years.

4. Irregular variations :- (I) irregular variations do not exhibit by definition pattern and there is no regular regular period or time of their occurrence. These are accidental changes which are purely random unforeseen and unpredictable, They turn the series first in one way and then in the otherway purely by chance with no regularity of occurrance. eg: earthquakes, wars, floods etc., Normally these are short term variations but sometimes their effect is so intense that they may give rise to new cyclical or other movements.

Measurement of Trend :- following are the methods which are used to measure the trend.

1) freehand or geographical method
2) method of semi - Averages.
3) method of moving averages
4) method of least squares.

methods of moving Average :- moving average method is a simple device of reducing fluctuations and obtaining trend values with a fair degree of accuracy. This period can be 3- yearly moving averages, 5 - yearly moving averages, 7-yearly averages, etc.,

for Example :-

3- yearly moving averages can be calculated from the data given by.

$\frac{a+b+c}{3}$, $\frac{b+c+d}{3}$, $\frac{c+d+e}{3}$, $\frac{d+e+f}{3}$ + ...

and these total to be written besides the year b, c, d, e and so on.

Merits:- (i) This method is simple.

(ii) This method is flexible enough to add more figures to the data because the entire calculations are not changed.

(iii) This method is objective in the sense that any body working on a problem with this method will get the same results.

(iv) If the period of moving averages coincides with the period of cyclic fluctuations in the data such fluctuations are automatically eliminated.

(v) This method is used for determining seasonal, cyclic and irregular variations besides the trend values.

Limitations:- (i) There are no trend value for some years in the begining and some in the end. for example, for 7-yearly moving averages there will be no trend values for the first three years and the last three years.

(ii) There is no functional relationship between the values and the time. Hence, this method is not helpful in forecasting and predicting the value on the basis of time.

(iii) The selection of the period of moving averages is a difficult task. Therefore, great care has to be taken in selecting the period, particularly where there is no business cycle in the time series.

(iv) In case of non-linear trend the values obtained by this method are baised in one or the other direction.

Method of Least Squares:- This is the best method for obtaining the trend values. It provides are convenient bias for obtaining the line of best fit in a series. Line of the best fit is a line from which the sum of the deviation could be the least as compared to the sum of square of the deviations obtained by using other linear. We know that the sum of the deviations from the line of the best fit is zero. for this reasons the sum of the squares of the deviations of various point from the line of best fit is the best.

(i) $\Sigma(y - y_c) = 0$,

(ii) $\Sigma(y - y_o)^2$

$y = a + bx$

where y represents estimated value x represents the deviations in time period.

$$\Sigma y = Na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

$$\Sigma y = Na \Rightarrow a = \frac{\Sigma y}{N} = \bar{y}$$

$$\Sigma xy = b\Sigma x^2 = b = \frac{\Sigma xy}{\Sigma x^2}$$

Merits :- (i) Ris method gives the trend value for the entire time period. (ii) Ris method can be used to forecast future trend because trend line establishes a functional relationship between the value and the time. (iii) this method is a completely objective method.

Limitions :- (i) It requires many calculations and is tedious and complicated. (ii) future forecasts made by this method are based only on trend values, seasonal.., cyclical or irregular variations are ignored. (iii) If even a single item is added to the series of a new equation has to be formed.

calculate 3 yearly moving average of the data given below :-

Year :	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Sales (in millions) :	3	4	8	6	7	11	9	10	14	12

calculation of 3 year moving Average.

Solution :-

year	sales (million in Rs)	3 yearly totals	3-yearly moving Average (trends)
1980	3	15	15/3 = 5
1981	4	18	18/3 = 6
1982	8	21	21/3 = 7
1983	6	24	24/3 = 8
1984	7	27	27/3 = 9
1985	11	30	30/3 = 10
1986	9	33	33/3 = 11
1987	10	36	36/3 = 12
1988	14		
1989	12	Total	68.

steps : 1 :-

odd the values of the 1st three year (normally 1980, 1981) 1982, that is 3+4+8 = 15) and place of the total agains the medil year 1981.

steps : 2 :- leave the first year value and odd up the values of the next three 1981, 1982, 1983 that is 4+6+8 = 18 and the place of the total agains the medical year 1980.

computation of 5-yearly moving average

year	values	5-yearly m.T	5-yearly M.A (3) ÷ 5	Trend value rounded off ④
1971	105			
1972	107			
1973	109	547	109.4	109
1974	112	558	111.6	112

1975	114	569	113.8	113
1976	116	581	116.2	116
1977	118	592	118.4	118
1978	121	602	120.4	120
1979	123	611	122.2	122
1980	124	620	124	124
1981	125	628	125.6	126
1982	127			
1983	128			

Fit a trend line the following data for the least squares method.

year	1985	1987	1989	1991	1993
product (in 1tones)	18	21	23	27	16

estimate the product in 1990 and 1995.

solution :-

year	product	x	x^2	xy
1985	18	-4	16	-72
1987	21	-2	4	-42
1989	23	0	0	0
1991	27	2	4	54
1993	16	4	16	64
	105	0	40	4.

$$y = a + bx \quad \text{— ①}$$

$$a = \frac{\Sigma y}{N}, \quad b = \frac{\Sigma xy}{\Sigma x^2} \quad \text{— ②}$$

$$a = \frac{105}{5} = 21, \quad b = 4/40 = 1/10 = 0.1$$

the linear trend equation $y = a + bx$ samasion the missing value for 1986. Hence x,

$x = (x - 1989)$ $\quad$ $y = 21 + 0.1x$

$= (1990 - 1989)$ $\quad$ $y = 21 + (0.1)(1)$

$= 1$ $\quad$ $= 21 + 0.1$

$x = (x - 1989)$ $\quad$ $y = 21.1.$

$= (1986 - 1989)$