UNIT 18 CORRELATION

Objectives

After completion of this unit, you should be able to :

- understand the meaning of correlation
- compute the correlation coefficient between two variables from sample observations
- test for the significance of the correlation coefficient
- identify confidence limits for the population correlation coefficient from the observed sample correlation coefficient
- compute the rank correlation coefficient when rankings rather than actual values for variables are known
- appreciate some practical applications of correlation
- become aware of the concept of auto-correlation and its application in time series analysis.

Structure

18.1 Introduction
18.2 The Correlation Coefficient
18.3 Testing for the Significance of the Correlation Coefficient
18.4 Rank Correlation
18.5 Practical Applications of Correlation
18.6 Auto-correlation and Time Series Analysis
18.7 Summary
18.8 Self-assessment Exercises
18.9 Key Words
18.10 Further Readings

18.1 INTRODUCTION

We often encounter situations where data appears as pairs of figures relating to two variables. A correlation problem considers the joint variation of two measurements neither of which is restricted by the experimenter. The regression problem, which is treated in Unit 19, considers the frequency distributions of one variable (called the dependent variable) when another (independent variable) is held fixed at each of several levels.

Examples of correlation problems are found in the study of the relationship between IQ and aggregate percentage marks obtained by a person in SSC examination, blood pressure and metabolism or the relation between height and weight of individuals. In these examples both variables are observed as they naturally occur, since neither variable is fixed at predetermined levels.

Examples of regression problems can be found in the study of the yields of crops grown with different amount of fertiliser, the length of life of certain animals exposed to different amounts of radiation, the hardness of plastics which are heat-treated for different periods of time, and so on. In these problems the variation in one measurement is studied for particular levels of the other variable selected by the experimenter. Thus the factors or independent variables in regression analysis are not assumed to be random variables, though the dependent variable is modelled as a random variable for which intervals of given precision and confidence are often worked out. In correlation analysis, all variables are assumed to be random variables.

For example, we may have figures on advertisement expenditure (X) and Sales (Y) of a firm for the last ten years, as shown in Table I. When this data is plotted on a graph as in Figure I we obtain a **scatter diagram.** A scatter diagram gives two very useful types of information. First, *we* can observe patterns between variables that indicate whether the variables are related. Secondly, if the variables are related we can get an idea of what kind of relationship (linear or non-linear) would describe the relationship. Correlation examines the first

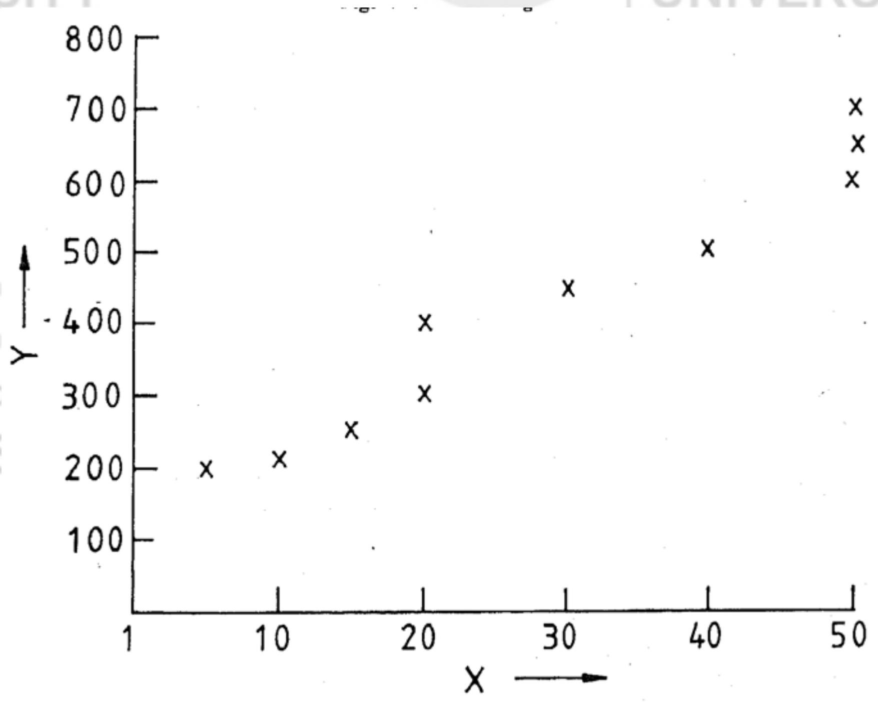
Table 1

Yearwise data on Advertisement Expenditure and Sales

Year	**Advertisement Expenditure in thousand Rs. (X)**	**Sales** in **Thousand Rs. (Y)**
1988	50	700
1987	50	650
1986	50	600
1985	40	500
1984	30	450
1983	20	400
1982	20	300
1981	15	250
1980	10	210
1979	5	200

question of determining whether an association exists between the two variables, and if it does, to what extent. Regression examines the second question of establishing an appropriate relation between the variables.

Figure I: Scatter Diagram



The scatter diagram may exhibit different kinds of patterns. Some typical patterns indicating different correlations between two variables are shown in Figure II.

What we shall study next is a precise and quantitative measure of the degree of association between two variables and the correlation coefficient.

18.2 THE CORRELATION COEFFICIENT

Definition and Interpretation

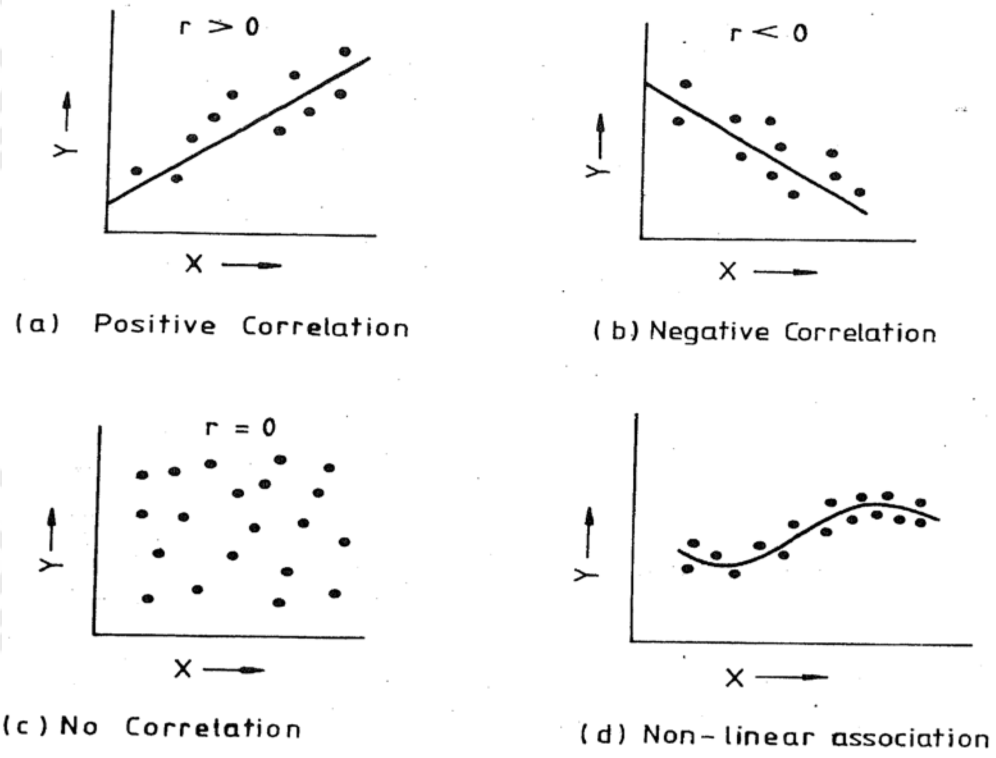
The correlation coefficient measures the degree of association between two variables X and Y. Pearson's formula for correlation coefficient is given as

$$r = \frac{\frac{1}{n}\Sigma(X - \overline{X})(Y - \overline{Y})}{\sigma_x \sigma_y}$$

……………………(18.1)

Where r is the correlation coefficient between X and Y, a% and a_y are the standard deviations of X and Y respectively and n is the number of values of the pair of

Figure II: Different Types of Association Between Variables



variable X and Y in the given data. The expression $\frac{1}{n}\Sigma(X-\overline{X})(Y-\overline{Y})$ is known as the **covariance** between X and Y. Here r is also called the Pearson's product moment correlation coefficient. You should note that r is a dimensionless number whose numerical value lies between +1 and -1. Positive values of r indicate positive (or direct) correlation between the two variables X and Y i.e. as X increases Y will also increase or as X decreases Y will also decrease. Negative values of r indicate negative (or inverse) correlation, thereby meaning that an increase in one variable results in a decrease in the value of the other variable. A zero correlation means that there is no association between the two variables. Figure H shows a number of scatter plots with corresponding values for the correlation coefficient r.

The following form for carrying out computations of the correlation coefficient is perhaps more convenient

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \qquad \ldots\ldots (18.2)$$

where

$x = X - \overline{X}$ = deviation of a particular X value from the mean $\overline{X}$

$y = Y - \overline{Y}$ = deviation of a particular Y value from the mean $\overline{Y}$

Equation (18.2) can be derived from equation (18.1) by substituting for σ_x and σ_y as follows:

$$\sqrt{\frac{1}{n}\Sigma(X-\overline{X})^2} \text{ and } \sigma_y = \sqrt{\frac{1}{n}\Sigma(X-\overline{Y})^2} \qquad \ldots\ldots (18.3)$$

Activity A

Suggest five pairs of variables which you expect to be positively correlated.

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Activity B

Suggest five pairs of variables which you expect to be negatively correlated.

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

A **Sample Calculation:** Taking as an illustration the data of advertisement expenditure (X) and Sales (Y) of a company for the 10-year period shown in Table 1, we proceed to determine the correlation coefficient between these variables :

Computations are conveniently carried out as shown in Table 2.

Table 2

Calculation of Correlation Coefficient

S.No.	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
1.	50	700	21	274	441	75,076	5,7
2.	50	650	21	224	441	50,176	4,7
3.	50	600	21	174	441	30,276	3,6
4.	40	500	11	74	121	5,476	814
5.	30	450	1	24	1	576	24
6.	20	400	–9	–26	81	676	234
7.	20	300	–9	–126	81	15,876	1,134
8.	15	250	–14	–176	196	30,976	2,464
9.	10	210	–19	–216	361	46,656	4,104
10.	5	200	–24	–226	576	51,076	5,424
Total	290	4,260	0	0	2,740	3,06,840	28,310

$$\bar{X} = \frac{290}{10} = 29$$

$$\bar{Y} = \frac{4260}{10} = 426$$

$$\therefore r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{28310}{\sqrt{2740 \times 306840}} = 0.976$$

This value of r (= 0.976) indicates a high degree of association between the variables X and Y. For this particular problem, it indicates that an increase in advertisement expenditure is likely to yield higher sales.

You may have noticed that in carrying out calculations for the correlation coefficient in Table 2, large values for x^2 and y^2 resulted in a great computational burden. Simplification in computations can be adopted by calculating the deviations of the observations from an assumed average rather than the actual average, and also scaling these deviations conveniently. To illustrate this short cut procedure, let us compute the correlation coefficient for the same data. We shall take U to be the deviation of X values from the assumed mean of 30 divided by 5. Similarly, V represents the deviation of Y values from the assumed mean of 400 divided by 10.

The computations are shown in Table 3.

Table 3

Short cut Procedure for Calculation of Correlation Coefficient

S.No	X	y	U	V	UV	U^2	V^2
1.	50	700	4	30	120	16	900
2.	50	650	4	25	100	16	625
3.	50	600	4	20	80	16	400
4.	40	500	2	10	20	4	100
5.	30	450	0	5	0	0	25
6.	20	400	-2	0	0	4	0
7.	20	300	-2	-10	20	4	100
8.	15	250	-3	-15	45	9	225
9.	10	210	-4	-19	76	16	361
10.	5	200	-5	-20	100	25	400
Total			-2	26	561	110	3,13

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

$$r = \frac{561 - \frac{(-2)(26)}{10}}{\sqrt{110 - \frac{(-2)^2}{10}}\sqrt{3136 - \frac{(26)^2}{10}}}$$

$$= \frac{566.2}{10.47 \times 55.39}$$

$$= 0.976$$

We thus obtain the same result as before.

Activity C

Use the short cut procedure to obtain the value of correlation coefficient in the above example using scaling factor 10 and 100 for X and Y respectively. (That is, the deviation from the assumed mean is to be divided by 10 for X values and by 100 for Y values.)

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

18.3 TESTING FOR THE SIGNIFICANCE OF THE CORRELATION COEFFICIENT

Once the correlation coefficient has been calculated from sample data one is normally interested in asking the question: Is there an association between the variables? Or with what confidence can we make a statement about the association between the variables?

Such questions are best answered statistically by using one of the following two commonly used procedures :

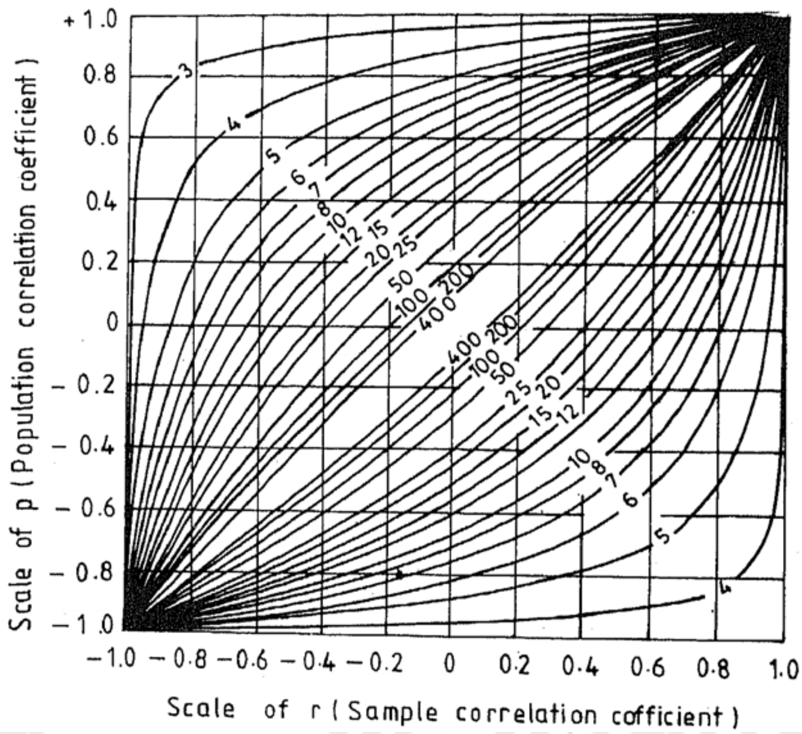
i) Providing confidence limits for the population correlation coefficient from the sample size n and the sample correlation coefficient r. If this confidence interval includes the value zero, then we say that r is not significant, implying thereby that the population correlation coefficient may be zero and the value of r may be due to sampling variability.

ii) Testing the null hypothesis that population correlation coefficient equals zero vs. the alternative hypothesis that it does not, by using the t-statistic.

The use of both these procedures is now illustrated.

The value of the sample correlation coefficient is used as an estimate of the true population correlation p. It is desirable to include a confidence interval for the true value along with the sample statistics. There are several methods for obtaining the confidence interval for p. However, the most straight forward method is to use a chart such as that shown in Figure III.

Figure III: Confidence Bands for the Population Correlation



Once r has been calculated, the chart can be used to determine the upper and lower values of the interval for the sample size used. In this chart the range of unknown values of p is shown in the vertical scale; while the sample r values are shown on the horizontal axis, with a number of curves for selected sample sizes. Notice that for every sample size there are two curves. To read the 95% confidence limits for an observed sample correlation coefficient of 0.8 for a sample of size 10, we simply look along the horizontal line for a value of 0.8 (the sample correlation coefficient) and construct a vertical line from there till it intersects the first curve for n =10. This happens for p = 0.2. This is the lower limit of the confidence interval. Extending the vertical line upwards, it again intersects the second n =10 line at p = 0.92, which represents the upper confidence limit. Thus the 95% confidence interval for the population correlation coefficient becomes

$$0.2 \leq p \leq 0.92$$

If a confidence interval for p includes the value zero, then r is not considered significant since that value of r may be due to nothing more than sampling variability.

This method of using charts to determine the confidence intervals is convenient, though of course we must use a different chart for different confidence limits (e.g. 90%, 95%, 99%).

The alternative approach for testing the significance of r is to use the formula

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} \quad \text{...... (18.3)}$$

Referring to the table of t-distribution for (n-2) degrees of freedom, we can find the critical value for t at any desired level of significance (5% level of significance is commonly used). If the calculated value oft (as obtained by equation 18.3) is less than or equal to the table value, we accept the hypothesis (H_o: the correlation coefficient equals zero), meaning that the correlation between the variables is not significantly different from zero:

Suppose we obtain a correlation coefficient of 0.2 for a sample of size 10.

$$t = \frac{0.2}{\sqrt{(1-0.04)/8}} \cong 0.577$$

And from the t-distribution with 8 degrees of freedom for a 5% level of significance, the table value = 2.306. Thus we conclude that this r of 0.2 for n = 10 is not significantly different from zero.

It should be mentioned here that in case the same value of the correlation coefficient of 0.2 was obtained on a sample of size 100 then

$$t = \frac{0.2}{\sqrt{(1-0.04)/98}} \cong 2.021$$

And the tabled value for a t-distribution with 98 degrees of freedom and a 5% level of significance = 1.99. Since the calculated t exceeds this figure of 1.99, we can conclude that this correlation coefficient of 0.2 on a sample of size 100 could be considered significantly different from zero, or alternatively that there is statistically significant association between the variables.

18.4 RANK CORRELATION

Quite often data is available in the form of some ranking for different variables. It is common to resort to rankings on a preferential basis in areas such as food testing, competitive events (e.g. games, fashion shows, or beauty contests) and attitudinal surveys. The primary purpose of computing a correlation coefficient in such situations is to determine the extent to which the two sets of rankings are in agreement. The coefficient that is determined from these ranks is known as Spearman's rank correlation coefficient, r.

This is given by the following formula

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \ldots\ldots (18.4)$$

Here n is the number of pairs of observations and d_i is the difference in ranks for the ith observation set.

Suppose the ranks obtained by a set of ten students in a Mathematics test (variable X) and a Physics test (variable Y) are as shown below :

Rank for variable X	1	2	3	4	5	6	7	8	9	10
Rank for variable Y	3	1	4	2	6	9	8	10	5	7

To determine the rank correlation, r_s we can organise computations as shown in Table 4 :

Table 4

Determination of Spearman's Rank Correlation

Individual	Rank in Maths(X)	Rank in Physics(Y)	d =Y -X	d^2
1	1	3	+2	4
2	2	1	-1	1
3	3	4	+1	1
4	4	2	-2	4
5	5	6	+1	1
6	6	9	+3	9
7	7	8	+1	1
8	8	10	+2	4
9	9	5	-4	16
10	10	7	-3	9
Total				50

Using the formula (18.4) we obtain

$$r_s = 1 - \frac{6 \times 50}{10(100-1)} = 1 - 0.303 = 0.697$$

We can thus say that there is a high degree of correlation between the performance in Mathematics and Physics.

We can also test the significance of the value obtained. The null hypothesis is that the two variables are not associated, i.e. $r_s = 0$. That is, we are interested to test the null hypothesis, H_o that the two variables are not associated in the population and that the observed value of r_s differs from zero only by chance. The t-statistic that is used to test this is

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

$$= 0.697 \sqrt{\frac{10-2}{1-(0.697)^2}}$$

$$= 2.75$$

Referring to the table of the t-distribution for n-2 = 8 degrees of freedom, the critical value for t at a 5% level of significance is 2.306. Since the calculated value of t is higher than the table value, we reject the null hypothesis concluding that the performances in Mathematics and Physics are closely associated.

When two or more items have the same rank, a correction has to be applied to $\sum d_i^2$. For example, if the ranks of X are 1, 2, 3, 3, 5, ... showing that there are two items with the same 3rd rank, then instead of writing 3, we write $3\frac{1}{2}$ for each so that the sum of these items is 7 and the mean of the ranks is unaffected. But in such cases the standard deviation is affected, and therefore, a correction is required. For this, $\sum d_i^2$ is increased by $(t^3-t)/12$ for each tie, where t is the number of items in each tie.

Activity D

Suppose the ranks in Table 4 were tied as follows: Individuals 3 and 4 both ranked 3rd in Maths and individuals 6, 7 and 8 ranked 8th in Physics. Assuming that other rankings remain unaltered, compute the value of Spearman's rank correlation.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

18.5 PRACTICAL APPLICATIONS OF CORRELATION

The primary purpose of correlation is to establish an association between any two random variables. The presence of association does not imply causation, but the existence of causation certainly implies association. Statistical evidence can only establish the presence or absence of association between variables. Whether causation exists or not depends merely on reasoning. For example, there is reason to believe that higher income causes higher expenditure on superior quality cloth. However, one must be on the guard against spurious or nonsense correlation that may be observed between totally unrelated variables purely by chance.

Correlation analysis is used as a starting point for selecting useful independent variables for regression analysis. For instance a construction company could identify factors like

- population
- construction employment
- building permits issued last year which it feels would affect its sales for the current year.

These and other factors that may be identified could be checked for mutual correlation by computing the correlation coefficient of each pair of variables from the given historical data (this kind of analysis is easily done by using an appropriate routine on a computer). Only variables having a high correlation with the yearly sales could be singled out for inclusion in a regression model.

Correlation is also used in factor analysis wherein attempts are made to resolve a large set of measured variables in terms of relatively few new Categories, known as factors. The results could be useful in the following three ways :

i) to reveal the underlying or latent factors that determine the relationship between the observed data, -

ii) to make evident relationships between data that had been obscured before such analysis, and

iii) to provide a classification scheme when data scored on various rating scales have to be grouped together.

Another major application of correlation is in forecasting with the help of time series models. In using past data (which is often a time series of the variable of interest available at equal time intervals) one has to identify the trend, seasonality and random pattern in the data before an appropriate forecasting model can be built. The notion of auto-correlation and plots of auto-correlation for various time lags help one to identify the nature of the underlying process. Details of time series analysis are discussed in Unit 20. However, some fundamental concepts of auto-correlation and its use for time series analysis-are outlined below.

18.6 AUTO-CORRELATION AND-TIME SERIES ANALYSIS

The concept of auto-correlation is similar to that of correlation but applies to values of the same variable at different time lags. Figure IV shows how a single variable such as income (X) can be used to construct another variable (XI) whose only difference from the first is that its values are lagging by one time period. Then, X and XI can be treated as two variables and their correlation found. Such a correlation is referred to as auto-correlation and shows how a variable relates to itself for a specified time lag. Similarly, one can construct X2 and find its correlation with X. This correlation will indicate how values of the same variable that are two periods apart relate to each other.

Figure IV: Example of the Same Variable with Different Time Lags

Time	X Original variable	X1 One time lag variable constructed from X	X2 Two time lags variable constructed from X
t = 1	13	8	15
2	8	15	4
3	15	4	4
4	4	4	12
5	4	12	11
6	12	11	7
7	11	7	14
8	7	14	12
9	14	12	
10	12		

One could construct from one variable another time-lagged variable which is twelve periods removed. If the data consists of monthly figures, a twelve-month time lag will show how values of 'the same month but of different years correlate with each other. If the auto-correlation coefficient is positive, it implies that there is a seasonal pattern of twelve months duration. On the other hand, a near zero auto-correlation indicates the absence of a seasonal pattern. Similarly, if there is a trend in the data, values next to each other will relate, in the sense that if one increases, the other too will tend to increase in order to maintain the trend. Finally, in case of completely random data, all auto-correlations will tend to zero (or not significantly different from zero).

The formula for the auto correlation coefficient at time lag k is:

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{n} (X_t - \overline{X})^2}$$

where

r_k denotes the auto-correlation coefficient for time lag k

k denotes the length of the time lag

n is the number of observations

X, is the value of the variable at time t and

X is the mean of all the data

Using the data of Figure IV the calculations can be illustrated.

$$\overline{X} = \frac{13 + 8 + 15 + \ldots + 12}{10} = \frac{100}{10} = 10$$

$$r_1 = \frac{(13-10)(8-10) + (8-10)(15-10) + \ldots + (14-10)(12-10)}{(13-10)^2 + (8-10)^2 + \ldots + (14-10)^2 + (12-10)^2}$$

$$= \frac{-27}{144} = -0.188$$

For k = 2, the calculation is as follows :

$$r_2 = \frac{\sum_{t=1}^{8} (X_t - 10)(X_{t-2} - 10)}{\sum_{t=1}^{10} (X_t - 10)^2}$$

$$= \frac{(13-10)(15-10) + (8-10)(4-10) + \ldots + (7-10)(12-10)}{(13-10)^2 + (8-10)^2 + \ldots + (14-10)^2 + (12-10)^2}$$

$$= \frac{-29}{144} = -.201$$

A plot of the auto-correlations for various lags is often made to identify the nature of the underlying time series. We, however, reserve the detailed discussion on such plots and their use for time series analysis for Unit 20.

18.7 SUMMARY

In this unit the concept of correlation or the association between two variables has been discussed. A scatter plot of the variables may suggest that the two variables are related but the value of the Pearson correlation coefficient r quantifies this association. The correlation coefficient r may assume values between -1 and 1. The sign indicates whether the association is direct (+ve) or inverse (-ve). A numerical value of r equal to unity indicates perfect association while a value of zero indicates no association.

Tests for significance of the correlation coefficient have been described. Spearman's rank correlation for data with ranks is outlined. Applications of correlation in identifying relevant variables for regression, factor analysis and in forecasting using time series have been highlighted. Finally the concept of auto-correlation is defined and illustrated for use in time series analysis.

18.8 SELF-ASSESSMENT EXERCISES

1 What do you understand by the term correlation? Explain how the study of correlation helps in forecasting demand of a product.

2 A company wants to study the relation between R&D expenditure (X) and annual profit (Y). The following table presents the information for the last eight years:

Year	R&D Expense (X) (Rs. in thousands)	Annual Profit (Y)
1988	9	45
1987	7	42
1986	5	41
1985	10	60
1984	4	30
1983	5	34
1982	3	25
1981		20

a) Plot the data on a scatter diagram.

b) Estimate the sample correlation coefficient.

c) What are the 95% confidence limits for the population correlation coefficient?

d) Test the significance of the cor[r]elation coefficient using a t-test at a significance level of 5%.

3 The following data pertains to length of service (in years) and. the annual income for a sample of ten employees of an industry:

Length of service in years (X)	Annual income in thousand rupees (Y)
6	14
8	17
9	15
10	18
11	16
12	22
14	26
16	25
18	30
20	34

Compute the correlation coefficient between X and Y and test its significance at levels of 0.01 and 0.05.

4 Twelve salesmen are ranked for efficiency and the length of service as below :

Salesman	Efficiency (X)	Length of Service (Y)
A	1	2
B	2	1
C	3	5
D	5	3
E	5	9
F	5	7
G	7	7
H	8	6
I	9	4
j	10	11
K	11	10
L	12	11

a) Find the value of Spearman's rank correlation coefficient, r_s

b) Test for the Significance of r_s

5 An alternative definition of the correlation coefficient between a two-dimensional random variable (X, Y) is

$$\rho = \frac{[(X - E(X))(Y - E(Y))]}{\sqrt{V(X)V(Y)}}$$

where E(.) represents expectation and V(.) the variance of the random variable. Show that the above expression can be simplified as follows :

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$$

(Notice here that the numerator is called the covariance of X and Y).

6 In studying the relationship between the index of industrial production and index of security prices the following data from the Economic Survey 1980-81 (Government of India Publication) was collected.

	70-71	71-72	72-73	73-74	74-75	75-76	76-77	77-78	78-79
Index of Industrial (1970-100)	101.3	114.8	119.6	122.1.	125.2	122.2	135.3	140.1	150.1
Index of Security Prices (1970-71-100)	100.0	95.1	96.7	116.0	113.2	96.9	102.9	107.4	130.4

a) Find the correlation between the two indices.

b) Test the significance of correlation coefficient at 0.01 level of significance.

7 Compute and plot the first five auto-correlations (i.e. up-to time lag 5 periods) for the time series given below :

t	1	2	3	4	5	6	7	8	9	10
dt	13	8	15	4	4	12	11	7	14	12

18.9 KEY WORDS

Auto-correlation: Similar to correlation in that it described the association or mutual dependence between values of the same variable but at different time periods. Auto-correlation coefficients provide important information about the structure of a data set.

Correlation: Degree of association between two variables.

Correlation Coefficient : A number lying between -1 (Perfect negative correlation) and + i (perfect positive correlation) to quantify the association between two variables.

Covariance: This is the joint variation between the variables X and Y. Mathematically defined as

$$\frac{\Sigma(X_i - \overline{X})(Y_j - \overline{Y})}{n}$$

for n data points.

Scatter Diagram: An ungrouped plot of two variables, on the X and Y axes.

Time Lag: The length between two time periods, generally used in time series where one may test, for instance, how values of periods 1, 2; 3, 4 correlate with values of periods 4, 5, 6, 7 (time lag 3 periods).

Time-Series: Set of observations at equal time intervals which may form the basis of future forecasting.

18.10 FURTHER READINGS

Box, G.E.P., and G.M. Jenkins, 1976. *Time Series Analysis, Forecasting and Control,* Holden-Day: San Francisco.

Draper, N. and H. Smith, 1966. *Applied Regression* Analysis, John Wiley: New

York.

Edwards, B. 1980. *The Readable Maths and Statistics Book,* George Allen and Unwin: London.

Makridakis, S. and S. Wheelwright, 1978. *Interactive Forecasting: Univariate and Multivariate Methods,* Holden-Day: San Francisco.

Peters, W.S. and G.W: Summers, 1968. *Statistical Analysis for Business Decisions,* Prentice Hall: Englewood-Cliffs.

Srivastava, U.K., G.V. Shenoy and S.C. Sharma, 1987. *Quantitative Techniques for Managerial Decision Making,Wiley* Eastern: New Delhi.

Stevenson, W.J. 1978. *Business Statistics-Concepts and Applications,* Harper and Row: New York.