The Motto of Our University
(SEWA)

SKILL ENHANCEMENT
EMPLOYABILITY
WISDOM
ACCESSIBILITY

JAGAT GURU NANAK DEV
PUNJAB STATE OPEN UNIVERSITY, PATIALA
(Established by Act No. 19 of 2019 of the Legislature of State of Punjab)

MASTER OF ARTS

CORE COURSE (CC): ECONOMICS

SEMESTER- I

MAEC24104T -QUANTITATIVE METHODS-I

Head Quarter: C/28, The Lower Mall, Patiala-147001
Website: www.psou.ac.in

The Study Material has been prepared exclusively under the guidance of Jagat Guru Nanak Dev Punjab State Open University, Patiala, as per the syllabi prepared by Committee of Experts and approved by the Academic Council.

COURSE COORDINATOR AND EDITOR:

Dr. Kuldeep Walia

Assistant Professor in Economics

Jagat Guru Nanak Dev Punjab State Open University, Patiala

JAGAT GURU NANAK DEV PUNJAB STATE OPEN UNIVERSITY,PATIALA

(Established by Act No. 19 of 2019 of the Legislature of State of Punjab)

PREFACE

Jagat Guru Nanak Dev Punjab State Open University, Patiala was established in December 2019 by Act 19 of the Legislature of State of Punjab. It is the first and only Open University of the State, entrusted with the responsibility of making higher education accessible to all, especially to those sections of society who do not have the means, time or opportunity to pursue regular education.

In keeping with the nature of an Open University, this University provides a flexible education system to suit every need. The time given to complete a programme is double the duration of a regular mode programme. Well-designed study material has been prepared in consultation with experts in their respective fields.

The University offers programmes which have been designed to provide relevant, skill-based and employability-enhancing education. The study material provided in this booklet is self-instructional, with self-assessment exercises, and recommendations for further readings. The syllabus has been divided in sections, and provided as units for simplification.

The University has a network of 100 Learner Support Centres/Study Centres, to enable students to make use of reading facilities, and for curriculum-based counselling and practicals. We, at the University, welcome you to be a part of this instituition of knowledge.

Prof. G.S Batra
Dean Academic Affair

•

M.A (ECONOMICS)
(MAEC24104T) QUANTITATIVE METHODS I
SEMESTER - I

MAX. MARKS:100
EXTERNAL:70
INTERNAL:30
PASS:40%
CREDITS:6

OBJECTIVES:

- To define the type and quantity of data that need to be collected.
- To Organize and summarize the data.
- To analyze the data and conclude it.
- This course introduces the students to applying mathematical techniques to economic theory also

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions each from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any 10 questions from this section.
4. The examiner shall give clear instructions to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:

Candidates are required to attempt any two questions each from sections A, and B of the question paper, and any ten short answer questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

•

Section - A

Unit 1: Unit 1: Differentiation of Functions: Simple and Partial Derivatives, Differentiation of Simple Functions – Polynomial (x) and Exponential Functions. Maxima and Minima of functions of one variable only and its applications.

Unit 2: Matrices: Definition and Types, Operations, Adjoint and inverse of a matrix (up to 3x3) Solution of simultaneous equations (up to 3) by matrix methods and Crammer's Rule.

Unit 3: Statistics: definition, importance and Scope, limitations, Distrust

Unit 4: Types of Data Collection and its Sources

Section – B

Unit 5: Classification and Tabulation of Data

Unit 6: Diagrammatic and Graphical presentation of data (with MS-Excel)

Unit 7: Sample, Population, Characteristics of good sample, type of sampling techniques, Sampling errors.

Unit 8: Index Numbers: Concepts, Simple Index Numbers, Laspeyre's, Paasche's and Fisher's index numbers only (among weighted index numbers) and Reversibility Tests.

Suggested Readings:

- A.M Goon, M.K Gupta and B. Dasgupta, fundamental of statistics Vol-I, World press Calcutta
- Gupta SC: Fundamental of statistics, S. Chand & Company. New Delhi
- Gupta, SP: Statistical Methods, S. Chand & Company. New Delhi
- Monga, GS: Mathematics and Statistics for Economics, Vikas Publishing House, New Delhi.
- Singh, D. and Chaudhary, F.S. (1986): Theory and Analysis of Sample Survey Designs. New Age International Publishers.

M.A (ECONOMICS)

(MAEC24104T) QUANTITATIVE METHODS I

SEMESTER - I

COURSE COORDINATOR AND EDITOR: DR. KULDEEP WAILA

SECTION A

Unit 1:	**Differentiation of Functions**: Simple and Partial Derivatives, Differentiation of Simple Functions – Polynomial (x) and Exponential Functions. Maxima and Minima of functions of one variable only and its applications.
Unit 2:	**Matrices**: Definition and Types, Operations, Adjoint and inverse of a matrix (up to 3x3) Solution of simultaneous equations (up to 3) by matrix methods and Crammer's Rule.
Unit 3:	Statistics: definition, importance and Scope, limitations, Distrust
Unit 4:	Types of Data Collection and its Sources

SECTION-B

Unit 5:	Classification and Tabulation of Data
Unit 6:	Diagrammatic and Graphical presentation of data (with MS-Excel)
Unit 7:	Sample, Population, Characteristics of good sample, type of sampling techniques,sampling errors.
Unit 8:	**Index Numbers**: Concepts, Simple Index Numbers, Laspeyre's, Paasche's and Fisher's index numbers only (among weighted index numbers) and Reversibility Tests.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 1: DIFFERENTIATION OF FUNCTIONS

STRUCTURE

1.0 OBJECTIVES

After studying the Unit, students will be able to:

- Meaning of determination
- Various functions in economics
- Methods to solve equations
- Limits of function and sequence
- Continuity of function and its types
- Types of differentiation

1.1 Functions in Economics

Knowledge of functions plays an important role in economics. It shows the relationship between dependent (output) variables and dependent (input) variables and is helpful to know how to obtain values for dependent variables from independent variables. The dependent variables are called effects, and the independent variables are called causes.

Generally, an independent variable is plotted on the horizontal axis, and a dependent variable is plotted on the y-axis. If one value of an independent variable maps two or more values of dependent variables, then it can't be considered a function as per the vertical line test method. These functions in set algebra can be classified as injective (one-one), many-one, surjective (onto), algebraic, into, periodic, integer, signum, polynomial, linear, identical, quadratic, cubic, composite, modulus, fractional, constant, and even and odd functions.

Functions are generally represented as follows:

$$Y=f(X)=2x^2$$

Where f of X is equal to the double of X^2. Here, we can input values for an independent variable X to obtain values for the dependent variable Y.

The algebraic operations performed to solve equations follow the BEDMAS principle. It means first completing inside bracket operations, followed by exponentiation, multiplication, division, addition, and subtraction. These are explained as follows:

Exercise 1: Let $Y=f(X)=5+3X^{2}$. Find value of f (2).

Solution: Here, we first substitute the value of X = 2 and square it, giving 4. Now we can multiply this 4 by 3 obtaining 12. Finally, we add this result to 5 giving value of 17. Hence, the value of Y at X = 2 is 17.

Exercise 2: Let $Y = f(X) = (5+3)X^{2}$. Find value of f (3).

Solution: In the above example, we need to do the bracket first. Add 5 and 3, giving a value of 8. Later, operate exponentially (X^2), put the value of X = 3, giving 9. Now multiply 8 by 9 giving 72 answers. Here, we need to be careful that $X^2=X.X$ and not 2X. The reciprocal of X means 1 divided by X and exponential means exponential power time multiplication of the value.

The values of variables can also be substituted to find the final value of any economic indicator. For example:

Exercise 3: Let, Import=X and Export=Z. The value of Import (X) is constant= Rs. 300 million and Exports in Rs. million is (Z)= 500+0.2Y. Calculate net income.

Solution: The net income can be calculated by finding the difference between exports and imports.

$$\text{Net Income} = \text{Exports} - \text{Imports}$$

$$\text{Net Income} = (500+0.2Y) - 300$$

$$= 200+0.2Y$$

We need to note that multiplication of a negative value with a positive value or multiplication of a positive value with a negative value will always give a negative result. If a negative value is multiplied even number of times, it will yield a positive value, and if a negative value is multiplied by odd number of times, it will yield a negative value. Also, if any value is multiplied with zero will yield a zero value, and if zero is divided by any value, it will also yield zero. However, any value divided by zero will yield an infinite value, which is not defined.

For example, total and average revenue can be calculated based on the quantity of goods sold at a given price. Let a firm sell Q quantities of goods for P. Total revenue is calculated as follows:

Total revenue (TR) = Price (P). Quantity (Q) sold = P.Q

The total Profit = TR-Total Cost (C) = P.Q-C

Average Profit or Per unit Profit = (P.Q-C)/Q

Students of economics need to deal with various types of functions. The cost function is a major function that gives the total cost of producing a Q number of the same goods. Generally, it consists of two costs components called fixed costs and variable costs. It is calculated as follows:

Total Cost (C)= Fixed Cost+ Variable Cost

Exercise 4: A company produces cold drinks. The building and infrastructure costs are fixed. The raw material, energy and labour costs are considered as variable costs. If the fixed cost is Rs. 20,000 and variable cost is Rs. 2 per bottle of soft drink. Calculate total cost.

Solution: The total cost is calculated as follows:

Total Cost (C)= Fixed Cost+ Variable Cost

Total Cost (C)= 20,000+2 times number of bottles produced (Q)

C=20,000+2Q

The variable cost associated with the production of one additional unit is called the marginal cost. The marginal revenue is the slope of linear revenue function, and the marginal profit is the slope of the profit function. Companies calculate the break-even point to make business decisions, where at the break-even point, profit is zero and cost is equal to revenue.

The major functions in the field of economics include constant functions, polynomials, and rationals; these are discussed in the next section.

1.1.1 Constant Function

The constant functions assume only one value. The output value is independent of the input value. The graphs of constant functions are as follows:

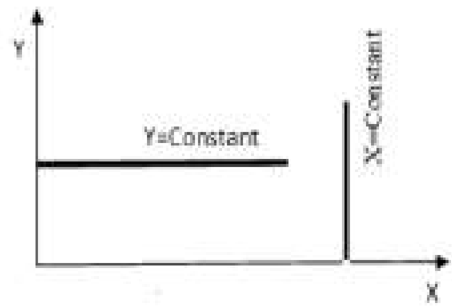


Figure 1: Constant functions

For example, the function Y = f(X) = 3 or Y = 3 shows that the domain of this function is set of all real numbers, but the co-domain is fixed equal at 3. Means f (2) = 3, f(5) = 3, and for any value of X, Y assumes only one value fixed to 3. In the case of one variable polynomial X, the non-zero constant assumes a polynomial of zero degree, and its line is parallel to the X-axis and never touches it. However, if f(X) = 0, is the identically zero function with line resting on the X-axis. Constant functions or even functions are symmetric with the Y-axis or have zero slope, as shown in the figure above. The derivative of such functions is zero. The real example is a store where every item costs a fixed price. The graph of such functions assumes a linear line parallel to the X-axis.

1.1.2 Polynomial Function

A polynomial function assumes positive powers only in the functional relationships. Generally, a polynomial is expressed as:

$$Y=f(X)=C_0+C_1X+C_2X^2+C_3X^3+C_4X^4+\ldots$$

Where C_0, C_1, C_2, etc. are constants and X and Y are independent and dependent variables. It will be a constant function if the power of X is zero and a linear function if the power of X is 1. If the power of X is 2, it is a quadratic, and if the power of X is 3, it is a cubic function.

1.1.3 Quadratic Functions

The word quadratic means square. It has a parabola graph and is useful to identify maxima and minima in economic optimisation problems by twice differentiating functions. These are useful in profit and loss, forecasting, trajectory plots, etc. The quadratic functions are expressed as:

$f(X) = aX^2 + bX + c$, where a, b, and c are constants not equal to zero. The value of X can be calculated for equation the $aX^2 + bX + c=0$ using the following formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

1.1.4 Cubic Functions

The cubic function was discovered by Scipione del Ferro, an Italian mathematician. In this function, each value of X has one unique Y-value, like a one-to-one function. It has no slope and assumes an S-shape. The change in volume of a ball or cube is a good example of cubic function. The cubic equation is indicated by the following:

$$f(X) = a X^3 + b X^2 + c X + d.$$

1.1.5 Rational Functions

A rational function is the ratio of two polynomial functions in the form $\frac{P}{Q}$ where Q is not equal to zero. Mathematically, a rational function can be represented as $f(x) = \frac{P(x)}{Q(x)}$ where $Q(x) \neq 0$. For example, $f(x) = \frac{x^2+x-1}{2x^2+2x-5}$ where $2x^2 - 2x - 5 \neq 0$, is a rational function. An asymptote of a curve is a line that shows the distance between the curve and the line approaches zero as one of the coordinates along the x-axis or y-axis approaches infinity. Rational function curves can have three types of asymptotes: horizontal asymptote, for example X=a; vertical asymptote, for example Y=b; and oblique asymptote for example Y=ax+b. The most common example of a rational function in economics is the average cost. Where, Average Cost= $\frac{C(x)}{x}$ Here, C(x) is the total cost of x number of items.

Exercise 5: A chocolate manufacturing plant has a fixed cost of ₹5000 and per unit variable cost of ₹25. Find the average cost when 200 chocolates are produced.

Solution: Here the total cost function C(x)= 5000+25x and Average Cost= $\frac{5000+25x}{x}$

Average Cost= $\frac{5000+25(200)}{200}$ = 50 per item

1.1.6 Non-algebraic Functions

Non-algebraic functions are also called transcendental functions. These transcendental functions can be expressed in algebra in the form of infinite sequences. Non-algebraic functions are polynomials in the form of trigonometric, for example, f(x)=Sin (2x+3); absolute value functions, for example, f(x)=|5|; logarithmic functions, for example, f(x)=log (x); and exponential functions, for example, $f(x)= 4x^3$; and powered as a root, for example, $f(x) = \sqrt{x}$. In these functions, the behavior of the left-hand side variable is explained by the variable(s) on the right-hand side. Knowledge of these functions will help in understanding business functional relationships for short-term and long-term decision-making.

1.2 Solving Equations

A function is used to define the relationship among dependent and independent variables. The solution of a functional relationship will be obtained with the help of an equation. An equation is a mathematical relation that shows the equality of two sides of an expression. For instance, 4X+3Y=15. The left and right sides are equal. The rate of change in the dependent variable with respect to the independent variable shall be calculated by taking the derivative of the dependent variables with respect to one of the independent variables. An equation that involves independent and dependent variables and has at least one derivate of the dependent variable with respect to the independent variable is called a differential equation. There are many methods to solve equations.

The main methods to solve equations include algebraic, graphical, substitution, elimination, and matrix methods. Here, we can add, subtract, divide and multiply both sides of an equation to simplify and solve it.

1.2.1 Algebraic Method

Consider the following example:

$$5x + 20 = 6x + 9$$

It has only one variable X. Here we can subtract 6X from both sides. This gives us the following:

$$5X + 20 - 6X = 6X + 9 - 6X$$

$$-X+20=9$$

Now we can subtract 20 from both sides yielding

$$-X + 20 - 20 = 9 - 20$$

$$-X = -11$$

Multiply both sides by -1 we get

$$-1(-X) = -1(-11)$$

$$X = 11$$

To test the value put X=11 in the equation 5X + 20 = 6X + 9 and we have 5 (11) + 20 = 6 (11) + 9 = 75. Now, both sides have value equal to 75 and solution value X = 11 is valid.

1.2.2 Graphical Method

The Graphical method is helpful in finding solutions to equations where two variables are involved because a graph paper has only two dimensions, i.e., length and breadth. This method helps in investigating economic phenomena using diagrams and graphs. The abstract graphical relations help strategists manipulate and gain better insights into the business environment.

For example:

$$Y=2X-1 \text{ and } 2Y= -X+8$$

Here, we can solve this function by equating these two equations and simplifying as follows:

$$2X-Y-1=0 \quad \text{---Equation (1)}$$

$$X+2Y-8=0 \quad \text{---Equation (2)}$$

$$Y=1-2X \quad \text{--from equation (1)}$$

$$Y = \frac{8-X}{2} \quad \text{---from equation (2)}$$

The graphical behaviour of variable Y can be plotted for equations (1) and (2). As it is a linear function with the exponential power of variables in both equations equal to 1. We need to plot at least two points to draw a line. Here, we can assume Y = 0 to get one point for X variable and assume X = 0 to get another point for Y the variable.

The X and Y values for equation (1) are as follows:

Variable	X (when Y=0)	Y (When X=0)
Y	0	1
X	1/2	0

Similarly, the X and Y values for equation (2) are as follows:

Variable	X (when Y=0)	Y (When X=0)
Y	0	4
X	8	0

Now we can plot the graph as shown below:

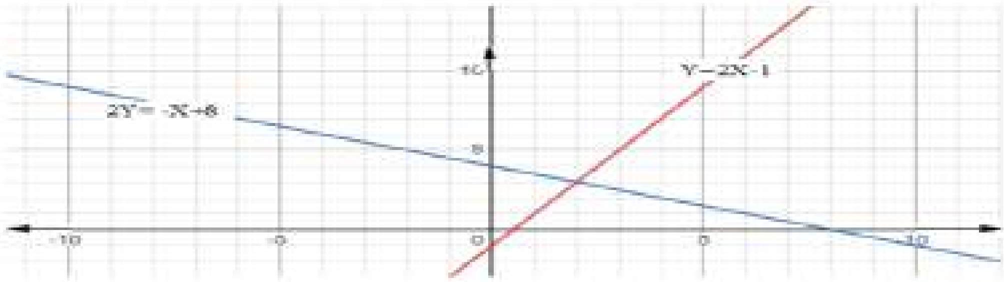


Figure 2: Graph of functions Y=2X-1 and 2Y= -X+8

1.2.3 Elimination by Substitution Method

This method involves eliminating one variable from another equation by substituting its value in the equation. For example, we have two linear equations:

$$3x+2y = 11 \text{ and } 2x-y = 4$$

We have

$$3x+2y = 11 \quad \dots (1)$$

$$2x-y = 4 \quad \dots (2)$$

We can modify the equation (2) as follow:

$$x = (y+4)/2 = y/2 + 2 \dots (3)$$

The value of x obtained in equation (3) will be substituted in equation (1) as follow:

$$3(Y/2 + 2) + 2y = 11$$

$$3Y/2 + 6 + 2Y = 11$$

$$Y(3/2 + 2) = 11-6$$

$$7Y/2 = 5$$

$$Y = 5*2/7 = 10/7$$

Now, using the distributive property of equations, we can substitute this value of Y in either equation (1) or equation (2). Let us substitute this value in equation (2), and we have the following:

$$2x - y = 4$$

$$2x - 10/7 = 4$$

$$2x = 4 + 10/7 = 38/7$$

$$x = 38/(2*7) = 38/14 = 19/7$$

Hence, x = 19/7 and y = 10/7 satisfy both the equations (1 and 2).

1.2.4 Matrix Method

The graphical method is limited to solving equations involving a maximum of two variables. The elimination method is quite a lengthy exercise to find the values of variables satisfying all equations. The matrix method has an advantage over these two methods as it is quick and easy to find solutions to equations involving two or more variables. This method is based on the Gaussian elimination method to find solutions. The following are important steps to solve equations using the matrix method:

1. Write variables in equations in the proper order.
2. The coefficients and constructs should be written on their respective sides.
3. It should be possible to find the inverse of the matrix, i.e., determinant of the matrix should not be zero.

The following example will explain this method:

Let there be three equations,

$$a_1x+a_2\ y+a_3z=k_1$$

$$b_1x+b_2\ y+b_3z=k_2$$

$$c_1x+c_2\ y+c_3z=k_3$$

Where a, b, c, and k are constants and x, y, and z are variables.

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

$$AX=K$$

Where Matrix A presents constants, matrix X presents variables and matrix K presents resource constraints.

$$A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

$$X = \begin{pmatrix} x \\ y \end{pmatrix}$$

z

$$K = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$$

$$X = A^{-1}K$$

Where $A^{-1} \neq 0$ and A^{-1} =Adjoint A/Determinant of A

For example, let us have,

$$2x+y+2z=2$$
$$2x-y+z=12$$
$$x+3y-z=6$$

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 3 & -1 \end{bmatrix} \qquad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad K = \begin{bmatrix} 2 \\ 12 \\ 6 \end{bmatrix}$$

A^{-1} =Adjoint A/Determinant of A

$|A|$ = Determinant of A= 2(-1*-1 - 3*1) - 2(1*-1 - 2*3) + 1(1*1 - (-1) *2)

=2(1-3) - 2(-1-6) + 1(1+2) = -4+14+3=13

The value of the determinant of A is not equal to zero, hence, we can find its inverse. We need to find determinant of the 2X2 matrix, which is as follows:

Row 1 elements are represented by 'a' with subscripts. The first subscript shows the row and the second subscript shows the column as follows:

a_{11}= (-1*-1-3*1) = 1-3=-2; a_{12}= (2*-1-1*1) =-2-1=-3 a_{13}= (2*3-(-1) *1) =6+1=7

Similarly for Row 2 and Row 3, co-factors are as follows:

a_{21}= (1*-1-3*2) = -1-6=-7; a_{22}= (-1*2 -1*2) =-2-2=-4 a_{23}= (2*3-1*1) =6-1=5

a_{31}= (1*1-2*-1) = 1+2=3; a_{32}= (1*2 -2*2) =2-4=-2 a_{33}= (2*-1-2*1) =-2-2=-4

The matrix of these determinants with cofactors shall be calculated by summing the row number and column number of the respective element. If the sum is even, we need to multiply the values of determinants obtained by the + sign, and if the sum of row plus column position comes to an odd number, we need to multiply by the - sign. Now the cofactor matrix obtained is as follows:

Cofactor matrix of A=

$$\begin{matrix} -2 & 3 & 7 \\ 7 & -4 & -5 \\ 3 & 2 & -4 \end{matrix}$$

The adjoint of A is formed by changing the first row into the first column, the second row into the second column, and the third row into the third column as follow:

$$\text{Adjoint of A=} \begin{pmatrix} -2 & 7 & 3 \\ 3 & -4 & 2 \\ 7 & -5 & -4 \end{pmatrix}$$

$$A^{-1} = \text{Adjoint A/Determinant of A} = 1/13 * \begin{pmatrix} -2 & 7 & 3 \\ 3 & -4 & 2 \\ 7 & -5 & -4 \end{pmatrix}$$

Now $X = A^{-1} K$. Here, we need to multiply 3by3 matrix with column matrix K the multiplication result is as follows:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1/13 \begin{pmatrix} -2 & 7 & 3 \\ 3 & -4 & 2 \\ 7 & -5 & -4 \end{pmatrix} \begin{pmatrix} 2 \\ 12 \\ 6 \end{pmatrix} = \begin{pmatrix} 98/13 \\ -30/13 \\ -70/13 \end{pmatrix}$$

Now, putting these values in all three equations in these questions makes the left side equal to the right side.

1.3 Limit Of A Function

The idea of a limit is the basis of calculus, and a limit is the function value as that function approaches the specific value of a decision variable. The function can be continuous or discontinuous. A function is continuous if the right-side limit and left-side limit values of a function at a specific point x = k exist and are equal. Otherwise, it is discontinuous. Let us take a function f(x)=x-2/x-2; here the function is not defined at x=2 because it will give a value of zero divided by zero, which is not defined. Otherwise, for any value of x, the function f(x) approaches 1. This is the limit; as x approaches 2 but is not equal to 2, the value of the function is 1. In another example, let us have the following function:

$$S(x) = \begin{cases} x^2 & x \neq 2 \\ 1 & x = 2 \end{cases}$$

This function is discontinuous at x = 2 as the given value of a function at x = 2 is 1. However, as x→ 2 the value of the function approaches 4. This can help us to understand the limits and continuity. This function is discontinuous at x = 2. A limit is represented as shown below:

$$\lim_{x \to c} = f(x) = k$$

In the above function, the value of f(x) approaches k as x approaches c. Limit is the output value of a function or sequence is an important part of integration and differentiation to graph functions to understand its real time behaviour. A limit could be one-sided, which means that a function or sequence gives output value as the input approaches a particular value either from the right-side (above) or the left side (below). These left and right-side limits are written as follows:

$$\text{Lim}_{x \to c^+} f(x) = k \text{ or } f(x) \to k \text{ as } x \to c^+$$

The above-mentioned limit is right-sided and left sided limit can be presented as follows:

$$\text{Lim}_{x \to c^-} f(x) = k \text{ or } f(x) \to k \text{ as } x \to c^-$$

For example, consider a function f(x)=1/x. This function is not defined at x=0 and it is discontinuous at x=0. However, as x approaches 0 from +∞ the value of function, f(x) approaches +∞. The right-sided limit of a function can be written as:

$$\lim_{x \to 0^+} f(x) = +\infty$$

The left-sided limit of a function can be written as:

$$\lim_{x \to 0^-} f(x) = -\infty$$

The above function has different values from the right-side and left-side. Hence, this function is not well defined at x=0.

Let us consider a function $f(x) = \frac{1}{x}$

This function is undefined at x=0. The graph of this function is shown below:

Table 1: Right sided and left sided limits

Variable & Function	Right-side limit					Right-side limit			
X	1	0.6	0.4	0.3	0.0001	0.001	-0.08	-0.1	-0.4
F(x)	1	1.67	2.5	3.3	10000	-10000	-12.5	-10	-2.5

The graph of the above function is shown below:

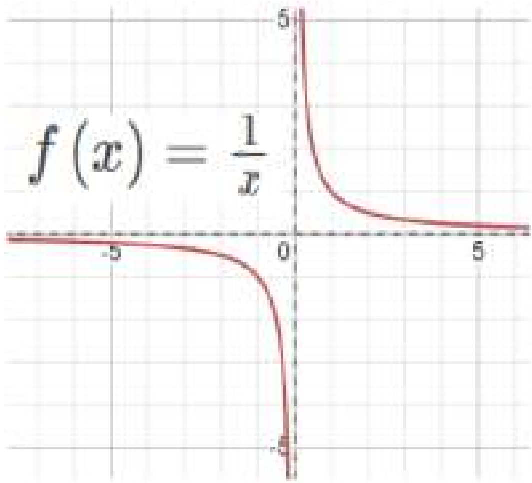


Figure 2: Graph of f(x)=1/x function

The above function, f(x)=1/x is discontinuous as x-→0 both from the left side as well as the right-side.

Similarly, consider another example f(x) = x^3. Here, also we can consider both left-side and right-side limits.

Table 2: Right-sided and left-sided limits of f(x)= x^3

	Right-side limit (x→+∞)					Left-side limit(x→-∞)				
X	0	1	5	10	100	-1	-5	-10	-100	-999
F(x)	0	1	125	1000	1000000	-1	-125	-1000	-1000000	-997002999

The plot of function f(x)=x^3 is shown in figure below:

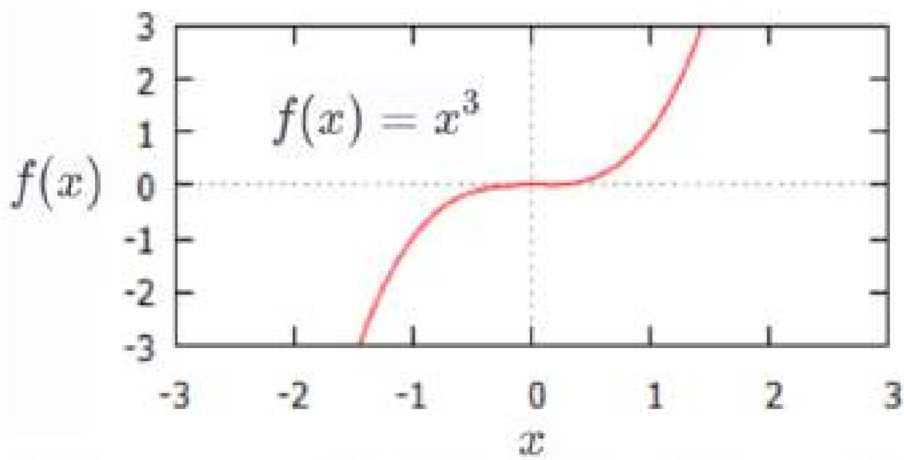


Figure 3: Graph of f(x)= x^3 function

It is evident from the above graph that as the value of x decreases from left, f(x) becomes more negative and as the value of x increases from right, it becomes more positive.

1.3.1 Properties of Limits

If the limits of two functions f(x) and h(x) exist, then the following properties can be applied as $x \to a$:

1. Law of addition: $\lim_{x\to a} [f(x) + h(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} h(x)$

2. Law of subtraction: $\lim_{x\to a} [f(x) - h(x)] = \lim_{x\to a} f(x) - \lim_{x\to a} h(x)$

3. Law of multiplication: $\lim_{x\to a} [f(x).h(x)] = \lim_{x\to a} f(x).$ limx $\lim_{x\to a} h(x)$

4. Law of Division: The division is applicable when limit of denominator is not equal to zero.

$$\lim x \to a \quad \frac{f(x)}{h(x)} = \frac{\lim x \to a \quad f(x)}{\lim x \to a \quad h(x)}$$

5. Law of root: The limit outside the root is applicable to the variable inside the root power

$\lim_{x\to a} f(x) = \frac{\sqrt{\lim x\to a \quad x^3}}{c}$ Here, c is a constant.

6. Law of power: According to the law of power, limit of a function raised to any power is equal to the limit of function raised to the same power.

$\lim_{x\to a} (f(x))^n = (\lim_{x\to a} f(x))^n$, where n is an integer.

7. Law of constant: According to law of constant, the limit of a constant function is equal to the constant value of the function. Let us suppose a constant function c, $\lim_{x\to a} f(c) = c$

8. Limit of constant multiplied function: If a function is multiplied with a constant has limit, it is equal to the limit value of function multiplied by the same constant:

$$\lim_{x \to a} c.f(x) = c. \lim_{x \to a} f(x)$$

9. Limits of trigonometric functions: The limits of trigonometric functions as $\lim x \to \pm\infty$ is not defined. However, limits of trigonometric numbers, if defined has a specific value. For example, $\lim_{x \to a} \text{Sin}(x) = \text{Sin } a$; $\lim_{x \to a} \text{Cos}(x) = \text{Cos } a$; $\lim_{x \to a} \tan(x) = \tan a$; $\lim_{x \to a} \text{Cosec}(x) = \text{Cosec } a$; $\lim_{x \to a} \text{Sec}(x) = \text{Sec } a$; $\lim_{x \to a} \text{Cot}(x) = \text{Cot } a$.

10. Limits of inverse trigonometric functions: Inverse trigonometric functions are called as Arc functions. Arcsine ($\sin^{-1}(-x) = -\sin^{-1}(x), x \in [-1, 1]$); Arccosine ($\cos^{-1}(-x) = \pi - \cos^{-1}(x), x \in [-1, 1]$); Arctangent ($\tan^{-1}(-x) = -\tan^{-1}(x), x \in R$); Arccotangent ($\cot^{-1}(-x) = \pi - \cot^{-1}(x), x \in R$); Arcsecant ($\sec^{-1}(-x) = \pi - \sec^{-1}(x), |x| \geq 1$); and Arccosecant ($\text{cosec}^{-1}(-x) = -\text{cosec}^{-1}(x), |x| \geq 1$).

1.4 Limit of a Sequence

A sequence is the chronological arrangement of two or more things, either in finite or infinite form. A finite sequence has an ending term, but an infinite sequence has no ending term. For example, if the 1, 2, 3, 4, 5 sequence has five entries and the last entry is 5, it is a finite sequence; however, a sequence for example 1, 2, 3, 4, 5,... has infinite terms, called an infinite sequence. Each entry in a sequence is called a term. A series can be the sum of all items in a sequence. Sequences can be also be classified as arithmetic, geometric, harmonic, and fibonacci numbers. Arithmetic sequences are formed by subtracting or adding a specific number from the preceding number of the arithmetic sequence. On the other side, harmonic sequences have reciprocal terms with the arithmetic sequence. The geometric sequences are formed by dividing or multiplying a specific number with the preceding number of the geometric sequence. A sequence formed by adding two preceding elements that start with 0 and 1 is called a fibonacci number. The following are a few examples of these sequences:

1.4.1 Arithmetic Sequences

An arithmetic sequence can have terms like $a, a + d, a + 2d, + a(n-1) d$. Where n is the nth term of the sequence and d is the difference between the succeeding and preceding terms. Any term (a_n) in the arithmetic sequence can be found using $a_n = a+(n-1) d$. The sum of all terms $Sn = \frac{n(2a+(n-1)d)}{2}$ The limit of a sequence can also be calculated using limits as the sequence approaches a specific value. For example, we have a sequence 1, 3/2, 5/3, 7/4, 9/5, 11/6, 13/7…2-1/n.

$$\lim_{n \to \infty} (2-1/n) = 2$$

As n approaches ∞ the limit 1/n approaches zero, and the value of the limit function is equal to 2. However, this sequence does not have any term=2. Hence, it is evident that a limit may or may not contain a limit as a term in the sequence.

1.4.2 Geometric Sequences

A geometric sequence can assume the terms a, ar, ar^2, ar^3, …$ar^{(n-1)}$. The limits of a geometric sequence can be calculated as shown in the following example:

Let us have an infinite series with sum G_∞ =1/2+ 1/4+ 1/8+1/16+1/32 … =1

The sum of this series can be calculated using the following formula:

$$G\infty = \frac{a(1-r^n)}{1-r}$$

If the value of r>1, r^n will grow exponentially as value of r increases to an infinitely large value and will decrease exponentially when r<1. So, in the limit notation, we can write when r<1 as:

$$\lim_{n \to \infty} r^n = 0 \text{ and can take limit value of } \quad G\infty = \frac{a}{(1-r)}$$

In a similar fashion, we can find limits to harmonic sequence and fibonacci numbers.

1.5 Continuity of Functions

We can see continuity in many natural phenomena around us. For example, water is flowing in rivers, blood circulates in the human body, and growth of plants, etc. It is just like drawing a graph without lifting the pencil. A continuous function in mathematics holds the following properties at a point x = b:

1. f(b) has a finite value, and it exists.

2. $\lim_{x \to b} f(x) = f(b)$

3. The right-side and left-side limits are finite and equal.

A function would be continuous in an interval if all three conditions are met. There are two important theorems regarding the continuity of functions. These are discussed in the next section.

1.5.1 Intermediate Value Theorem

According to this theorem, a function f(x) is continuous in the interval [a, b], if there exists a minimum and maximum value in the interval [a, b] as follows:

$$a \le x_{min} \le b \text{ and } a \le x_{max} \le b$$

Also, the function f(x) has maximum $f(x_{max})$ and minimum $f(x_{min})$ values as follow:

$$f(x_{min}) \le f(x) \le f(x_{max}) \text{ when } a \le x \le b.$$

This theorem advocates that continuous functions always have minimum and maximum values.

1.5.2 Extreme Value Theorem

This theorem helps to find minimum and maximum values of a continuous function. According to this theorem, if a real-valued function is continuous in the closed interval [c, d] where c<d, then there exist two real numbers 'a' and 'b' in this closed interval [c, d] such that f(a) is the minimum and f(b) is the maximum value of the function f(x). Mathematically, it is shown as follows:

$$f(a) \le f(x) \le f(b), \forall x \in [a, b]$$

1.6 TYPES OF DISCONTINUITY

There are three different types of discontinuity: removable discontinuity, infinite discontinuity, and jump discontinuity. These are discussed as follows:

1.6.1 Removable Discontinuity

This type of discontinuity occurs when $\lim_{x \to a} f(x) \ne f(a)$. For example, show that the following function is not continuous:

$$f(x) = \frac{4x + 10}{x^2 - 2x - 15}$$

Now, we can factorise the denominator $x^2 - 2x - 15 = (x-5)(x+3)$. As we know, rational functions become discontinuous when divided by zero. The denominator shows that function is discontinuous when (x-5) (x+3) =0. It will happen when x=5 and/or x=-3.

Exercise 6: Show that the $f(k) = 2k^3 - 5k^2 - 10k + 5$ has a root within the interval range [-2, 2]

Solution: Here we need to check if f(k)=0 between interval [-2, 2]. Now suppose a number t between such that -2< t < 2 and f(k) = 0

The value of $f(-2) = 2(-2)^3 - 5(-2)^2 - 10(-2) + 5 = -16 - 20 + 20 + 5 = -11$

and $f(2) = 2(2)^3 - 5(2)^2 - 10(2) + 5 = 16 - 20 - 20 + 5 = -19$

Now, as per Intermediate Value theorem t point that shall exist between the interval -2 < t < 2

This solution did not provide us any such point. Now, we can use graphical method to locate this point as shown in figure below:

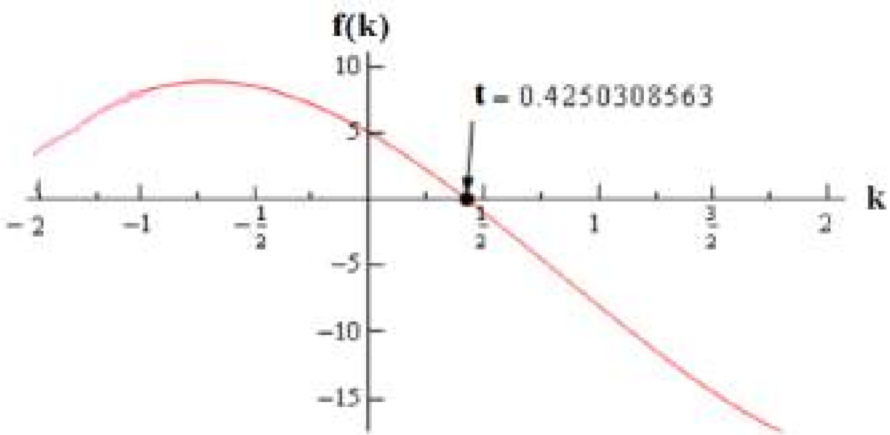


Figure 4: Graph of function f(k)= 2k³ - 5k² - 10k + 5

Let us take another example,

$$f(x) = \frac{(x+3)(x+1)}{(x+1)}$$

This function looks like f(x)=x+3. However, if we plot the curve, we will find that the function is not continuous at x=-1 which is evident from the following graph:

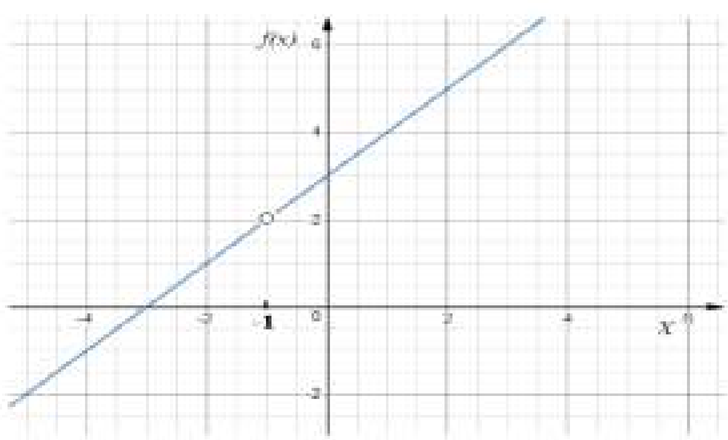


Figure 5: Plot of removable discontinuity graph

The hole in the graph is shown by the circle on the line. It shows the discontinuity of this function.

1.6.2 Infinite Discontinuity

It is also called as essential discontinuities due to severity of the infinite discontinuity points. This type of discontinuity arises when a function is not defined at a specific value, with the right-side and left-side limits tending to ±∞ values. Economists can classify it as large a scale discontinuity or a small-scale discontinuity. They can split this discontinuity to find a solution to their problem. For example, we have a function:

$$f(x) = \frac{1}{x}$$

With limit x→0 either from the right-side or from the left-side, the function is not defined at x = 0 which is evident from the graph below showing infinite discontinuity. The limit on the left side

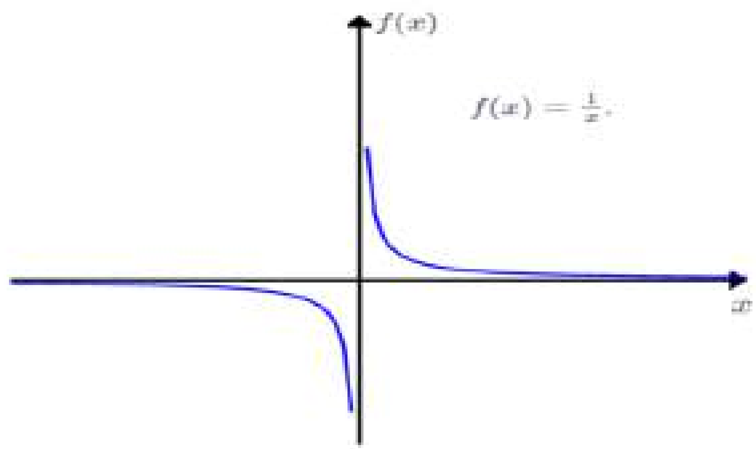


Figure 6: Graph of infinite discontinuity of function f(x)=1/x

of y-axis, as the value of x approaches 0, the infinitely smaller on the negative y-axis. Similarly, on the right-side of y-axis as x approaches zero, the function becomes infinitely large. The infinite discontinuity shall be visible if we the graph of function under evaluation. In this example, the function has infinite discontinuity at a value of x=0. It is called a pole in the graph.

1.6.3 Jump Discontinuity

This type of discontinuity is never infinite because the limits from both the right and left are real numbers. Let us consider the example given below:

0, x<0

$f(x) = \quad 2, x \geq 0$

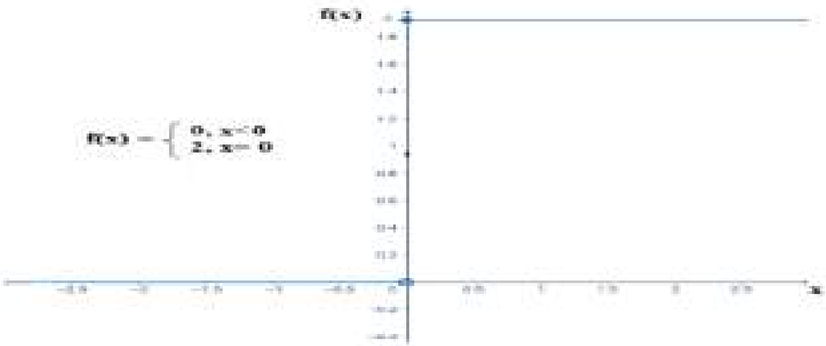


Figure 7: Graph of jump discontinuity

1.7 Differentiation

Differentiation is the finding derivative of a function. It is used to know the instantaneous (very small interval) rate of change of any quantity. For example, velocity of a scooter tells the rate of change in distance over time. Differentiation in mathematics is defined as the rate of change of an independent variable.

If a function y = f(x) goes through an infinitesimal change at any point, then it can show as follows:

$$y = f(x)$$

$y + \Delta y = f(x + \Delta x)$, then limiting relative change is given by

$$\text{Lim } \Delta \rightarrow 0 \ \frac{(y+\Delta y)-y}{\Delta x} = \frac{f(x+\Delta x)-f(x)}{\Delta x} = dy/dx$$

We need to note here that a differentiable function at any point is continuous at that point. However, a continuous function at a point may or may not be differentiable at that particular point. The following are important rules of differentiation:

Let the two functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $k : \mathbb{R} \rightarrow \mathbb{R}$ are differentiable functions at point x, then the following properties hold:

1. Differentiation of a polynomial: $(x) = x^n \Rightarrow f'(x) = nx^{n-1}$

$$\text{Given } (x) = x^{-10} \Rightarrow f'(x) = -10x^{-10-1} = -10x^{-11}$$

2. Differentiation of Constant multiplier of a function: Let α be a constant multiplied to a function $\alpha(x)$. The derivative $(\alpha(x))' = \alpha f'(x)$, if $\alpha \in \mathbb{R}$ (constant).

$$\text{Given } (x) = 3x^5 \Rightarrow f'(x) = 3 .5. x^{5-1} = 15x^4.$$

3. Differentiation of addition or subtraction of functions: The derivative of the sum or subtraction of two or more functions is equal to the sum or subtraction of their individual derivatives:

$$(f(x) \pm k(x))' = f'(x) \pm k'(x) \text{ (addition/subtraction).}$$

Given $(x) = (x^2 \pm 3x^6)$:

$$f'(x) = f'(x^2) \pm f'(3x^6)$$

$$= 2x + 3.6\, x^{6-1}$$

$$= 2x \pm 18\, x^5$$

Given $(x) = (x^3 + 3x^5)^4$:

$$\text{Let } (x) = (x^3 + 3x^5) \Rightarrow g'(x) = 3x^2 + 15x^4$$

$$\therefore f'(x) = 4\,(x^3 + 3x^5)^{4-1}.\,(3x^2 + 15x^4)$$

$$= 4(x^3 + 3x^5)^3.\,(3x^2 + 15x^4)$$

4. Derivative of multiplication of two functions: Let f and g be two functions and their product is fg. The derivative of two multiplying functions is given by:

$$(f(x).\,k(x))' = f'(x)\,k(x) + f(x)\,k'(x)$$

Given $(x) = (x)^3\,.\,(5x)$:

$$\text{We know that } f'(x) = h(x).\,g'(x) + g(x).\,h'(x)$$

$$= x^3(5) + 5x(3x^2)$$

$$= 5x^3 + 15x^3$$

$$= 20x^3$$

5. Derivative of dividing functions: The derivative of division of the functions is given as:

$$\textbf{derivative } (\mathbf{f(x)/k(x)}) = \frac{\mathbf{f'(x).\,k(x) - f(x).\,k'(x)}}{\textbf{square of } \mathbf{k(x)}}$$

Given $f(x) = \frac{x+1}{x-2}$ Where $\neq 2$

As we know that $f'(x) = [h(x).\,g'(x) - (x).h'(x)] \,/\, [h(x)]^2$

Let $s(x) = x + 1$ and $t(x) = x - 2$

Differentiate them individually we have. $s'(x) = 1$ and $t'(x) = 1$ and substitute the value in the formula we get the following:

$$f'(x) = \frac{1.(x-2)-(x-1).1}{(x-2).(x-2)}$$

$$= \frac{x-2-x+1}{(x-2).(x-2)}$$

$$= -1/(x-2)^2$$

1.7.1 Types of Differentiation

There are five basic types of functions, where we can apply differentiation. These are trigonometric functions, algebraic functions, logarithmic functions, exponential functions, and mixed functions. The differentiation of these functions is discussed as follows:

Differentiation of trigonometric functions: The differentiation of trigonometric functions is shown below:

SN	Function f(x)	Differentiation f '(x)	SN	Function f(x)	Differentiation f '(x)
1.	$\sin x$	$\cos x$	7.	$\cos x$	$-\sin x$
2.	$\tan x$	$\sec^2 x$	8.	$\cot x$	$-\csc^2 x$
3.	$\sec x$	$\sec x \tan x$	9.	$\csc x$	$-\csc x \cot x$
4.	$\sin^{-1} x$	$\frac{1}{\sqrt{(1-x^2)}}$	10.	$\text{Cos}^{-1} x$	$-\frac{1}{\sqrt{(1-x^2)}}$
5.	$\tan^{-1} x$	$1/(1+x^2)$	11.	Cot^{-1}	$-1/(1+x^2)$
6.	$\sec^{-1} x$	$\frac{1}{\lvert x\rvert\sqrt{(1-x^2)}}$	12	$\text{Csc}^{-1} x$	$-\frac{1}{\lvert x\rvert\sqrt{(1-x^2)}}$

Exercise 7: Given f (x)=Cos x- Sin x, find its derivative.

Solution: Let f(x) = cos x; and g(x) = sin x

Using the subtraction rule of differentiation,

d/dx [f(x) – g(x)] = d/dx f(x) – d/dx g(x)

d/dx (cos x – sin x) = d/dx (cos x) – d/dx (sin x)

= - sin x – cos x

Differentiation of algebraic functions:

Differentiation Rule	f (x)	f '(x)
Sum & Difference	$\frac{d[f(x) \pm g(x)]}{dx}$	$\frac{d}{dx} f(x) \pm \frac{d}{dx} g(x)$
Product	$\frac{d[f(x).g(x)]}{dx}$	$g(x) \frac{d}{d(x)} f(x) + f(x) \frac{d}{d(x)} g(x)$
Quotient	$\frac{d[f(x)]}{g(x)}$	$\frac{g(x) \frac{f(x)}{dx} - f(x). \frac{g(x)}{dx}}{[g(x)]^2}$

Exercise 8: Given f (x)= 3x + (1/x), find its derivative.

Solution: Let us assume two parts of this function g(x) = x and h(x) = 1/x

According to sum rule of differentiation the derivative is equal to the sum derivative of individual terms. Hence,

d/dx [g(x) + h(x)] = d/dx g(x) + d/dx h(x)

d/dx [3x + (1/x)] = d/dx (3x) + d/dx (1/x)

$= 3 + (-1/x^2)$

$= 3 - (1/x^2)$

Exercise 9: Given f (x) = $(4x^3 - 6x + 1)(x + 1)$, find its derivative.

Solution: Let us assume two parts of this function g(x) = x and h(x) = 1/x

Let f(x) = $(4x^3 - 6x + 1)$ and g(x) = (x + 1).

According to the product rule of differentiation,

d/dx [f(x). g(x)] = f(x) [d/dx g(x)] + g(x) [d/dx f(x)]

$= (4x^3 - 6x + 1)$ [d/dx (x + 1)] + (x + 1) [d/dx $(4x^3 - 6x + 1)$]

$= (4x^3 - 6x + 1)(1 + 0) + (x + 1)[4(3x^2) - 6(1) + 0]$

$= (4x^3 - 6x + 1) + (x + 1)(12x^2 - 6)$

$= 4x^3 - 6x + 1 + 12x^3 - 6x + 12x^2 - 6$

$= 16x^3 + 24x^2 - 12x - 5$

Exercise 10: Given f (x) = $3x^2/(x+1)$, find its derivative

Solution: Here, f(x) = $3x^2/(x + 1)$

$g(x) = 3x^2$ $\Rightarrow g'(x) = 6x$

$h(x) = (x + 1)$ $\Rightarrow h'(x) = 1$

According to the quotient rule of differentiation $\Rightarrow f'(x) = [h(x)g'(x) - g(x)h'(x)]/[h(x)]^2$

$\Rightarrow f'(x) = [(x+1).6x - 3x^2 \cdot 1] / (x + 1)^2$

$\Rightarrow f'(x) = (6x^2 + 6x - 3x^2) / (x + 1)^2$

$\Rightarrow f'(x) = (3x^2 + 6x)/(x + 1)^2$

Differentiation of logarithmic functions: Chain rule is used for differentiation logarithmic functions. Important logarithmic differentiation formulas are given below:

1. $\log xy = \log x + \log y$

2. $\text{Log } x/y = \log x - \log y$

$\log x^y = y \log x$

$\log_y x = \log (x) / \log (y)$

Note: Derivative of exponential functions is calculated using logarithmic formulas.

Exercise 11: Given, $y = x^x$, find its derivative.

Solution: Take log on both sides: $\log(y) = \log(x^x)$

Using logarithmic rules $\Rightarrow$ $\log(y) = x \cdot \log(x)$ [Using property $\log(a^x) = x \cdot \log(a)$]

Step 3: Now differentiate the equation with respect to x $\Rightarrow$

$$\frac{d}{dx}\log(y) = \frac{d}{dx}(x.\log(x))$$

$$\frac{d}{dx}\log(y) = x.\frac{d}{dx}\log(x) + \log(x).\frac{dx}{dx}$$

$$\frac{1}{y}\frac{dy}{dx} = x.\frac{1}{x} + \log(x)$$

$$\frac{dy}{dx} = y.(1 + \log(x))$$

$$\frac{dy}{dx} = x^x(1 + \log(x))$$

1.8 Maxima And Mimima

Maxima and minima show the maximum or minimum value of a function. The following procedure is adopted to find these two points:

1. Find the derivative of the given function.

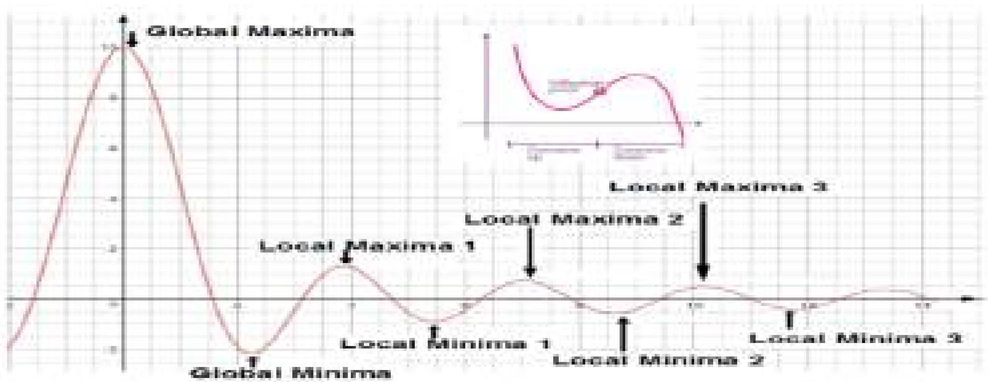
2. Put this derivative = 0.

3. Find a second differentiation.

4. Put these values in the second derivative. If second derivative $f''(x)<0$, is the maxima, and if $f''(x)>0$, is the minima.

5. To find maximum and minimum values put the calculated values in the function f(x).

Maxima and minima are classified (**Figure 8**) as local maxima or minima and absolute or global maxima and minima. Local maxima or minima give the maximum or minimum value of a point when compared to other points in the domain of a function. The global maxima or minima is a point that doesn't have any points in the domain of the function whose value is greater or lesser than the value of the global maxima or minima. The point of inflexion shows the change from maxima to minima or vice versa.The knowledge of maxima and minima can help economists to know how profit functions limit the resources or expenditures.



(Source: Byjus and topper)

Figure 8: Maxima and minima of a function

Exercise 12: Let a function $y = 3x^3 - 2x^2 + 8$, find maxima and minima.

Solution: Let f (x)=y= $3x^3 - 2x^2 + 8$

Calculate first derivative, $f'(x) = 3.3\,x^2 - 2.2x^1 + 0$

$$=9x^2-5x=x\,(9x-5)$$

$$\text{Put } f'(x)=0=x(9x-5)=0$$

$$\Rightarrow \text{either } x=0 \text{ and or } 9x-5=0$$

$$9x=5 \Rightarrow x=5/9$$

Now we have two values of x where x=0 and x=5/9

Take second derivative and put the values

$f''(x)= d(9x^2-5x)/dx =18x-5$

$f''(x=0) =18x-5=18.0-5= -5 <0$ (x=0 is point of maxima)

$f''(x=5/9) = (18.5/9)-5=10-5=+5>0$ (x=5/9 is point of minima)

Exercise 13: Let $f(x)= x^4 - 54x^2+11$, find point of inflexion.

Solution: $f(x) = x^4 - 54x^2+11$ (given)

The first derivative of the function $f'(x) = 4x^3 - 108x$

The second derivative of the function $f''(x) = 12x^2 - 108$

Let us assume second derivative $f''(x) = 12x^2 - 108 =0$

Divide by 12 on both sides, we get $x^2 - 9= 0$

$\Rightarrow x^2 = 9$. Therefore, $x = \pm 3$

To check inflexion at x = 3, substitute two adjoining points x= 2 and 4 in $f''(x)$

So, $f''(2) = 12(2)^2 - 108 = -60$ (negative)

$f''(4) = 12(4)^2 - 108 = 84$ (positive)

To check for x = -3, substitute x= 0 and -4 in $f''(x)$

So, $f''(0) = 12(0)^2 - 108 = -108$ (negative)

$f''(3) = 12(-4)^2 - 108 = 84$ (positive)

Hence, proved

Now, substitute $x = \pm 3$ in $f''(x)$

Therefore, it becomes

$f''(3) = 12(3)^2 - 108 = 0$

$f''(-3) = 12(-3)^2 - 108 = 0$

Therefore, the inflection points are (3, 0), and (-3, 0).

Exercise 14: A toy manufacturing company manufactures and sells 3000 units per month at a price of ₹100. Reducing the price by ₹ 5 shall help the company to sell 300 additional units per month. Predict the maximum possible revenue of the company.

Solution: Suppose that p is number of deductions of ₹ 5 from the base price of ₹ 100.

Then the unit price is 100 − 5p

The total amount of products sold for the month is 3000 + 300p

Hence the total revenue is given by (R) = (100 − 5p). (3000 + 300p)

$=300,000 + 30,000p - 15,000p - 1500p^2$

$= -1500p^2 +15,000p + 300,000$

Assume f (R)= $-1500p^2 +15,000p + 300,000$

The first differentiation of the revenue function f '(R)= -3000p+15000

Put f ' (R)=0

$\Rightarrow 3000p=15000$

p=5, Now take the second derivative f ''(R)

f ''(R)=d (-3000p+15000)/dx = -3000 <0

So, p= 5 is a point of maximum as the second derivative is -ve.

The maximum possible revenue per month=New price* No of units sold

= (100-5). (3300) = ₹3,13,500.

Increase in revenue= (100-5).3300-100.3000

=3,13,500-300000=₹13,500.

1.9 Partial Differentiation

The partial derivative of a function is used to take one tangent line of a given function to know its slope, is called partial differentiation. If a function f (x, y) depends on two independent variables x, y and we find its derivative with respect to x, keeping y as a constant, and its derivative with respect to y, keeping x as a constant, is called partial differentiation. It is shown as follows:

The partial derivative with respect to x while keeping y constant is given as:

$$f_x = \frac{\partial f}{\partial x} = \lim_{k\to 0}\frac{f(x+k, y) - f(x, y)}{k}$$

The partial derivative with respect to y, keeping x as a constant, is given as:

$$f_{xy} = \frac{\partial f}{\partial y} = \lim_{k\to 0}\frac{f(x, y+k) - f(x, y)}{k}$$

Rules of partial differentiation: The following rules are applicable to partial differentiation:

1. Multiplication rule of partial differentiation: Let we have a function f=g(x,y). h(x,y) . The value of partial differentiation of the product of these two functions g and h is give below:

$$f_x = \frac{\partial f}{\partial x} = g(x, y)\frac{\partial h}{\partial x} + h(x, y)\frac{\partial g}{\partial x}$$

$$f_y = \frac{\partial f}{\partial y} = g(x, y)\frac{\partial h}{\partial y} + h(x, y)\frac{\partial g}{\partial y}$$

2. Division or quotient rule: Let us assume a function $f = \frac{g(x,y)}{h(x,y)}$. Its partial derivative with respect to x and y is given below:

$$f_x = \frac{h(x, y)\frac{\partial g}{\partial x} - g(x, y)\frac{\partial h}{\partial x}}{[h(x, y)]^2}$$

$$f_y = \frac{h(x, y)\frac{\partial g}{\partial y} - g(x, y)\frac{\partial h}{\partial y}}{[h(x, y)]^2}$$

3.Power Rule of partial differentiation: Let $f= [g(x,y)]^n$, its partial derivative with respect to x and y is given below: $f_x=n[g(x,y)]^{n-1}. \partial g/\partial x$

$$f_y=n[g(x,y)]^{n-1}. \partial g/\partial y$$

4.Chain rule of independent and dependent variables: Let us have x=j(t) and y=k(t) are two single independent functions differentiable with respect to t. Also, there is another differentiable function z=f(x,y), differentiable with respect to x, and y. Now, we can write z=f(j(t), k(t)), which is differentiable with respect to t. The partial derivative of z with respect to 't' is given below:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}.\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}.\frac{\partial y}{\partial t}$$

Let us have another functions x = j (u, v) and y = k (u, v) are two dependent differentiable functions and z = f (x, y) is a differentiable with respect to x and y. We can write z = f (j (u, v), k (u, v)). The partial derivative of z with respect to u and v is given below:

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}.\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}.\frac{\partial y}{\partial u}$$

$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}.\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}.\frac{\partial y}{\partial v}$$

The partial differentiation of logarithmic and trigonometric functions is done under the same procedure as discussed for normal differentiation.

Exercise 15: let $f(x, y) = 4x + 3y$, find its partial derivative with respect to x and y.

Solution: The partial derivative is calculated by differentiating the function with respect to one variable and keeping another variable as a constant.

$$f(x, y) = 4x + 3y$$

$$\frac{\partial f}{\partial x} = \frac{\partial(4x + 3y)}{\partial x} = \frac{\partial 4x}{\partial x} + \frac{\partial 3y}{\partial x} = 4 + 0 = 4$$

$$\frac{\partial f}{\partial y} = \frac{\partial(4x + 3y)}{\partial y} = \frac{\partial 4x}{\partial y} + \frac{\partial 3y}{\partial y} = 0 + 3 = 3$$

Exercise 16: Let $f(x, y) = x^3y + \sin x + \cos y$, find its partial derivative with respect to x and y.

Solution: We can differentiate this function with respect to one variable keeping all other variables constant

$$f_x = \partial f/\partial x = (3x)\, y + \cos x + 0$$

$$= 3xy + \cos x$$

Similarly, $$f_y = \partial f/\partial y = x^3 + 0 + (-\sin y)$$

$$= x^3 - \sin y$$

1.10 Suggested Readings

- K. Sydsaeter and P. Hammond (2002). Mathematics for Economic Analysis. Pearson Educational Asia: Delhi.
- Wainwright, Chiang (2013). Fundamental Methods of Mathematical Economics, Tata McGraw Hill, 2013.
- Archibald, C.C and Lipsey, R.G (1977). An Introduction to a Mathematical Treatment of Economics, 1977, English Language Book Society.
- Sanchati, D.C. and Kapoor, V.K (1993). Business Mathematics. New Delhi, Sultan Chand & Sons, 1993.
- Gupta, S.C.: Fundamentals of Statistics, Bombay, Himalaya Publishing House.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 2 – MATRICES

STRUCTURE

2.0 Objectives

2.1 Introduction

2.2 Definition of Matrix

2.3 Notations

2.4 Types of Matrices

2.5 Equality of Matrices

2.6 Sub-Matrix of a Matrix

2.7 Multiplication of Matrix by a Scalar

2.8 Addition of Matrices

2.9 Difference of Matrices

2.10 Properties of Matrix Addition

2.11 Multiplication of Matrices

2.12 Properties of Matrix Multiplication

2.13 Applications of Matrices to Business and Economic Problems

2.14 Transpose of a Matrix

2.15 Symmetric Matrix

2.16 Skew-Symmetric Matrix

2.17 Orthogonal Matrix

2.18 Sum Up

2.19 Key Terms

2.20 Questions for Practice

2.21 Further Readings

2.0 Objectives

After studying the Unit, students will be able to:

- Define the Meaning and types of a Matrix
- Apply Addition, Subtraction and Multiplication of Matrices
- Apply the concept of Matrices to Business and Economic Problems
- Distinguish between Single and Double Entry Systems
- Find out Transpose a Matrix
- Understand the meaning of Symmetric, Skew-Symmetric, and Orthogonal Matrices.

2.1 Introduction

Matrices are one of the most important and powerful business mathematics tools. Matrices applications help solve linear equations and with the help of this cost estimation, sales projection, etc. can be predicted. A matrix consists of a rectangular presentation of numbers arranged systematically in rows and columns describing the various aspects of a phenomenon inter-related in some manner.

For example: Marks obtained by two students, say, Ram and Shyam, in English, Mathematics, and Statistics are as follows:

	English	Mathematics	Statistics
Ram	60	80	85
Shyam	65	90	80

These marks may be represented by the following rectangular array enclosed by a pair of brackets

$$\begin{bmatrix} 60 & 80 & 85 \\ 65 & 90 & 80 \end{bmatrix} \begin{matrix} \leftarrow \textit{First Row} \\ \leftarrow \textit{Second Row} \end{matrix}$$

↑ First Column ↑ Second Column ↑ Third Column

.

Each horizontal line is called a row and each vertical line is called a column. The first row indicates the marks obtained by Ram in English, Mathematics and Statistics respectively and the second by Shyam in the three respective subjects.

Such a rectangular array is called a Matrix.

2.2 DEFINITION OF A MATRIX

An $m \times n$ matrix is a rectangular array of mn numbers (or elements) arranged in the form of an ordered set of m rows and n columns. A matrix A having m rows and n columns is typically written as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n}$$

The horizontal lines are called rows and the vertical lines columns.

The numbers $a11, a12, \cdots,$ and amn belonging to the matrix are called elements.

A matrix having m rows and n columns is said to be of *order* $m \times n$ (read as 'm' by 'n').

The order may be written on the right of the matrix, as shown above.

2.3 Notations

Matrices are denoted by capital letters such as $A, B, C, \ldots,$ X, Y, Z and their elements by small letters $a, b, c, \ldots, a11, a12,$ etc. There are different nations enclosing the elements constituting a matrix in common use, *viz*., [ ], ( ), and { }, but we shall use the first one throughout the chapter. The suffixes of the element aij depict that the elements lie in the i^{th} row and j^{th} column. Note that we always write the row number first and column number afterward. Also, note that, for the sake of brevity, the matrix A given above may also be written as $A = [aij]\ m \times n$.

2.4 Types of Matrices

A. Row Matrix: If a matrix has only one row, it is called a *row matrix*. Thus, any $1 \times n$ matrix is called a row matrix, for example,

$$A = [a1\ a2\ a3\ \ldots\ an]$$

is a row matrix of order $1 \times n$.

B. Column Matrix: A matrix consisting of only one column is called a columnatrix. In other words, any m × 1 matrix is called a column matrix, for example

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

is a column matrix of order m × 1.

C. Zero or Null Matrix (0): If every element of an m × n matrix is zero, the matrix is called a zero matrix of order m × n, and is denoted by $0_{m \times n}$ or 0_{mn} or 0 simply. For example,

$$0_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{2\times 3}$$

is a zero matrix of order 2 × 3.

D. Square Matrix: Any matrix in which the number of rows is equal to the number of columns is called a Square matrix. Thus any n × n matrix is a square matrix of order n. Generally, we denote the order of a square matrix by a single number n, rather than n × n.

Remark: The elements a_{ij} for which iI = j in $A = [a_{ij}]_{n \times n}$ are called the diadonal lements and the line along which the elements $a_{11}, a_{12}, \dots, a_{nn}$ lie is called the leading diagonal or principal diagonal or diagonal simply. In a square matrix the pair of elements a_{ij} and a_{ji} are said to be conjugate elements.

E. Diagonal Matrix: A square in which all elements except those in the leading diagonal are zero, is called a diagonal matrix. Thus, a diagonal matrix of order n will be:

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Sometimes a diagonal matrix of order n with diagonal elements $a_{11}, a_{22}, \dots, a_{nn}$ is denoted by $A = \text{diag}\,(a_{11}, a_{22}, \dots, a_{nn})$.

F. Scalar Matrix: A diagonal matrix whose diagonal elements are all equal is called a scalar matrix. For example,

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

is a scalar matrix.

G. Identity or Unit Matrix (I): A scalar matrix in which each of its diagonal elements is unity is called an identity or unit matrix.

Thus, a square matrix $A = [a_{ij}]_{n \times n}$ is called identity matrix, if

$$a_{ij} = \begin{cases} 1, & \text{when } i = j, \\ 0, & \text{when } i \neq j. \end{cases}$$

An identity matrix of order n is denoted by I_n. Thus, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a unit matrix of order 2.

H.Triangular Matrix: If every element above or below the leading diagonal is zero, the matrix is called a triangular matrix. If the zero element is lie below the leading diagonal, the matrix is called upper triangular matrix; If the zero elements is lie above the leading diagonal, the matrix is called lower triangular matrix. The matrices A_1 and A_2 given below are the examples of upper and lower triangular matrices respectively:

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} \text{(Upper triangular matrix)}$$

$$A_2 = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \text{(Lower triangular matrix)}$$

2.5 Equality Of Matrices

Two matrices are called comparable, if each of them consists of as many rows and columns as the other. Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be equal, if (i) they have the same order and (ii) have equal corresponding elements throughout ($a_{ij} = b_{ij}$ for every i and j). The equality of matrices A and B.

Thus, the matrices $\begin{bmatrix} 1 & 7 \\ 3 & 5 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 9 \end{bmatrix}$ are not comparable while $\begin{bmatrix} 1 & 7 & 8 \\ 3 & 5 & 4 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 6 \\ 3 & 2 & 9 \end{bmatrix}$ are comparable but not equal.

The matrices $\begin{bmatrix} 1 & 5 & 9 \\ 3 & 4 & 12 \end{bmatrix}$ and $\begin{bmatrix} 1 & 5 & 3 \times 3 \\ 3 & 2 \times 2 & 3 \times 4 \end{bmatrix}$ are equal.

2.6 Sub Matrix Of A Matrix

A matrix which is obtained from a given matrix by deleting any number of rows or columns is called a sub-matrix of the given matrix.

For example, the matrix $\begin{bmatrix} 2 & 3 \\ 9 & 3 \end{bmatrix}$ is a sub-matrix of $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 9 & 3 \end{bmatrix}$. Obtained by deleting 2nd row and 1st column.

2.7 Multiplication Of A Matrix By A Scalar

Let $A = [a_{ij}]$ be an $m \times n$ matrix and let k be any real number (called scalar). Then the product of k and A denoted by kA is defined to be the $m \times n$ matrix (i, j)th element is ka_{ij}, i.e.,

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}$$

Thus, if $$A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}, 3A = \begin{bmatrix} 6 & 9 \\ 12 & 15 \end{bmatrix}$$

Thus, we notice that, to get the scalar product each element of the given matrix is multiplied by the given scalar.

Properties of Scalar Multiplication:

(i) The product of a matrix with a scalar is commutative, i.e., $kA = Ak$.

(ii) If $k = -1, (-1)A = [-a_{ij}]$. Generally, $(-1)A$ is denoted by $-A$ and is called the negative of matrix A. Thus, $-[a_{ij}] = [-a_{ij}]$.

(iii) If A and B are comparable matrices and k is any scalar, we have $k(A + B) = kA + kB$.

(iv) If k and l are any two scalars and A is any matrix, we have $(k + l)A = kA + lA$.

(v) If k and l are any two scalars, we have $k(lA) = (kl)A$.

Example 1. Read the following elements $a_{21}, a_{32}, a_{22}, a_{11}$ in

$$A = \begin{bmatrix} 4 & 2 & 3 \\ 9 & 0 & 4 \\ 8 & 7 & 9 \end{bmatrix}$$

Solution: a_{21} denotes element in second row and first column,

$a_{21} = 9, a_{32} = 7, a_{22} = 0, a_{11} = 4.$

Example 2. Construct a 2×3 matrix $A = [a_{ij}]_{2\times3}$ whose general element is giving by

$$a_{ij} = (i - j)^2/_2$$

Solution: $a_{11} = \frac{(1-1)^2}{2} = 0,\ a_{12} = \frac{(1-2)^2}{2} = \frac{1}{2},\ a_{13} = \frac{(2-1)^2}{2} = \frac{4}{2} = 2$

$a_{21} = \frac{(2-1)^2}{2} = \frac{1}{2},\ a_{22} = \frac{(2-2)^2}{2} = 0,\ a_{23} = \frac{(2-3)^2}{2} = \frac{1}{2}$

$$\therefore \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}_{2\times 3}$$

$$\Rightarrow \quad A = \begin{bmatrix} 0 & \frac{1}{2} & 2 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}_{2\times 3}$$

Example 3. If $A = \begin{bmatrix} 2 & 3 & 5 \\ 3 & -6 & 2 \end{bmatrix}$, find $-3A$.

Solution:
$$-3A = \begin{bmatrix} -3\times 2 & -3\times 3 & -3\times 5 \\ -3\times 3 & -3\times -6 & -3\times 2 \end{bmatrix} = \begin{bmatrix} -6 & -9 & -15 \\ -9 & 18 & -6 \end{bmatrix}$$

2.8 Addition of Matrices

The matrices A and B are conformable for addition, if they are comparable. i.e., B has the same number of rows and the same number of columns as A. Their sum, denoted by A + B, is defined to be the matrix obtained by adding the corresponding elements of A and B.

For example, if $A = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 3 & 7 \end{bmatrix}$

and $B = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 1 & 2 \end{bmatrix}$, we have

$$A + B = \begin{bmatrix} 2+1 & 3+3 & 0+5 \\ 3+3 & 3+1 & 7+2 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 3 & 6 & 5 \\ 6 & 4 & 9 \end{bmatrix}$$

In general, if $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$, which is a matrix obtained by adding the elements in the corresponding positions. Thus, from

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m\times n} \quad \text{and}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}_{m\times n}$$

We get
$$A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}_{m\times n}$$

2.9 DIFFERENCE OF MATRICES

If A and B are two comparable matrices, then their difference A − B is matrix whose elements are obtained by subtracting the elements of B from the corresponding elements of A.

If $A = [a_{ij}]_{m \times n}$ and $B = [b_{ij}]_{m \times n}$

Then $A - B = [a_{ij} - b_{ij}]_{m \times n}$.

2.10 PROPERTIES OF MATRIX ADDITION

Suppose A, B, C are three matrices of the same order $m \times n$. Then the matrix addition has following properties:

A. Associativity

$$A + (B + C) = (A + B) + C$$

i.e., the addition of matrices is associative.

B. Commutativity

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}]$$

$$= [b_{ij} + a_{ij}] = [b_{ij}] + [a_{ij}]$$

$$= B + A$$

i.e., matrix addition is commutative.

C. Distributive law

$m(A + B) = mA + mB$ (m being an arbitrary scalar), because

$$m(A + B) = m[a_{ij} + b_{ij}]$$

$$= [m\,a_{ij} + m\,b_{ij}]$$

$$= m[a_{ij}] + m[b_{ij}] = mA + mB.$$

D. Existence of additive identity

Let A be any $m \times n$ matrix, and 0 the $m \times n$ null matrix. Then, we have

$$A + 0 = 0 + A = A$$

i.e., the null matrix is the identity for the matrix addition.

E. Existence of additive inverse

−A is the additive inverse of A, because

$$(-A) + A = [-a_{ij}] + [a_{ij}]$$

$$= [-a_{ij} + a_{ij}] = 0 = A + (-A)$$

Thus, for any matrix A, there exists a unique additive inverse −A.

F. Cancellation law

$$A + B = A + C \Rightarrow [a_{ij} + b_{ij}] = [a_{ij} + c_{ij}]$$

$$\Rightarrow a_{ij} + b_{ij} = a_{ij} + c_{ij}$$

$$\Rightarrow b_{ij} = c_{ij}$$

$$\Rightarrow [b_{ij}] = [c_{ij}]$$

$$\Rightarrow B = C$$

This is said to be left cancellation.

Similarly, right cancellation, namely,

$B + A = C + A \Rightarrow B = C$ can be proved.

Example 4. If $A = \begin{bmatrix} 1 & 5 & 6 \\ -6 & 7 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -5 & 7 \\ 8 & -7 & 7 \end{bmatrix}$, find A + B and A − B.

Solution: Here $A + B = \begin{bmatrix} 1 & 5 & 6 \\ -6 & 7 & 0 \end{bmatrix} + \begin{bmatrix} 1 & -5 & 7 \\ 8 & -7 & 7 \end{bmatrix}$

$$= \begin{bmatrix} 1+1 & 5-5 & 6+7 \\ -6+8 & 7-7 & 0+7 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 13 \\ 2 & 0 & 7 \end{bmatrix}$$

and $A - B = \begin{bmatrix} 1 & 5 & 6 \\ -6 & 7 & 0 \end{bmatrix} - \begin{bmatrix} 1 & -5 & 7 \\ 8 & -7 & 7 \end{bmatrix}$

$$= \begin{bmatrix} 1-1 & 5-(-5) & 6-7 \\ -6-8 & 7-(-7) & 0-7 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 10 & -1 \\ -14 & 14 & -7 \end{bmatrix}$$

Example 5. If $A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & -1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & -6 \\ 0 & -1 & 3 \end{bmatrix}$, evaluate 3A − 4B.

Solution: $3A - 4B = 3\begin{bmatrix} 2 & 3 & 1 \\ 0 & -1 & 5 \end{bmatrix} - 4\begin{bmatrix} 1 & 2 & -6 \\ 0 & -1 & 3 \end{bmatrix}$

$$= \begin{bmatrix} 6 & 9 & 3 \\ 0 & -3 & 15 \end{bmatrix} - \begin{bmatrix} 4 & 8 & -24 \\ 0 & -4 & 12 \end{bmatrix}$$

$$= \begin{bmatrix} 6-4 & 9-8 & 3-(-24) \\ 0-0 & -3-(-4) & 15-12 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 27 \\ 0 & 1 & 3 \end{bmatrix}.$$

Example 6. If X, Y are two matrices given by the equations

$$\mathbf{X + Y} = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \text{ and } \mathbf{X - Y} = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}, \text{ find } \mathbf{X, Y}.$$

Solution: We have

$$X + Y = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}$$

...(i)

$$X - Y = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$$

...(ii)

By adding equations (i) and (ii),

$$2X = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1+3 & -2+2 \\ 3-1 & 4+0 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 2 & 4 \end{bmatrix},$$

$$\therefore \quad X = \frac{1}{2}\begin{bmatrix} 4 & 0 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}.$$

From equation (i),

$$Y = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} - X = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1-2 & -2-0 \\ 3-1 & 4-2 \end{bmatrix} = \begin{bmatrix} -1 & -2 \\ 2 & 2 \end{bmatrix}.$$

CHECK YOUR PROGRESS (A)

1. Find the elements $a_{31}, a_{24}, a_{34}, a_{22}$ in each of the following matrices given below. Also give their diagonal elements.

$$A = \begin{bmatrix} 8 & 6 & -3 & 1 \\ 2 & 3 & 5 & 0 \\ 3 & 9 & 5 & -7 \\ 5 & -3 & 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 7 & 3 & 5 \\ 2 & 3 & -1 & 0 \\ 3 & 5 & 6 & 8 \\ 4 & 3 & 0 & 0 \\ 2 & 1 & 9 & 8 \end{bmatrix}$$

2. Write the matrix $A = [a_{ij}]$ of order 2×3 whose general elements is given by (i) $a_{ij} = ij$ (ii) $a_{ij} = (-1)^{ij}(i + j)$

3. Find x and y, if

$$\begin{bmatrix} x+y & z \\ 1 & x-y \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 1 & 7 \end{bmatrix}$$

4. If $A = \begin{bmatrix} 0 & 2 & 3 \\ 2 & 1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 6 & 3 \\ 1 & 4 & 5 \end{bmatrix}$, find the value of $2A + 3B$.

5. Given $A = \begin{bmatrix} 1 & 2 & -3 \\ 5 & 0 & 2 \\ 1 & -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -1 & 2 \\ 4 & 2 & 5 \\ 2 & 0 & 5 \end{bmatrix}$, find the matrix C such that $A + 2C = B$.

6. Solve the following equations for A and B:

$$2A - B = \begin{bmatrix} 3 & -3 & 0 \\ 3 & 3 & 2 \end{bmatrix}, 2B + A\begin{bmatrix} 4 & 1 & 5 \\ -1 & 4 & -4 \end{bmatrix}$$

ANSWERS

1. 3, 0, -7, 3 and 3, 0, 8, 3; the diagonal elements are 8, 3, 8, 0 and 1, 3, 6, 0.

2. (i) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$ (ii) $\begin{bmatrix} -2 & 3 & -4 \\ 3 & 4 & 5 \end{bmatrix}$

3. $x = 5, y = -2$

4. $\begin{bmatrix} 21 & 22 & 15 \\ 7 & 14 & 23 \end{bmatrix}$

5. $\begin{bmatrix} 1 & -\frac{3}{2} & \frac{5}{2} \\ -\frac{1}{2} & 1 & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} & -2 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & -2 \end{bmatrix}$

2.11 Multiplication of Matrix

The multiplication of one matrix by another is possible, if and only if the number of columns of first matrix is equal to the number of rows of the second matrix. The resulting matrix will have the number of rows equal to those in the first matrix and the number of columns equal to those in the second matrix. Thus, if matrix A is of order $m \times n$ and B is of order $n \times p$, the product AB is possible, i. e., the matrices A and B are conformable for multiplication in the order A, B. The order of the resulting matrix AB will be $m \times p$. The (i, k)th element (i. e., the element laying ith row and kth column) of AB is given by

$$a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

Thus, to obtain the (i, k)th element of the product AB, we multiply the elements of the ith row of A to the corresponding elements of the kth column of B and add the products thus obtained. The resulting sum is the (i, k)th element of AB.

If AB is denoted by $C = [c_{ij}]$, i. e., $AB = [c_{ij}]$, we have

$$\begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1k} & \cdots & b_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{j1} & \cdots & b_{jk} & \cdots & b_{jp} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & \cdots & b_{nk} & \cdots & b_{np} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & \cdots & c_{1k} & \cdots & c_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{i1} & \cdots & c_{ik} & \cdots & c_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{m1} & \cdots & c_{mk} & \cdots & c_{mp} \end{bmatrix}$$

Where $c_{ij} = \sum_{j=1}^{n} a_{ij} b_{jk}$. Putting $i = 1, 2, \cdots, m$ and $k = 1, 2, 3, \ldots, p$, all the elements of C will be found.

In the product AB, A is said to be pre-multiplier or pre-factor while B is said to be post-multiplier or post-factor. It is to be noted that in multiplying one matrix by another, unlike ordinary numbers, the placement of matrices as pre factor and post factor is very important. Thus AB is not the same as BA.

2×2 Matrix Multiplication Formula

Let us consider two matrices A and B of order "2 × 2". Then its multiplication is achieved using the formula.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} andB = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

$$AB = \begin{bmatrix} (ap + br) & (aq + bs) \\ (cp + dr) & (cq + ds) \end{bmatrix}$$

3x3 Matrix Multiplication Formula

Let us consider two matrices P and Q of order "3 × 3". Now, the matrix multiplication formula of "3 × 3" matrices is,

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix}$$

$$XY = \begin{bmatrix} (x_{11}y_{11} + x_{12}y_{21} + x_{13}y_{31}) & (x_{11}y_{12} + x_{12}y_{22} + x_{13}y_{32}) & (x_{11}y_{13} + x_{12}y_{23} + x_{13}y_{33}) \\ (x_{21}y_{11} + x_{22}y_{21} + x_{23}y_{31}) & (x_{21}y_{12} + x_{22}y_{22} + x_{23}y_{32}) & (x_{21}y_{13} + x_{22}y_{23} + x_{23}y_{33}) \\ (x_{31}y_{11} + x_{32}y_{21} + x_{33}y_{31}) & (x_{31}y_{12} + x_{32}y_{22} + x_{33}y_{32}) & (x_{31}y_{13} + x_{32}y_{23} + x_{33}y_{33}) \end{bmatrix}$$

2.12 Properties Of Matrix Multiplication

1. Matrix multiplication is not commutative. To verify the above statement, let us take an example. Consider the matrices.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

It can easily find that

$$AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \text{ while } BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so that $AB \neq BA$.

This shows that matrix multiplication is not commutative. Actually speaking, for a given pair of matrices A and B, the products AB and BA may not be even comparable. For example, if A is an $m \times n$ matrix and B is an $n \times m$ matrix, AB would be an $m \times m$ matrix and BA would be an $n \times n$ matrix.

It may also happen that for a pair of matrices A and B the product AB may be defined but the product BA may not be defined. For example, it A is an $m \times n$ matrix and B is an $n \times p$ matrix, AB would be an $m \times p$ matrix, but it is not meaningful to talk of BA unless $m = p$.

Note 1. It is worthwhile to note that the statement 'matrix multiplication is not commutative' does not mean that there are no matrices A and B such that $AB = BA$. It simply means that generally $AB \neq BA$. Thus, we wish to convey that there do exist some pairs of matrices A and B for which $AB = BA$.

Note 2. It is also to be noted that in matrices, $AB = 0$ need not always imply that either $A = 0$ or $B = 0$. This will be clear, if we consider the matrices,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}.$$

For these matrices, $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, but none of A and B is zero matrix.

Note 3. The familiar cancellation law of multiplication for numbers fails to be true for matrix multiplication.

Below we give the properties which hold good for matrices.

2. **Associative law:** Let A, B and C be the matrices of suitable size for the products A(BC) and (AB)C to exist. Then, $A(BC) = (AB)C$.
3. **Distributive law:** $A(B+C) = AB + AC$, (left distributive) and $(B+C)D = BD + CD$ (right distributive), provided that the matrices A, B, C and D are of the sizes that they are conformable for the operations involved so that the above relations are meaningful.
4. **Multiplication of a matrix by a unit matrix:** If A is square matrix of order $n \times n$ and I is the unit matrix of the same order, we get

   $AI = A = IA$.
5. **Multiplication of a matrix by itself:** The product A, A is defined, if the number of column is equal to the number of rows of A, i.e., if A is a square matrix and in that case A. $A = A^2$ will also be a square matrix of the same type. Also,

   $A.A.A = A^2A = A^3$.

Similarly, $A.A.A \dots n$ times$= A^n$.

Note: If I is a unit matrix, we have $I = I^2 = I^3 = \cdots = I^n$.

Example 7. If $A = \begin{bmatrix} 2 & 2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix}$

and $B = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}$. Obtain the product AB and BA and show than $AB \neq BA$.

Solution:

$$AB = \begin{bmatrix} 2 & 2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix}_{3\times3} \begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}_{3\times3}$$

Number of columns of A = Number of rows of B

$\therefore$ AB is defined

$$AB = \begin{bmatrix} 0+0+3 & 0+2+6 & 4+4+0 \\ 0+0-1 & 0+3-2 & 4+6+0 \\ 0+0+2 & 0+1+4 & -6+2+0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 8 & 8 \\ -1 & 1 & 10 \\ 2 & 5 & -4 \end{bmatrix}$$

$$BA = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}_{3\times3} \begin{bmatrix} 2 & 2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix}_{3\times3}$$

$$= \begin{bmatrix} 0+0-6 & 0+0+2 & 0+0+4 \\ 0+2-6 & 0+3+2 & 0-1+4 \\ 2+4+0 & 2+6+0 & 3-2+0 \end{bmatrix}_{3\times3}$$

$$= \begin{bmatrix} -6 & 2 & 4 \\ -4 & 5 & 3 \\ 6 & 8 & 1 \end{bmatrix}$$

Hence $AB \neq BA$.

Example 8. Write down the products AB and BA of the two matrices A and B, where,

$$\mathbf{A} = [\mathbf{1} \quad \mathbf{2} \quad \mathbf{3} \quad \mathbf{4}] \text{ and } \mathbf{B} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{bmatrix}$$

Solution: Since **A** is a 1×4 matrix and **B** is a 4×1 matrix. **AB** will be a 1×1 matrix.

$$\mathbf{AB} = [1 \quad 2 \quad 3 \quad 4] \times \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$= [1.1 + 2.2 + 3.3 + 4.4] = [30]$$

BA will be a 4×4 matrix.

$$\mathbf{BA} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1\times1 & 2\times1 & 3\times1 & 4\times1 \\ 1\times2 & 2\times2 & 3\times2 & 4\times2 \\ 1\times3 & 2\times3 & 3\times3 & 4\times3 \\ 1\times4 & 2\times4 & 3\times4 & 4\times4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix}$$

Example 9. Evaluate $\mathbf{A}^2 - 4\mathbf{A} - 5\mathbf{I}$, where

$$\mathbf{A} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{2} \\ \mathbf{2} & \mathbf{1} & \mathbf{2} \\ \mathbf{2} & \mathbf{2} & \mathbf{1} \end{bmatrix} \text{ and } \mathbf{I} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Solution: We have

$$A^2 = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+4+4 & 2+2+4 & 2+4+2 \\ 2+2+4 & 4+1+4 & 4+2+2 \\ 2+4+2 & 4+2+2 & 4+4+1 \end{bmatrix} = \begin{bmatrix} 9 & 8 & 8 \\ 8 & 9 & 8 \\ 8 & 8 & 9 \end{bmatrix}$$

$$A^2 - 4A - 5I = \begin{bmatrix} 9 & 8 & 8 \\ 8 & 9 & 8 \\ 8 & 8 & 9 \end{bmatrix} - 4\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} - 5\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 9-4-5 & 8-8+0 & 8-8+0 \\ 8-8+0 & 9-4-5 & 8-8+0 \\ 8-8+0 & 8-8+0 & 9-4-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0.$$

Where 0 is the null matrix.

Example 10. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, show that

$$\mathbf{A^2 - (a+d)A = (bc - ad)I.}$$

Solution: $\mathbf{A^2 = A.A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$= \begin{bmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix}$$

$$A^2 - (a+d)A = \begin{bmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix} - (a+d)\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{bmatrix} - \begin{bmatrix} a^2 + ad & ab + bd \\ ac + cd & ad + d^2 \end{bmatrix}$$

$$= \begin{bmatrix} bc - ad & 0 \\ 0 & bc - ad \end{bmatrix} = (bc - ad)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= (bc - ad)\mathbf{I}.$$

Example 11. If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 2 & -3 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, show that $A(B + C) = AB + AC$.

Solution: We have

$$B + C = \begin{bmatrix} 1 & 0 \\ 2 & -3 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+1 & 0-1 \\ 2+0 & -3+1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 2 & -2 \end{bmatrix}$$

$$A(B + C) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 2 & -1 \\ 2 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 2+4 & -1-4 \\ 6+8 & -3-8 \end{bmatrix} = \begin{bmatrix} 6 & -5 \\ 14 & -11 \end{bmatrix}.$$

...(i)

Again, $$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 2 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1+4 & 0-6 \\ 3+8 & 0-12 \end{bmatrix} = \begin{bmatrix} 5 & -6 \\ 11 & -12 \end{bmatrix}$$

and $$AC = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+0 & -1+2 \\ 3+0 & -3+4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix},$$

∴ $$AB + AC = \begin{bmatrix} 5 & -6 \\ 11 & -12 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -5 \\ 14 & -11 \end{bmatrix}$$

...(ii)

From (i) and (ii), we have

$$A(B + C) = AB + AC.$$

Example 12. $A = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix}$, $B = \begin{bmatrix} a & 1 \\ b & -1 \end{bmatrix}$ **and** $(A + B)^2 = A^2 + B^2$ **find a and b.**

Solution: We have

$$A^2 = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} \times \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 1-2 & -1+1 \\ 2-2 & -2+1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

and $$B^2 = \begin{bmatrix} a & 1 \\ b & -1 \end{bmatrix} \times \begin{bmatrix} a & 1 \\ b & -1 \end{bmatrix} = \begin{bmatrix} a^2 + b & a - 1 \\ ab - b & b + 1 \end{bmatrix}$$

∴ $$A^2 + B^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} a^2 + b & a - 1 \\ ab - b & b + 1 \end{bmatrix} = \begin{bmatrix} a^2 + b - 1 & a - 1 \\ ab - b & b \end{bmatrix}$$

...(1)

Also $$A + B = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} + \begin{bmatrix} a & 1 \\ b & -1 \end{bmatrix} = \begin{bmatrix} 1+a & -1+1 \\ 2+b & -1-1 \end{bmatrix} = \begin{bmatrix} 1+a & 0 \\ 2+b & -2 \end{bmatrix}$$

∴ $$(A + B)^2 = \begin{bmatrix} 1+a & 0 \\ 2+b & -2 \end{bmatrix} \times \begin{bmatrix} 1+a & 0 \\ 2+b & -2 \end{bmatrix} =$$

$$\begin{bmatrix} (1+a)^2 + 0 & 0+0 \\ (2+b)(1+a) - 2(2+b) & 0+4 \end{bmatrix}$$

$$= \begin{bmatrix} (1+a)^2 & 0 \\ (2+b)(a-1) & 4 \end{bmatrix}.$$

...(2)

Now given that $(A+B)^2 = A^2 + B^2$.

Hence from (1) and (2), we get

$$\begin{bmatrix} (1+a)^2 & 0 \\ (2+b)(a-1) & 4 \end{bmatrix} = \begin{bmatrix} a^2+b-1 & a-1 \\ ab-b & b \end{bmatrix}$$

or $a-1=0$ and $b=4$.

Hence $a=1, b=4$.

2.13 Applications Of Matrices To Business And Economic Problems

Matrices play a crucial role in various business and economic applications due to their ability to represent and solve complex systems of equations, model relationships between variables, and analyze data. Here are some key applications of matrices in business and economics:

1. Linear Programming Matrices are extensively used in linear programming, a mathematical technique for optimizing resource allocation. Businesses can use matrices to represent constraints, objective functions, and decision variables, allowing them to find the optimal solution for maximizing profits or minimizing costs.

2. Input-Output Analysis: Input-output models in economics use matrices to represent the relationships between different sectors of an economy. These models help analyze the interdependencies between industries, consumption, and production, aiding in policy decisions and economic planning.

3. predicting future economic conditions: Matrices is useful in predicting future economic conditions, such as consumer behavior, investment decisions, and market trends.

4. Financial Analysis: Matrices are used in finance to model and analyze investment portfolios. The risk and return associated with different assets can be represented in matrix form, facilitating the calculation of portfolio diversification and optimization.

6. Supply Chain Management: Matrices can represent the flow of goods, information, and resources within a supply chain. This allows businesses to optimize logistics, minimize costs, and enhance efficiency in the production and distribution processes.

7. Network Analysis: Matrices are used to model and analyze networks, such as social networks or transportation networks. This can be applied to understand the flow of information, goods, or services within a network, aiding in decision-making for resource allocation and optimization.

8. Time Series Analysis: Matrices are used to represent and analyze time series data in economics. Techniques like matrix decomposition can help identify trends, seasonal patterns, and underlying structures in economic time series

9. Risk Management: Matrices are used to model and analyze risk in financial and business contexts. For example, covariance matrices can be employed to assess the risk associated with different assets in a portfolio.

These applications highlight the versatility of matrices in addressing various quantitative aspects of business and economic problems, making them a valuable tool for decision-makers and analysts.

Example 14. A manufacturer produces three items P, Q and R and sells them in two markets I and II. Annual sales are given below:

	P	Q	R
I	6,000	2,000	3,000
II	8,000	4,000	2,000

If sales price of each unit of P, Q, R is Rs. 4, Rs. 3, Rs. 2 respectively, then find the total revenue of each market using matrix.

Solution: Let $A = \begin{matrix} I \\ II \end{matrix}\begin{bmatrix} 6,000 & 2,000 & 3,000 \\ 8,000 & 4,000 & 2,000 \end{bmatrix}$ is the sales matrix and

$B = \begin{matrix} P \\ Q \\ R \end{matrix}\begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$ is the price matrix

$\therefore$ Revenue matrix $AB = \begin{bmatrix} 6,000 & 2,000 & 3,000 \\ 8,000 & 4,000 & 2,000 \end{bmatrix}\begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$

$$= \begin{bmatrix} 6,000 \times 4 & + & 2,000 \times 3 & + & 3,000 \times 2 \\ 8,000 \times 4 & + & 4,000 \times 3 & + & 2,000 \times 2 \end{bmatrix} = \begin{bmatrix} 36,000 \\ 48,000 \end{bmatrix}$$

Hence

Total Revenue of Market I = Rs. 36,000

Total Revenue of Market II = Rs. 48,000

CHECK YOUR PROGRESS (B)

1. If $\mathbf{A} = \begin{bmatrix} 3 & 6 & -5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 4 \\ 7 \\ 2 \end{bmatrix}$, find **AB** and **BA.**

2. If $\mathbf{A} = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & -1 \\ -3 & 2 \end{bmatrix}$, find **AB** and **BA.** Is = **BA** ?

3. If $\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ -4 & 2 & 5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 2 & 1 \end{bmatrix}$, find **AB** and show that **AB = BA.**

4. When $\mathbf{A} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} i & -1 \\ -1 & -i \end{bmatrix}$ and $i = \sqrt{(-1)}$, determine **AB**. Also compute **BA**.

5. If $\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & -3 & 4 \\ 3 & -2 & 3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -1 & -2 & -1 \\ 6 & 12 & 6 \\ 5 & 10 & 5 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} -1 & -1 & 1 \\ 2 & 2 & -1 \\ -3 & 3 & 3 \end{bmatrix}$,

Show that **AB** and **CA** are the null matrices but **BA** and **AC** are not the null matrices.

6. If $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} a & b \\ 3 & 5 \end{bmatrix}$.

Find a and b such that **AB = BA.** Also compute **3A + 5B.**

7. If $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} 5 & 1 \\ 7 & 4 \end{bmatrix}$, show that $\mathbf{A(B + C) = AB + AC}$.

8. If $\mathbf{A} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$, prove by mathematical induction that $\mathbf{A}^n = \begin{bmatrix} 1 & na \\ 0 & 1 \end{bmatrix}$

9. Given $\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 3 \\ 3 & -1 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 3 \\ 0 & 2 \\ -1 & 4 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} 1 & 2 & 3 & -4 \\ 2 & 0 & -2 & 1 \end{bmatrix}$.

Show that $\mathbf{(AB)C = A(BC)}$.

10. If $A = \begin{bmatrix} -5 & -2 & 1 \\ 4 & 3 & 3 \\ -6 & 6 & -2 \end{bmatrix}$

Find the matrix B such that A + B = unit matrix.

11. If $\begin{bmatrix} 1 & 1 & x \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0$ **find x**

12. There are two families A and B. In family A, there are 4 men, 6 women and 2 children; and in family B there are 2 men, 2 women and 4 children. The recommended daily requirement for calories is:

Calories:	Man 2,400;	Woman 1,900;	Child 1,800
Protein:	Man 55 gm;	Woman 45 gm;	Child 33 gm

Calculate the total requirements of calories and proteins for each of the two families using matrix method.

13.The co-operative store of a particular school has 10 dozen books of physics, 8 dozen of chemistry books and 5 dozen of mathematics books. Their selling price are Rs. 65.70, Rs. 43.20 and Rs. 76.50 respectively. Find by matrix method the total amount, the store will receive from selling all three items.

Answers

1. $AB = [44], BA = \begin{bmatrix} 12 & 24 & -20 \\ 21 & 42 & -35 \\ 6 & 12 & -10 \end{bmatrix}$, **2.** $AB = \begin{bmatrix} -13 & 8 \\ -8 & 5 \end{bmatrix}, BA = \begin{bmatrix} 1 & 2 \\ -4 & -9 \end{bmatrix}$,

3. $AB = \begin{bmatrix} 0 & -4 \\ 10 & 3 \end{bmatrix}$ **4.** $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, BA = \begin{bmatrix} 2i & -2 \\ -2 & -2i \end{bmatrix}$,

6. $a = 65, b = 15, \begin{bmatrix} 41.5 & 13.5 \\ 27 & 28 \end{bmatrix}$ **10.** $\begin{bmatrix} 6 & 2 & -1 \\ -4 & -2 & -3 \\ 6 & -6 & -1 \end{bmatrix}$

11. -2

12. Calories for family A and B are 24,600 and 15,800 and proteins are 556 gms and 332 gms respectively.

13. Rs. 16,621.20

2.14 Transpose Of A Matrix

A matrix obtained by interchanging the corresponding rows and columns of a given matrix A is called the **transpose matrix** of **A.** The transpose of a matrix is denoted by A^T or A'.

For example,

If $A = [1 \quad 5]$, then $A' = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

If $A = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$, then $A' = [2 \quad -3]$

If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $A' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$

Properties of the transpose matrix

(i) $(A')' = A$

(ii) $(A + B)' = A' + B'$

(iii) $(\lambda A)' = \lambda A'$

(iv) If A and B are two matrices which are conformable for multiplication, then

$$(AB)' = B'A'$$

This is called 'Reversal Law'.

Example 15: If $A = \begin{bmatrix} 1 & 2 & -3 \\ 3 & 0 & 2 \\ 4 & 5 & 0 \end{bmatrix}$

and $B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix}$

Then, verify that $(AB)^T = B^T A^T$

Solution: $A = \begin{bmatrix} 1 & 2 & -3 \\ 3 & 0 & 2 \\ 4 & 5 & 0 \end{bmatrix}_{3\times3}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix}_{3\times3}$

$$AB = \begin{bmatrix} 1 & 2 & -3 \\ 3 & 0 & 2 \\ 4 & 5 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} (1+4+0) & (0+2-3) & (0+0-9) \\ (3+0+0) & (0+0+2) & (0+0+6) \\ (4+10+0) & (0+5+0) & (0+0+0) \end{bmatrix}$$

$$= \begin{bmatrix} 5 & -1 & -9 \\ 3 & 2 & 6 \\ 14 & 5 & 0 \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} 5 & 3 & 14 \\ -1 & 2 & 5 \\ -9 & 6 & 0 \end{bmatrix}$$

…(i)

$$A^T = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 0 & 5 \\ -3 & 2 & 0 \end{bmatrix}_{3\times3}, B^T = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix}_{3\times3}$$

$$B^T A^T = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 3 & 4 \\ 2 & 0 & 5 \\ -3 & 2 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} (1+4-0) & (3+0+0) & (4+10+0) \\ (0+2-3) & (0+0+2) & (0+5+0) \\ (0+0-9) & (0+0+6) & (0+0+0) \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 3 & 14 \\ -1 & 2 & 5 \\ -9 & 6 & 0 \end{bmatrix}$$

…(ii)

From (i) and (ii),

$$\mathbf{(AB)^T = B^T A^T}$$

2.15 Symmetric Matrix

A square matrix A such that $A' = A$ is called symmetric matrix, i. e., matrix $[a_{ij}]$ is symmetric provided $a_{ij} = a_{ji}$ for all values of i and j.

For example:

$$A = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \text{ is a symmetric matrix.}$$

2.16 Skew-Symmetric Matrix

A square matrix A such that $A' = -A$ is called skew-symmetric matrix, i. e., matrix $[a_{ij}]$ is skew-symmetric provided $a_{ij} = -a_{ji}$ for all values of i and j.

For example:

$$A = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & 5 \\ 3 & -5 & 0 \end{bmatrix} \text{ is a skew-symmetric matrix.}$$

Remarks. In a skew-symmetric matrix, we have $a_{ij} = -a_{ji}$. For diagonal elements

$a_{ii} = -a_{ii}$, i. e., $2a_{ii} = 0$, or $a_{ii} = 0$.

Thus, every diagonal element of a skew-symmetric matrix is zero.

Example 16. Show that every matrix can be unequally expressed as the sum of a symmetric and a skew-symmetric matrix.

Solution: Let A be any square matrix.

Now, we have $A = \frac{1}{2}A + \frac{1}{2}A = \frac{1}{2}(A + A') + \frac{1}{2}(A - A')$

...(i)

Now, $(A + A')' = A' + (A')' = A' + A = A + A'$

(Since matrix addition is commutative)

and $(A - A')' = A' - (A')' = A' - A = -(A - A')$

Hence $(A + A')$ is symmetric and $(A - A')$ is skew-symmetric.

Consequently $\frac{1}{2}(A + A') = P$ (say) is a symmetric matrix and $\frac{1}{2}(A - A') = Q$ (say) is a skew-symmetric matrix.

Hence, $A = P + Q$.

Thus, any square matrix, can be expressed as the sum of symmetric and skew-symmetric matrix.

Uniqueness. To show that this representation is unique, let us suppose that another representation $A = R + S$ is possible, where R is symmetric and S is skew-symmetric, i.e. $R = R'$ and $= S'$.

Now $\quad A' = (R + S)' = R' + S' = R - S$

Also, $\quad A + A' = (R + S) + (R - S) = 2R$

and $\quad A - A' = (R + S) - (R - S) = 2S$

or $\quad R = \frac{1}{2}(A + A')$ and $S = \frac{1}{2}(A - A')$

Hence $\quad A = \frac{1}{2}(A - A') + \frac{1}{2}(A - A')$, is a unique representation.

Example 17. Express $\begin{bmatrix} 2 & 6 & -8 \\ 4 & 2 & 1 \\ -8 & 6 & 13 \end{bmatrix}$ as a sum of a symmetric and skew-symmetric matrix.

Solution: Let $\quad A = \begin{bmatrix} 2 & 6 & -8 \\ 4 & 2 & 1 \\ -8 & 6 & 13 \end{bmatrix} \quad \therefore \quad A' = \begin{bmatrix} 2 & 4 & -8 \\ 6 & 2 & 6 \\ -8 & 1 & 13 \end{bmatrix}$

$$\therefore \quad A + A' = \begin{bmatrix} 4 & 10 & -16 \\ 10 & 4 & 7 \\ -16 & 7 & 26 \end{bmatrix} \therefore \frac{1}{2}(A + A') = \frac{1}{2}\begin{bmatrix} 4 & 10 & -16 \\ 10 & 4 & 7 \\ -16 & 7 & 26 \end{bmatrix}$$

Which is a symmetric matrix.

Again $\quad \frac{1}{2}(A - A') = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ -2 & 0 & -5 \\ 0 & 5 & 0 \end{bmatrix}$

Which is a skew-symmetric matrix.

Now $\quad A = \frac{1}{2}(A + A') + \frac{1}{2}(A - A')$

i.e., $\quad A = \frac{1}{2}\begin{bmatrix} 4 & 10 & -16 \\ 10 & 4 & 7 \\ -16 & 7 & 26 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ -2 & 0 & -5 \\ 0 & 5 & 0 \end{bmatrix}$

Example 18. If A and B are both symmetric then, show that AB is symmetric iff A and B commute.

Solution: Since A and B are symmetric, we have

$A' = A$ and $B' = B$

Then $\quad (AB)' = B'A' \quad$ (reversal law)

$= BA = AB$, iff A and B commute

Thus AB is symmetric iff A and B commute.

52

2.17 Orthogonal Matrix

A square matrix A is said to be orthogonal if $A'A = AA' = I$

Now, we know that

$$|A'| = |A|$$

Also $$|A'A| = |A'||A|$$

or $$|I| = |A|^2 \quad \text{or} \quad |A|^2 = 1$$

This shows that the matrix A should be non–singular and invertible, if it is orthogonal matrix. Hence the condition

$A'A = I$ Implies that $A^{-1} = A'$

Example 19. Verify that $\mathbf{A = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ -2 & 2 & -1 \end{bmatrix}}$ **is orthogonal.**

Solution: We have $$A = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ -2 & 2 & -1 \end{bmatrix}$$

Then $$A' = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ -2 & 2 & -1 \end{bmatrix}$$

We have, $$= \frac{1}{9}\begin{bmatrix} 1+4+4 & 2+2-4 & -2+4-2 \\ 2+2-4 & 4+1+4 & -4+2+2 \\ -2+4-2 & -4+2+2 & 4+4+1 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Hence, by definition, matrix A is orthogonal.

CHECK YOUR PROGRESS (C)

23

1. If A and B are symmetric, then show that A + B is symmetric.
2. If $A = \begin{bmatrix} 5 & 3 \\ 2 & 4 \end{bmatrix}$, verify that $(6A)' = 6A'$.
3. Given: $A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}$ obtain the matrix $2A' + 3B'$.
4. Express $\begin{bmatrix} 2 & 1 & -11 \\ 6 & 13 & 7 \\ -4 & -2 & 1 \end{bmatrix}$ 3 as a sum of a symmetric and skew-symmetric matrix.

Answers

3. $\begin{bmatrix} 10 & 5 \\ -3 & 10 \end{bmatrix}$

4. $\frac{1}{2}\begin{bmatrix} 4 & 7 & -15 \\ 7 & 26 & 5 \\ -15 & 5 & 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & -5 & -7 \\ 5 & 0 & 9 \\ 7 & -9 & 0 \end{bmatrix}$

2.18 SUM UP

- A Matrix is a particular type of arrangement of m n numbers which are arranged in the form of an ordered set of m rows and n columns.
- There are different types of Matrices like Row, Column, Square, Rectangular, Null, Diagonal, Scalar, Unit, Triangular matrix etc.
- Matrix algebra includes addition of matrices, difference of matrices, multiplication of a matrix by a scalar and matrix multiplication.
- We can obtain the Transpose of a Matrix by interchanging the corresponding rows and columns of a given matrix.

2.19 Key Terms

- **Row Matrix:** If a matrix has only one row, it is called a row matrix.
- **Column Matrix:** A matrix consisting of only one column is called a column matrix.
- **Zero or Null Matrix (0):** If every element of an $m \times n$ matrix is zero, the matrix is called a zero matrix.
- **Square Matrix:** Any matrix in which the number of rows are equal to the number of columns is called a Square matrix.
- **Diagonal Matrix:** A square in which all elements except those in the leading diagonal are zero, is called a diagonal matrix.
- **Scalar Matrix:** A diagonal matrix whose diagonal elements are all equal is called a scalar matrix.
- **Identity or Unit Matrix (I):** A scalar matrix in which each of its diagonal elements is unity is called an identity or unit matrix.
- **Trace of a Matrix:** The trace of any square matrix is the sum of its main diagonal elements.
- **Triangular Matrix:** If every element above or below the leading diagonal is zero, the matrix is called a triangular matrix.
- **Transpose of a Matrix:** A matrix obtained by interchanging the corresponding rows and columns of a given matrix.

- **Symmetric Matrix:** A matrix which is equal to its own transpose.
- **Skew Symmetric Matrix:** A matrix which is equal to negative of its own transpose.

2.20 Questions For Practice

Q1.**What is matrix? Explain its types.**

Q2.What do you mean by equality of matrices?

Q3.Explain addition and difference of matrix.

Q4.What are the properties of Addition matrix?

Q5.What is Multiplication of Matrices? Give its properties.

Q6.Applications of Matrices to Business and Economic Problems

Q7.What do you mean by transpose of a matrix? Give an example

Q8.Explain symmetric matrix and Skew-symmetric matrix.

2.21 Further Readings

- Mizrahi and John Sullivan. Mathematics for Business and Social Sciences. Wiley and Sons.
- N. D. Vohra, Business Mathematics and Statistics, McGraw Hill Education (India) Pvt Ltd
- J.K. Thukral. Mathematics for Business Studies. Mayur Publications
- J. K. Singh, Business Mathematics, Himalaya Publishing House.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 3: STATISTICS: DEFINITION, IMPORTANCE AND SCOPE, LIMITATIONS, DISTRUST

STRUCTURE

3.0 Objectives

After reading this unit, learner will be able to learn:

- definitions given by different aspects
- statistics along with its applications
- limitations and misuse of statistics
- various types of data used in statistics will also be explored.

3.1 Introduction

This module is designed to know about the development of statistics using historical background. Statistics is not a subject that can be studied alone, rather it proves to be the basis for almost all other subjects as data handling is essential in almost all fields of life. Due to this reason, a number of definitions are given to statistics. Some of the major fields where statistics are prominently used are planning, finance, business, agriculture, biology, economics, industry, education, etc. Actually, a country's growth is very much dependent on statistics as without statistics it would not be possible to estimate the requirements of the country. However, statistics are based on probabilistic estimations and therefore not actual (in some cases), therefore can't be believed with 100% guarantee. Also, some people may misuse statistics for their own benefit. But still statistics is very essential and very much needed of life. There are various classifications of the data used in statistics viz., continuous, discrete, nominal, ordinal, etc. The data can be used as per requirement for a particular application.

3.2 Meaning of Statistics

Let us look into the meaning of the word 'Statistics'. It conveys different meaning to different people. A common man may simply interpret it as a mass of figures, graphs or diagrams relating to an economic, business or some other scientific activity. However, for an expert, it may also imply a statistical method of investigation in addition to a mere mass of figures. Let us discuss each of these.

3.2.1 Statistics in Plural Sense

Statistics in plural sense means the mass of quantitative information called 'data'. For example, we talk of information on population or demographic features of India available from the Population Census conducted every ten years by the Government of India. Similarly, we can have statistics (quantitative data or simply data) on Also referred to as Statistical Data, Horace Secrist describes statistics in plural sense as follows: "By Statistics we mean aggregates of facts affected to a marked extent .by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other."

This definition of statistics in plural sense highlights the following features:

a) **Statistics are numerical facts**: In order that information obtained from an investigation can be called as statistics or data, it must be capable of being represented by numbers. The collected data may be obtained either by the measurement of characteristics (like data on heights, weights, etc.) Or by counting when the characteristics (like honesty, smoking habit, beauty, etc.) is not measurable.

b) **Statistics are aggregates of facts**: Single and unrelated figures even though expressed as quantities are not statistics. For example, in a university examination Mr. Sharma secures 65% marks does not make statistics or data However, if we find that out of 3 lakh university students whose average marks were 55%, Mr. Sharma secured 65% marks, then these figures are statistics. So, no single figure in any sphere of statistical inquiry, say production, employment, wage and income constitutes statistics.

c) **Statistics are affected to a marked extent by multiplicity of causes**: In physical sciences it is possible to isolate the effect of various forces on a particular event. But in Statistics' facts and figures, that is, the collected information, are greatly influenced by a number of factors and forces working together. For example, the output of wheat in a year is affected by various factors like the availability of irrigation, quality of soils, method of cultivation, type of seed, amount of fertilizer used, etc. In addition to this there may be certain factors which are even difficult to identify.

d) **Statistics are numerically expressed**: Statistics are statements of facts expressed numerically or in numbers. Qualitative statements like "the students of a school ABC are more intelligent than those of school XYZ" cannot be regarded statistics. Contrary to this the statement that 'the average marks in school ABC are 90% compared with 60% in school XYZ, and that the former had 80% first division compared with only 50% in the latter", is a statistical statement.

e) **Statistics are enumerated or estimated with a reasonable degree of accuracy**: While enumerating or estimating statistics, a reasonable degree of accuracy must be achieved. The degree of accuracy needed, in an investigation, depends upon the nature and objective of investigation on one hand and upon the time and resources on the other. Thus, it is necessary to have a reasonable degree of accuracy of data, keeping in mind the nature and objective of investigation and availability of time and resources. The degree of accuracy once decided must be uniformly maintained throughout the investigation.

f) **Statistics are collected in a systematic manner:** Before the collection of statistics, it is necessary to define the objective of investigation. The objective of investigation must be specific and well defined. The data are then collected in systematic manner by proper planning which involves finding of answers to questions such as: Whether to use sample or census investigation, how to collect, arrange, present and analyse data, etc.)

g) **Statistics should be placed in relation to one another:** Only comparable data make some sense. 'Unrelated and incomparable data are no data. They are just figures. For example, heights and weights of students of a class do not have any relation with the income and qualification of their parents. For comparability, the data should be homogeneous; that is, it should belong to the same subject or class or phenomenon. For example, pocket money of the students of a class is certainly related to the income of their parents. Prices of onions and potatoes in Delhi can certainly be related to their prices in other cities of India.

Thus, it will not be wrong to say that "all statistics are numerical statements of facts but all numerical statements of facts are not statistics".

3.2.2 Statistics in Singular Sense

In the singular sense, Statistics refers to what is called statistical methods which means the ever-growing body of techniques for collection, condensation, presentation, analysis and interpretation of statistical data/quantitative information. In simple language, it means the subject of Statistics like any other subject such as Mathematics or Economics.

We can now take up definitions given by some famous statisticians.

A. L. Bowley gave a few definitions but none of them was complete and satisfactory. However, his two definitions make some sense even though incomplete. For example, he says, "Statistics may be called the science of counting". Here he is emphasizing on enumeration aspect of statistics, which no doubt is important at another place he describes statistics as "the science of measurement of the social organism...". He is also of the view that "Statistics may rightly be called the science of average". Although measurement, enumeration and averages (Arithmetic, Geometric and Harmonic means; Mode and Median which we will discuss in the next Block) are important, yet they are not the only concern of Statistics, as we shall study in the subsequent units.

Croxton and Cowden have put forward a very simple and precise definition of Statistics as "Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data". This definition lays emphasis on five important aspects, which in fact, constitute the very scope of the subject called Statistics or Statistical Methods. These are:

A) Collection of data: In any statistical inquiry, the collection of data is the first basic step. They form the foundation of statistical analysis, and therefore utmost care should be taken in collecting data. Faulty data will certainly lead to misleading results and can do more harm than good. The data can be drawn from two sources:

- Primary source where data are generated by the investigator himself through various methods

- Secondary source where data are extracted from the existing published or unpublished source, that is, from the data already collected by others. It saves a lot of time, effort and money of the investigator; but then he has to be conscious and judicious in their use.

B) Arrangement of Data: Data from the secondary source are already arranged or organised like population data from Census of India. A minor rearrangement to suit our needs can be undertaken. However, primary data are in a haphazard form and need some arrangement so that it makes some sense. The steps involved in this process are: -

- Editing: This involves the removal of omissions and inconsistencies involved in the collected information.
- Classification of data: It follows editing. It involves arranging data according to some common characteristic/s. Normally the raw information received from the respondents is put on the master sheets. For example, we may conduct a survey on, say, metal-based engineering industries of Orissa, from where information are collected on capital structure, output of different types of products, employment of unskilled, semi-skilled and skilled workers, cost and price structure, technology aspects, etc. All this information can be put on master sheets.

C) Tabulation: It is the last step in the arrangement process. From the master sheets (or coded sheets) information is tabulated in the form of frequency distributions or tables, where information is arranged in columns and rows.

D) Presentation of data: After the data have been arranged and tabulated, they can now be presented in the form of diagrams and graphs to facilitate the understanding of various trends as well as the process of comparison of various situations. Two different types of presentation of data are normally used, these are:

- Statistical tables
- Graphs including line graphs.

E) Analysis of data: It is the most important step in any statistical inquiry. A major portion of this course in Statistics is devoted to the methods used for analyzing the collected data to derive some policy conclusions.

3.3 Definitions Of Statistics

Statistics has been defined by the number of authors in different ways. The main reason for the various definitions is the changes that have taken place in statistics from time to time. Statistics, in general, is defined in two different ways viz., as "statistical data", i.e., based on the numerical statement of data and facts, and as 'statistical methods, i.e., based on the principles and techniques used in collecting and analyzing such data. Some of the important definitions under these two categories are given below.

Statistics as Statistical Data Webster defines Statistics as "classified facts representing the conditions of the people in a State, especially those facts which can be stated in numbers or in any other tabular or classified arrangement." Bowley defines Statistics as "numerical statements of facts in any department of enquiry placed in relation to each other."

A more exhaustive definition is given by Prof. Horace Secrist as follows: "By statistics we mean aggregation of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

According to Boddington, "Statistics is the science of estimates and probabilities." Again, this definition is not complete as statistics is not just probabilities and estimates but more than that. Some other definitions are: "The science of Statistics is the method of judging collective, natural or social phenomenon from the results obtained from the analysis or enumeration or collection of estimates."- as provided by King. "Statistics is the science which deals with collection, classification, and tabulation of numerical facts as the basis for explanation, description and comparison of the phenomenon." as given by Lovitt. But the best definition is the one given by Croxton and Cowden, according to whom Statistics may be defined as "the science which deals with the collection, analysis and interpretation of numerical data."

CHECK YOUR PROGRESS (A)

Q1. Statistics in Plural Sense.

Ans:____

Q2. Statistics in Singular Sense.

Ans:____

Q3: Collection of data.

Ans:____

3.4 Importance And Scope Of Statistics

Statistics is primarily used either to make predictions based on the data available or to make conclusions about a population of interest when only sample data is available. In both cases, statistics tries to make sense of the uncertainty in the available data. When making predictions statisticians determine if the difference in the data points is due to chance or if there is a systematic relationship. The more the systematic relationship that is observed the better the prediction a statistician can make. The more random error that is observed the more uncertain the prediction. Statisticians can provide a measure of the uncertainty to the prediction. When making inferences about a population, the statistician is trying to estimate how good a summary statistic of a sample really is at estimating a population statistic. For computer students, knowing the basic principles and methods in statistics could help them in doing their research work like comparing the speed of internet connection in different countries and the probability of how many times does each experience the same level of internet connection speed in a week, month or year. It could also be helpful in determining the best operating system to use. Whenever there is the need to compare data and know the best option that we should take statistics can give the answer. Statistics is having applications in almost all sciences - social as well as physical such as biology, psychology, education, economics, business management, etc. It is hardly possible to think of even a single department of human activity where statistics is not involved. It has rather become indispensable in all phases of human endeavor.

1. **Statistics and Planning:** Statistics is mother of planning. In the modern age which is termed as 'the age of planning', almost all over the world, particularly of the upcoming economies, are resorting to planning for economic development. In order that planning is successful, it must be based soundly on the correct analysis of complex statistical data.
2. **Statistics and Economics**: Statistical data and technique of statistical analysis have proved immensely useful in solving various economic problems, such as wages, prices, analysis of time series and demand analysis. A number of applications of statistics in the study of economics have led to the development of new disciplines called Economic Statistics and Econometrics.
3. **Statistics and Business**: Statistics is an essential tool for production control. Statistics not only helps the business executives to know the requirements of the consumers, but also for many other purposes. The success of a business actually depends upon the accuracy and precision of his statistical forecasting. Wrong analysis, due to faulty and inaccurate analysis of various causes affecting a particular phenomenon, might prove to be a disaster. Consider an examples of manufacturing readymade garments. Before starting one must have an overall idea as to 'how many garments are to be manufactured', 'how

much raw material and labour is needed for that', and 'what is the quality, shape, color, size, etc., of the garments to be manufactured'. If these questions are not analysed statistically in a proper manner, the business is bound to be failed. Therefore, most of the large industrial and commercial enterprises are employing trained and efficient statisticians.

4. **Statistics and Industry**: In industry, statistics is very widely used in 'Quality Control'. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, viz. inspection plans, control charts, etc., are of extreme importance.
5. **Statistics and Mathematics**: Statistics and mathematics are very intimately related. Recent advancements in statistical techniques are the outcome of wide applications of advanced mathematics. Main contributors to statistics, namely, Bernoulli, Pascal, Laplace, De-Moirve, Gauss, R. A. Fisher, to mention only a few, were primarily talented and skilled mathematicians. Statistics may be regarded as that branch of mathematics which provided us with systematic methods of analyzing a large number of related numerical facts. According to Connor, " Statistics is a branch of Applied Mathematics which specialise in data."
6. **Statistics and Biology, Astronomy and Medical Science:** The association between statistical methods and biological theories was first studied by Francis Galton in his work in Regression. According to Prof. Karl Pearson, the whole 'theory of heredity' rests on statistical basis. He said, "The whole problem of evolution is a problem of vital statistics, a problem of longevity, of fertility, of health, of disease and it is impossible 6 to discuss the national mortality without an enumeration of the population, a classification of deaths and knowledge of statistical theory." In astronomy, the theory of Gaussian 'Normal Law of Errors' for the study of the movement of stars and planets is developed by using the 'Principle of Least Squares'. In medical science also, the statistical tools for the collection, presentation, and analysis of observed facts relating to the causes and diseases and the results obtained from the use of various drugs and medicines, are of great importance. Moreover, the efficacy of a manufactured drug or injection, or medicine is tested by analyzing the 'tests of significance'.
7. **Statistics and Psychology and Education:** In education and psychology, too, statistics have found wide applications, e.g., to determine the reliability and validity of a test, 'Factor Analysis', etc., so much so that a new subject called 'Psychometry' has come into existence.
8. **Statistics and War:** In war, the theory of 'Decision Functions' can be of great assistance to military and technical personnel to plan 'maximum destruction with minimum effort'.

Thus, we see that the science of Statistics is associated with almost all the sciences - social as well as physical. Bowley has rightly said, "A knowledge of Statistics is like a knowledge of foreign language or algebra; it may prove of use at any time under any circumstance."

9. **Statistics and Physical Sciences:** Statistics has proved to be useful in physical sciences like Physics, Geology, Astronomy, Biology, Medicine, etc. A modern doctor relies heavily on the information on various parameters of a patient in diagnosing his disease. These include his body temperature behaviour, blood pressure and blood sugar level, ECG, etc. Doctor needs this information all the more when performing surgery. Further, before introducing a new drug, data are collected and analysed for its effects on rats, monkeys, rabbits, etc. If found statistically satisfactory, the experiments are then conducted on human beings. The efficacy of the medicine is studied statistically. For example, researchers may be interested in finding whether quinine is still effective in the control of malaria with a new strain of mosquito. They may conduct the experiment on, say, 1000 patients selected at random. If the percentage of success is quite high, researchers may declare that quinine is still effective in the control of malaria.

3.5 Limitations of Statistics

Statistics, with its wide applications in almost every sphere of human activity; is not without limitations. The following are some of its important limitations:

1. **Statistics is not suited to the study of the qualitative phenomenon:** Statistics, being a science dealing with a set of numerical data, is applicable to the study of only those subjects of enquiry which are capable of quantitative measurement. As such; qualitative phenomena like honesty, poverty, culture, etc., which cannot be expressed numerically, are not capable of direct statistical analysis. However, statistical techniques may be applied indirectly by first reducing the qualitative expressions to precise quantitative terms. For example, the intelligence of a group of candidates can be studied on the basis of their scores on a certain test.
2. **Statistics does not study individuals:** Statistics deals with an aggregate of objects and does not give any specific recognition to the individual items of a series. Individual items, taken separately, do not constitute statistical data and are meaningless for any statistical enquiry. For example, the individual figures of agricultural production, industrial output, or national income of any country for a particular year are meaningless unless, to facilitate comparison, similar figures of other countries or of the same country for different years are given. Hence, statistical analysis is suited to only those problems

where group characteristics are to be studied.

3. **Measurement errors:** During the data collection process, measurement errors might happen, resulting in inaccurate results being reported. These errors can be caused by factors such as faulty instruments, human error, respondent bias, or misunderstanding of survey questions. The validity and reproducibility of statistical analysis can be impacted by measurement mistakes.
4. **Sampling bias:** Sampling bias occurs when certain groups or individuals are systematically overrepresented or underrepresented in the sample. This may occur as a result of errors in the sampling procedure or nonresponse bias, where some people decide not to take part in the study. Sampling bias can result in erroneous inferences and unreliable generalizations.
5. **Misuse and misinterpretation:** Statistics can be misused or misinterpreted, leading to incorrect conclusions. Improper statistical analysis, intentional manipulation of data, or selective reporting of results can distort the findings and mislead the audience. It is important to use statistics appropriately and interpret them critically.
6. **Statistical laws are not exact**: Unlike the laws of physical and natural sciences, statistics are only approximations and not exact. Based on statistical analysis, we can talk only in terms of probability and chance and not in terms of certainty. Statistical conclusions are not universally true; rather they are true only on average.

3.6 Distrust of Statistics

Distrust of statistical data refers to a lack of confidence toward the information and findings derived from statistical analysis. This distrust can stem from various factors, including concerns about data collection methods, biases in data interpretation, manipulation or misrepresentation of data, towards the reliability of statistical methods.

There are several reasons why individuals or groups may exhibit distrust of statistical data:

1. **Methodological concerns:** Some people question the accuracy and reliability of data collection methods used to gather statistical information. They may believe that the sampling methods are faulty, the sample size is too small, or the data collection process is biased.
2. **Biases and manipulation**: Skepticism can arise when people suspect that statistical data is manipulated or biased to serve a particular agenda. This can occur through selective data presentation, cherry-picking data, or altering the analysis to support a predetermined conclusion.
3. **Lack of transparency:** When statistical data is not accompanied by transparent and

detailed information about the methodology, sources, and assumptions used, it can lead to distrust. People may be suspicious of hidden biases or vested interests that could influence the results.

4. **Complexity and misinterpretation:** Statistical analysis can be complex, and the interpretation of data requires expertise. Misinterpretation or misrepresentation of statistical findings by individuals or the media can contribute to distrust. People may feel overwhelmed or confused by the numbers, leading them to question their validity.
5. **Historical mistrust**: Historical events or instances of statistical manipulation can contribute to a general distrust of statistical data. Past cases of misleading or falsified statistics erode public trust and make people more cautious about accepting statistical information at face value.

It's essential to remember that critical thinking is necessary for analyzing all data, including statistical data, a total rejection of all statistical data can hinder policy creation, decision-making, and the comprehension of many phenomena. Transparency, clear communication, and strong procedures are needed to address the issues with statistical data in order to deliver accurate and trustworthy information.

3.7 Misuse of Statistics

Statistics is liable to be misused. As they say, "Statistical methods are the most dangerous tools in the hands of the experts. Statistics is one of those sciences whose adepts must exercise the self-restraint of an artist." The use of statistical tools by inexperienced and untrained persons might lead to very fallacious conclusions. One of the greatest shortcomings of statistics is that by just looking at them one can't comment on their quality and as such can be represented in any manner to support one's way of argument and reasoning. The requirement of experience and judicious use of statistical methods restricts their use to experts only and limits the chances of the mass popularity of this useful and important science. It may be pointed out that Statistics neither proves anything nor disproves anything. It is only a tool which if rightly used may prove extremely useful and if misused might be disastrous. According to Bowley, "Statistics only furnishes a tool necessary though imperfect, which is dangerous in the hands of those who do not know its use and its deficiencies." It is not the statistics that can be blamed but those persons who twist the numerical data and misuse them either due to ignorance or deliberately for personal selfish motives. As King pointed out, "Science of Statistics is the most useful servant but only of great value to those who understand its proper use."

CHECK YOUR PROGRESS (B)

Q1. Explain the importance of the statistics

Ans:____

____

Q2. Give any two limitations of statistics

Ans:____

____

Q3. Explain the distrust of statistics

Ans:____

____

Q4. Explain how statistics are misused.

Ans:____

____

3.8 Sum Up

A modern man must possess knowledge of Statistics like that of reading and writing. The word Statistics in a singular sense implies statistical methods aimed at collecting, arranging, presenting, analyzing and interpreting data. In the plural sense, it means mass of quantitative information like population data. The word statistic (as against statistics) means an estimator obtained from a sample with the purpose of inferring about the population value called a parameter. Statistics has utility in almost all branches of knowledge. In Economics and Business, it has a special utility. A combination of Economics, Statistics and Mathematics has led to a new subject called Econometrics. Despite immense utility, some unscrupulous persons have misused statistics driving it to the level that is worse than damned lies. Because of this, sometimes, it has been termed as unscrupulous science.

3.9 Questions For Practice

1. Give a historical background of statistics.
2. Write various definitions of statistics and discuss these definitions in brief.
3. State and explain various applications of statistics.
4. What are the various limitations of statistics?
5. Provide a few examples which can lead to incorrect conclusions due to wrong analysisof statistics.
6. Give any two examples of collecting data from day-to-day life.

3.10 Suggested Readings

- A. Abebe, J. Daniels, J.W. Mckean, "Statistics and Data Analysis".
- Clarke, G.M. & Cooke, D., "A Basic Course in Statistics", Arnold.
- David M. Lane, "Introduction to Statistics".
- S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", Sultan Chand & Sons, New Delhi.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 4: COLLECTION OF DATA: TYPES AND SOURCES

STRUCTURE

4.0 Learning Objectives

On going through this unit, you will be able to:

- Explain the concept and types of data collection
- Sources of Data collection
- various survey techniques under primary data and secondary data
- limitations of primary data and secondary data
- Precautions to Collect Secondary Data

4.1 Introduction

We face problems in various fields of our life, which force us to think and discover their solutions. When we are genuinely serious about the solution of a problem faced, a thinking process starts. Statistical Thinking or Statistical Inquiry is one kind of thinking process that requires evidence in the form of some information, preferably quantitative, which is known as data/statistical information. In a statistical inquiry, the first step is to procure or collect data. Every time the investigator may not start from the very beginning. He must try to use what others have already discovered, this will save us in cost, efforts and time.

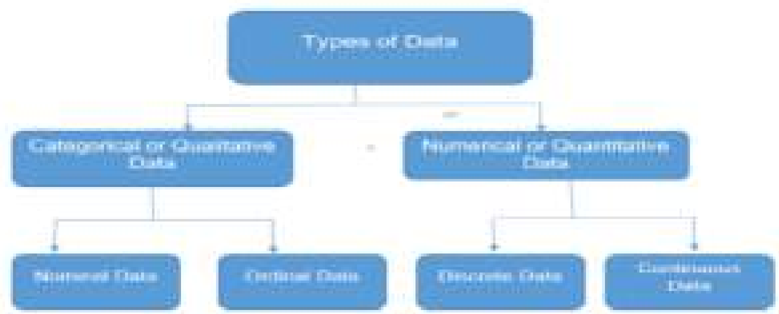
They are collections of any number of related observations with a predetermined goal. We can collect information on the number of T.V. sets sold by a particular salesman or a group of salesmen, on weekdays in different parts of Delhi to study the pattern of sales, lean days, effect of competitive products, income behaviour and other related matters. The information thus collected is called a data set and a single observation a data point. All types of information collected without proper aim or objective is of no use. For example, John's height is 5'6" or monthly wage of Mr. X on 1st January 2004 were Rs.15000/- are not data. Not all quantitative information is statistical. Isolated measurements are not statistical data. Statistics (that is in singular sense) is concerned with collection of data relevant to the solution of a particular problem. According to Simpson and Kafka (Basic Statistics),"Data have no standing in themselves; they have a basis for existence only where there is a problem".

4.2 Types of Data Collection

By now you have known that data could be classified in the following three ways:

a) Quantitative and Qualitative Data
b) Sample and Census Data
c) Primary and Secondary data

a) Quantitative and Qualitative data: Quantitative data are those set of information that are quantifiable and can be expressed in some standard units like rupees, kilograms, liters, etc. For example, pocket money of students of a class and the income of their parents can be expressed in so many rupees; the production or import of wheat can be expressed in so many kilograms or lakh quintals; the consumption of petrol and diesel in India as so many lakh liters in one year and so on. In other words, Qualitative data, also known as the categorical data, describes the data that fits into the categories. Qualitative data are not numerical. The categorical information involves categorical variables that describe the features such as a person's gender, home town etc. Categorical measures are defined in terms of natural language specifications, but not in terms of numbers



1. **Qualitative data:** on the other hand, are not quantifiable, that is, cannot be expressed in standard units of measurement like rupees, kilograms, liters, etc. This is because they are 'features', 'qualities', or 'characteristics' like eye colours, skin complexion, honesty, good or bad, etc. These are also referred to as attributes. In this case, however, it is possible to count the number of individuals (or items) possessing a particular attribute.
   - Nominal Data: Nominal data is one of the types of qualitative information which helps to label the variables without providing the numerical value. Nominal data is also called the nominal scale. It cannot be ordered and measured. But sometimes, the data can be qualitative and quantitative. Examples of nominal data are letters, symbols, words, gender etc. The nominal data are examined using the grouping method. In this method, the data are grouped into categories, and then the frequency or the percentage of the data can be calculated. These data are visually represented using the pie charts.
   - Ordinal Data: Ordinal data is a type of data that follows a natural order. The significant feature of the nominal data is that the difference between the data values is not determined. This variable is mostly found in surveys, finance, economics, questionnaires, and so on. The ordinal data is commonly represented using a bar chart. These data are investigated and interpreted through many visualization tools. The information may be expressed using tables in which each row in the table shows a distinct category. Quantitative or Numerical Data.
2. **Quantitative Data:** Quantitative data is also known as numerical data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of aspecific thing. Some examples of numerical data are height, length, size, weight, and so on. Quantitative data can be classified into two different types based on the data sets. Thetwo different classifications of numerical data are discrete data and continuous data.

- **Discrete Data:** Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in whole numbers e.g., the number of students in the class
- **Continuous Data:** Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range e.g., Temperature range.

The quantitative and qualitative data can be represented as in figure 1.1.

Figure 1.1: Quantitative and Qualitative Data

Figure 1.2 shows the types of qualitative data i.e., discrete and continuous data.

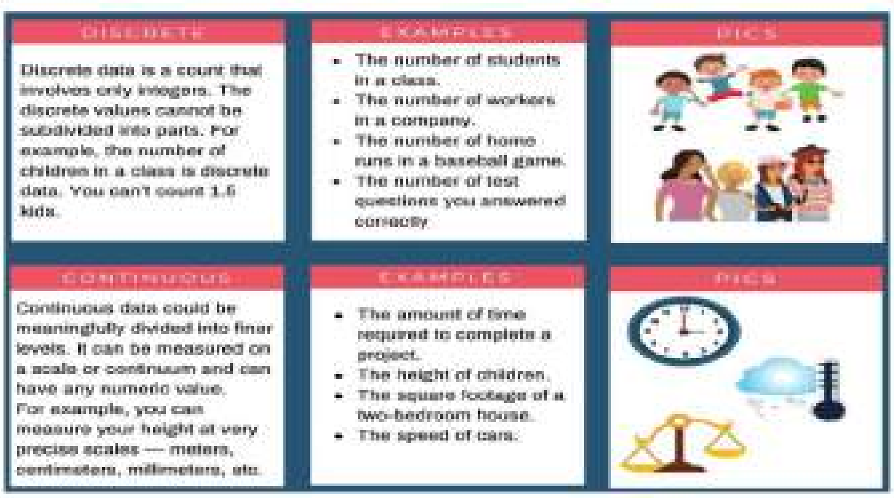


Figure 1.2: Types of Qualitative Data viz., Discrete and Continuous

Figure 1.3 shows the types of quantitative data i.e. nominal and ordinal data.

Figure 1.3: Types of Quantitative Data viz., Nominal and Ordinal

b) Sample and Census Data: Data can be collected either by census method or sample method. Information collected through sample inquiry is called sample data and the one collected through census inquiry is called census data. Population census data are collected every ten years in India.

c) Primary and Secondary Data: Primary data are collected by the investigator through field surveys. Such data are in raw form and must be refined before use. On the other hand, secondary data are extracted from the existing published or unpublished sources, that is; from the data already collected by others. The collection of data is the first basic step towards the statistical analysis of any problem. The collected data are suitably transformed and analyzed to draw conclusions about the population.

These conclusions may be either or both of the following:

(i) To estimate one or more parameters of a population or the nature of the population itself. This forms the subject matter of the theory of estimation.

(ii) To test a hypothesis. A hypothesis is a statement regarding the parameters or the nature of the population.

4.3 Sources of Data Collection

A pertinent question that arises now is how and from where to get data? Data are obtained through two types of investigations, namely,

1) Direct Investigation **or Primary Data** which implies that the investigator collects information by observing the items of the problem under investigation. As explained above, it is the primary source of getting data or the source of getting primary data and can be done through observation or inquiry. In the former we watch an event happening, for example, the number and type of vehicles passing through Vijay Chowk in New Delhi during different hours of the day and night. In the latter, we ask questions from the respondents through a questionnaire (personally or through mail). It is a costly method in terms of money, time, and effort.

2) Investigation through **Secondary Source** which means obtaining data from the already collected data. Secondary data are the other people's statistics, where other people include governments at all levels, international bodies or institutions like IMF, IBRD, etc., or other countries, private and government research organisations, Reserve Bank of India and other banks, research scholars of repute, etc. Broadly speaking we can divide the sources of secondary data into two categories: published sources and unpublished sources. A) Published Sources

1) Official publications of the government at all levels — Central, State, Union

2) Official publications of foreign countries.

3) Official publications of international bodies like IMF, UNESCO, WHO, etc.

4) Newspapers and Journals of repute, both local and international.

5) Official publications of RBI, and other Banks, LIC, Trade Unions, Stock exchanges, Chambers of Commerce, etc.

6) Reports submitted by reputed economists, research scholars, universities, and commissions of inquiry, if made public.

Data Collection Methods

Some main source of published data in India are Central Statistical Organisation (C.S.0.): Which publishes data on national income, savings, capital formation, etc. in a publication called National Accounts Statistics.

National Sample Survey Organisation (N.S.S.O.): Under the Ministry of Statistics and Programme Implementation, this organisation provides us with data on all aspects of the national economy, such as agriculture, industry, labor and consumption expenditure.

Reserve Bank of India Publications (R.B.I.): It publishes financial statistics. Its publications are Report on Currency and Finance, Reserve Bank of India Bulletin, Statistical Tables Relating to Banks in India, etc.

Labour Bureau: Its publications are Indian Labour Statistics, Indian Labour Year Book, Indian Labour Journal, etc.

Population Census: Undertaken by the office of the Registrar General India, Ministry of Home Affairs. It provides us with different types of statistics about the population.

B) Un-published Sources

1) Unpublished findings of certain inquiry committees.

2) Research workers' findings.

3) Unpublished material found with Trade Associations, Labour Organisations and Chambers of Commerce.

CHECK YOUR PROGRESS (A)

Q1) Explain the term quantitative data.

Ans:____

____

Q2) What is primary data?

Ans:____

____Q

3) What is secondary data set?

Ans:____

____

Q4) Define sample and population.

Ans:____

____

4.4 COLLECTION OF PRIMARY DATA SURVEY TECHNIQUES

After the investigator is convinced that the gain from primary data outweighs the money cost, effort and time, she/he can go in for this. She/he can use any of the following methods to collect primary data:

a. Direct Personal Investigation
b. Indirect Oral Investigation
c. Use of Local Reports/ agencies to get information
d. Mailed Questionnaire Method
e. Schedules sent through enumerators

a) **Direct Personal Investigation**: Here the investigator collects information personally from the respondents. She/ he meets them personally to collect information. This method requires much from the investigator such as:

- She/he should be polite, unbiased and tactful.
- She/he should know the local conditions, customs and traditions
- She/he should be intelligent possessing good observation power. Data Collection Methods
- She/he should use simple, easy and meaningful questions to extract information.

This method is suitable only for intensive investigations. It is a costly method in terms of money, effort and time. Further, the personal bias of the investigator cannot be ruled out and it can do a lot of harm to the investigation. The method is a complete flop if the investigator does not possess the above-mentioned qualities.

b) Indirect Oral Investigation: Method This method is generally used when the respondents are reluctant to part with the information due to various reasons. Here, the information is collected from a witness or from a third party who are directly or indirectly related to the problem and possess sufficient knowledge. The person(s) who is/are selected as informants must possess the following qualities:

- They should possess full knowledge about the issue.
- They must be willing to reveal it faithfully and honestly.
- They should not be biased and prejudiced.
- They must be capable of expressing themselves to the true spirit of the inquiry.

c) Use of Local Reports: This method involves the use of local newspaper, magazines and journals by the investigators. The information is collected by local press correspondents and not by the investigators. Needless to say, this method does not yield sufficient and reliable data The method is less costly but should not be adopted where high degree of accuracy or precision is required.

d) Mailed Questionnaire Method: It is the most important and systematic method of collecting primary data, especially when the inquiry is quite extensive. This method entails creating a questionnaire (a collection of questions pertaining to the research area with a chance for respondents to fill in their replies) and mailing it to the respondents with a deadline for responding quickly. the respondents are asked to extend their full cooperation by providing accurate responses and timely submission of the completed questionnaire. By assuring them that the information they provided in the questionnaire will be kept totally secure and hidden, respondents are also given a sense of security. The investigator typically pays the return postal costs by mailing a self-addressed, stamped envelope to achieve a speedy and better response. Researchers, individuals, non-governmental organisations, and occasionally even the government are involved in this technique.

e) Schedules sent through Enumerators: Using enumerators for primary data collection is a common practice in various research studies, surveys, and data collection efforts. Enumerators are individuals responsible for collecting data directly from respondents in the field. this is the method of obtaining answers to the questions in a form that is filled out by the interviewers or enumerators (the field agents who put these questions) in a face-to-face situation with the respondents. The questionnaire is a list of questions that the respondent himself answers in his own handwriting. 'Schedules sent through the enumerators' is the major data collecting technique that is commonly used. This is the case because the earlier ways that have been explained thus far have some drawbacks that this method does not. With the schedule (a list of questions), the enumerators directly contact the respondents, ask them questions, and record their responses.

The questionnaire in primary data is divided into two parts:

1) General introductory part which contains questions regarding the identity of the respondent and contains information such as name, address, telephone number, qualification, profession, etc.
2) Main question part containing questions connected with the inquiry. These questions differ from inquiry-to-inquiry Preparation of the questionnaire is a highly specialized job and is perfected with experience. Therefore, some experienced persons should be associated with it.

Drafting and framing a questionnaire is a critical step in primary data collection. A well-designed questionnaire ensures that you gather relevant and reliable data to address your research objectives. The following few important points should be kept in mind while drafting a questionnaire:

(i) Clearly outline the research objectives and the specific information you want to collect through the questionnaire. Identify the key research questions that need to be answered.

(ii) Make sure your questions are easy to understand. Avoid nonsense and complex language. Keep sentences and questions short and to the point.

(iii) The task of soliciting information from people in the desired form and with sufficient accuracy is the most difficult problem. By their nature people are not willing to reveal any information because of certain fears. Many times they provide incomplete and faulty information. Therefore, it is necessary that the respondents be taken into confidence. They should be assured that their individual information will be kept confidential and no part of it will be revealed to tax and other government investigative agencies. This is very essential indeed. Where providing information is not legally binding, the informant has to be sure and convinced that the results of the survey will help the authorities to frame policies which will ultimately benefit them. It is obvious that some element of good salesmanship is also required in the investigation.

(iv) Make a decision regarding the questions that will be included in the questionnaire. Typical sorts of queries include:

- Closed-ended inquiries: Those who respond select from a set of predetermined responses (such as multiple-choice inquiries).
- Open-ended inquiries: Those that respond provide their own, individual responses.
- Questions using a Likert scale: Determine if respondents agree or disagree with a statement using a scale (such as 1 to 5).
- Semantic differential questions: Request a rating on a scale of good to bad or satisfied to dissatisfied from respondents.

(v) Questions hurting the sentiments of respondents should not be asked. These include questions on his gambling habits, sex habits, indebtedness, etc.

(vi) Questions involving lengthy and complex calculations should be avoided because they require tedious extra work in which the respondent may lack both interests as well as capabilities.

4.5 Limitations Of Primary Data Collection

Primary data refers to data collected firsthand through direct observation, surveys, interviews, experiments, or other data collection methods. While primary data can be valuable for research and analysis, it also has certain limitations. Here are some common limitations of primary data:

1. **Cost and time**: Collecting primary data can be a time-consuming and costly process. It requires resources to design research instruments, recruit participants, conduct data collection, and analyze the data. Therefore, primary data collection may be impractical or unaffordable.
2. **Limited sample size:** Primary data collection often involves a smaller sample size compared to secondary data sources. The sample size may be constrained by factors such as time, budget, or accessibility of the target population. A small sample size may limit the generalizability of the findings to a larger population.
3. **Sampling bias**: Similar to the limitations of statistics, primary data collection can be susceptible to sampling bias. If the sample is not representative of the population of interest, the findings may not accurately reflect the characteristics or behaviors of the larger population. Careful attention must be given to sampling methods to minimize bias.
4. **Response bias**: Response bias occurs when participants in a study provide inaccurate or misleading responses. It can be influenced by factors such as social desirability bias (participants providing responses they think are socially acceptable) or recall bias (participants inaccurately remembering past events). Response bias can undermine the validity and reliability of primary data.
5. **Subjectivity and researcher bias**: Primary data collection methods often involve interaction between the researcher and participants. The subjective interpretation and biases of the researcher can unintentionally influence the data collection process and the responses obtained. Researchers need to be aware of their own biases and take steps to minimize their impact on the data.
6. **Limited scope**: Primary data collection typically focuses on specific research questions or objectives. While this targeted approach can yield detailed insights into specific areas of interest, it may not capture a broader range of factors or provide a comprehensive understanding of the phenomenon being studied. Using secondary data or employing a mixed-methods approach can help overcome this limitation.
7. **Ethical considerations:** Primary data collection involves ethical considerations regarding participant privacy, informed consent, and data protection. Researchers must adhere to ethical guidelines and obtain necessary approvals, which can introduce additional time and logistical constraints.

Understanding these limitations of primary data can help researchers and analysts make informed decisions about data collection methods and consider the strengths and weaknesses of primary data about their research objectives. It may also be beneficial to supplement primary data with secondary data sources to enhance the breadth and depth of the analysis.

CHECK YOUR PROGRESS (B)

Q1. Explain direct personal investigation and indirect oral investigation

Ans.____

____

Q2. Define the mailed questionnaire method and schedules sent through enumerators

Ans.____

____

Q3. Give limitations of primary data

Ans.____

____

4.6 Collection of Secondary Data Sources

The direct investigation, though desirable, is costly in terms of money, time and effort. Alternatively, information can also be obtained through a secondary source. It means drawing or collecting data from the already collected data of some other agency. Technically, the data so collected are called secondary data.

Secondary data sources in statistics refer to existing data that has been collected by someone else or for a different purpose but can be utilized for statistical analysis. These sources provide a wealth of information that can be used to explore research questions, test hypotheses, and derive insights. Here are some common secondary data sources used in statistics:

1. **Government agencies**: Government agencies at the local, national, and international levels collect and maintain a vast amount of statistical data. Examples include census data, labor statistics, economic indicators, crime rates, health statistics, and demographic information. These datasets are often publicly available and can provide valuable insights into various social, economic, and demographic trends.

2. **Research organizations and institutes:** Many research organizations and institutes conduct surveys, studies, and data collection efforts for specific research purposes. These organizations may focus on topics such as education, public health, social issues, or specific industries. Their datasets can provide detailed information on specific domains or research areas.

3. **International organizations**: International organizations, such as the World Bank, International Monetary Fund (IMF), United Nations (UN), and World Health Organization (WHO), collect and maintain extensive datasets on global development, economics, health, and social indicators. These datasets cover a wide range of countries and can be used for comparative analysis and cross-country studies.

4. **Academic institutions**: Universities and research institutions often conduct research studies and surveys, resulting in datasets that can be valuable for statistical analysis. These datasets may cover various disciplines, including social sciences, psychology, economics, education, and more. Academic institutions often make their datasets available to researchers, subject to certain restrictions and ethical considerations.

5. **Nonprofit organizations**: Nonprofit organizations focused on specific causes or social issues often collect data related to their mission. These organizations may conduct surveys, compile reports, or collaborate with other entities to collect data. Their datasets can provide insights into areas such as poverty, environmental issues, human rights, and social justice.

6. **Commercial data providers**: There are commercial entities that collect, aggregate, and sell datasets on various industries, market trends, consumer behavior, and more. These datasets can be useful for market research, business analytics, and understanding consumer preferences and trends.

7. **Online platforms and social media:** Online platforms and social media networks generate vast amounts of data. This data includes user-generated content, interactions, behaviors, and demographic information. While accessing and analysing this data may require specific permissions and compliance with privacy regulations, it can offer insights into online behavior, sentiment analysis, and social network analysis.

When using secondary data sources, researchers should consider factors such as the data quality, reliability, representativeness, and potential limitations or biases. It is essential to critically evaluate the data source and ensure that it aligns with the research objectives and analytical requirements

4.7 Limitations Of Secondary Data

Although the secondary source is cheap in terms of money, time and effort, utmost care should be taken in their use. It is desirable that such data should be vast and reliable and the terms and definitions must match the terms and definitions of the current inquiry. The suitability of the data may be judged by comparing the nature and scope of the present inquiry with that of the original inquiry. Secondary data will be reliable if these were collected by unbiased, intelligent and trained investigators. The time period to which these data belong should also be properly scrutinized.

Secondary data refers to data that is collected by someone else for a different purpose but can be utilized for research or analysis. While secondary data can be convenient and cost-effective, it also has certain limitations. Here are some common limitations of secondary data collection:

1. **Lack of control over data collection:** Since secondary data is collected by others, researchers have no control over the data collection process. This can result in data that may not perfectly align with the research objectives or may lack specific variables or measures that the researcher requires. The data may not have been collected with the same level of rigor or precision as desired.
2. **Data relevance and accuracy**: The relevance and accuracy of secondary data can vary. It may be challenging to find secondary data that precisely matches the research needs, as the data may be outdated or collected using different methodologies. In some cases, the data may contain errors, inconsistencies, or missing values, which can affect its reliability and validity.
3. **Limited contextual information**: Secondary data may lack detailed information about the context in which it was collected. Understanding the specific circumstances, conditions, or nuances surrounding the data collection process may be crucial for accurate interpretation and analysis. Without sufficient contextual information, the researcher may face challenges in fully understanding and interpreting the data.
4. **Potential bias and validity concerns**: Secondary data may contain inherent biases or limitations introduced by the original data collection process. The biases could be due to the research design, sampling methods, or data collection instruments used. Researchers must critically evaluate the reliability and validity of the secondary data source to ensure its suitability for their research objectives.
5. **Incompatibility and inconsistency**: When working with secondary data from multiple sources, researchers may encounter issues of incompatibility and inconsistency. The data may

have been collected using different formats, classifications, or units of measurement, making it challenging to combine or compare the data effectively. Harmonization or standardization efforts may be necessary to address these issues.

6. **Limited control over variables**: Secondary data may not include all the variables of interest to the researcher. Certain variables that are critical for the research objectives may be missing, limiting the scope of analysis or preventing the investigation of specific relationships or factors.
7. **Data availability and access**: Accessing certain types of secondary data can be challenging due to restrictions, copyright issues, or proprietary considerations. Researchers may face limitations in obtaining the specific data they need or may need to rely on aggregated or summarized data, which may not provide the level of detail required for the research.

Despite these limitations, secondary data can still be a valuable resource for researchers, providing a foundation for analysis, hypothesis generation, and comparison with primary data. Researchers should critically evaluate the quality and relevance of the secondary data and consider its limitations in the interpretation and analysis process.

4.8 Precautions To Collect Secondary Data

According to Prof. A.L. Bowley, "It is never safe to take the published statistics at their face value without knowing their meaning and limitations and it is always necessary to criticize the arguments that can be based upon them." In using secondary data, we should take special note of the following factors.

1) Reliable, 2) Suitable, and 3) Adequate.

Firstly, the reliability of data has to be the obvious requirement of any data, and more so of secondary data. The user must make himself/herself sure about it. For this (s)he must check whether data were collected by reliable, trained and unbiased investigators from dependable sources or not.

Second, we should see whether data belong to almost the same type of class of people or not. (1) To look at and compare the given inquiry's objectives, nature, and scope with the original research. To verify that all of the terms and units were uniformly defined throughout the previous investigation and that these definitions are appropriate for the current investigation as well. For instance, a unit can be defined in multiple ways depending on its context, such as a household, wage, price, farm, etc. The secondary data will be considered inappropriate for the present research if the units were identified differently in the original investigation than what we want. lastly, consider the variations in data collecting periods and consistency of conditions comparing the original investigation and the present investigation.

Third, even if the secondary data are reliable and suitable in, it might not be adequate for the particular inquiry's objectives. This happens if the original data refers to an area or a period that is much larger or smaller than the needed one, or when the coverage given in the initial research was too narrow or too wide than what is desired in the current research. Therefore, it is making sure that due to the gap of time, the conditions prevailing then are not much different from the conditions of today with respect to habits, customs, fashion, etc. Of course, we cannot hope to get the same conditions.

The suitability of data is another requirement. The research worker must ensure that the secondary data he plans to use suits his inquiry. He must match the class of people, geographical area, definitions of concepts, unit of measurement, time and other such parameters of the source he wants to use with those of his inquiry. Not only this, but the aim and objectives should also be matched for suitability.

Secondary data should not only be reliable and suitable, but also adequate for the present inquiry. It is always desirable that the available data be much more than required by the inquiry. For example, data on, say, the consumption pattern of a state cannot be derived from the data on its major cities and towns.

CHECK YOUR PROGRESS (C)

Q1. Explain the method of collect secondary data.

Ans.____

Q2. Define the Mailed questionnaire method and schedules sent through the enumerator? Give two limitations of primary data.

Ans.____

Q3. Give two limitations of secondary data.

Ans.____

4.9 Sum Up

Data Collection Methods Data / Statistics are quantitative information and can be distinguished as sample or census data; primary or secondary data. We require information for an investigation that can be gathered from either a primary source or a secondary source. Both require statistical surveys, which have two stages: planning and execution. The investigator should choose the primary or secondary sources, census or sample inquiry, type of statistical units and measurement units, level of precision desired, and other factors during the design stage. In the execution stage, the chief investigator has to set up administration, select and train field staff and supervise the entire process of data collection. Using secondary data from published or unpublished sources requires caution because they can lead to a number of problems. The questionnaire method is the most crucial of all survey methods. A questionnaire provides a list of relevant inquiries, which should be short, clear, and of the Yes/No variety with illustrative responses. They shouldn't have a lengthy list. Questions that are private or humiliating should be avoided.

4.10 Questions For Practice

1. What are the techniques to collection of data
2. What do you mean by primary data?
3. What are the sources of primary data?
4. What is questioner. What are the points to kept in mind before drafting questioner?
5. Explain the term secondary data with its sources
6. limitations of primary data and secondary data
7. Precautions to Collect Secondary Data

4.11 Suggested Readings

- A. Abebe, J. Daniels, J.W. Mckean, "Statistics and Data Analysis".
- Clarke, G.M. & Cooke, D., "A Basic course in Statistics", Arnold.
- David M. Lane, "Introduction to Statistics".
- S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", Sultan Chand & Sons, New Delhi.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 5: CLASSIFICATION AND TABULATION OF DATA

STRUCTURE

5.0 Objectives

On going through this Unit, you will be able to explain:

- stages of statistical inquiry after data have been collected
- classification of data
- methods of organizing (classification and arrangement) and condensing statistical data
- concepts of frequency distribution for individual, discrete and continuous series
- Tabulation of data
- Parts of table in statistics

5.1 Introduction

Data collected either from census or sample inquiry, that is from primary source, are always hotchpotch and in rudimentary form. To start with, they are contained in hundreds and thousands of questionnaires. To make a head and tail out of them, they must be organized, (i.e., classified and arranged) summarised. For this purpose, we can use various methods like preparing master sheets in which various information are recorded directly hm the questionnaires. From these sheets small summary tables can be prepared manually. Now-a-days computers can be used for organisation and condensation of data more swiftly, efficiently and in much less time. Some computer softwares are available which helps us to construct various types of graphs and diagrams. Data can be summarized numerically also. Here we use summary measures like measures of central tendency (such as Arithmetic, Geometric and Harmonic Means, Mode and Median); measures of dispersion (such as Range, Quartile Deviation, Mean Deviation, and Standard Deviation); measures of association in bivariate analysis (such as Correlation and Regression), Index Numbers, etc. In this Unit, we plan to discuss how data can be summarized using tables and graphs. It must be kept in mind that a good summarization and presentation of data is not undertaken for its own sake. It is not an end in itself. It sets the stage for useful analysis and interpretation of data. Again, a good presentation helps us to highlight significant facts and their comparisons. Figures can be made to speak out thereby making possible their intelligent use.

This module is designed to know about the representation of data in tabular and graphical forms. For analyzing the statistical data, it must be represented in a tabular form and this module does the same, i.e., describes the techniques to convert the data into tabular forms. To plan and interpret the data, visual effects are very useful and necessary. The visual effects in statistics can be obtained by representing the data through graphs.

5.2 Classification Of Data

According to Tuttle A.M., "A classification is a scheme for breaking a category into a set of parts, called classes, according to some precisely defined differing characteristics possessed by all the elements of the category"

Thus classification impresses upon the arrangement of the data into different classes, which are to be determined depending upon the nature, objectives and scope of the enquiry. This classification is based on sex, age, religion, weight, height, and no. of other factors.

5.3 Functions Of Classification

The functions of classification are as follows:

1. **Summarization:** Classification presents the heavy raw data in a reduced form readily comprehensible to the mind and attempts to highlight the significant features contained in the data.

2. **To make data comparable:** Classification enables us to make meaningful comparisons depending on the basis or criterion of classification. For example, the classification of the students in the university according to sex enables us to make a comparative study of the prevalence of university education among males and females.
3. **To make relationships among data:** The classification of the given data w.r.t. two or more criteria. say, the sex of the students and the faculty they join in the university will enable us to study the relationship between these two criteria.
4. **Statistical treatment of the data:** arrangement of the big heterogeneous data into relatively homogeneous groups as per their points of similarities makes it more intelligible, useful, and readily willing for further processing like tabulation, analysis, and interpretation of the data heterogeneous data into relatively homogeneous groups or classes according to their points of similarities introduces homogeneity or uniformity for further processing like tabulation, analysis and interpretation of the data.
5. **Decision Making:** Classification of data-informed decision-making by providing structured information about different groups or categories. This is especially valuable in business, healthcare, and other fields where decisions are based on data-driven insights.
6. **Report Generation:** When presenting statistical results, classification of data simplifies the presentation of complex data by presenting it in a categorized format. This aids in conveying information to different audiences clearly and effectively.
7. **Data Visualization:** in the case of visualization of data, classification facilitates the creation of various types of data visualizations, such as bar charts, pie charts, and histograms. These visualizations help convey the distribution and characteristics of data within each category.

5.4 Basis Of Classification

The basis of classification refers to the criteria or attributes used to group and categorize data into distinct classes or categories. The choice of basis for classification depends on the nature of the data and the goals of the analysis. Here are some common bases of classification in statistics:

- **Geographical Basis:** Data can be classified based on geographical regions, such as countries, states, cities, or other specific locations. This is useful when analyzing data that varies across different locations.
- **Chronological Basis:** Data can be classified based on differences in time. For example, lose of different years, profit, production, demand and supply, etc. for different periods either by increasing or decreasing time. This is also called time series data usually used in economics and business.

- **Quantitative Basis:** Data can be classified based on qualitative data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of aspecific thing. Some examples of numerical data are height, length, size, weight, and so on. Quantitative data can be classified into two different types based on the data sets. The two different classifications of numerical data are discrete data and continuous data.
- **Qualitative data:** Data can be classified based on qualitative data, and cannot be expressed in standard units of measurement like rupees, kilograms, liters, etc. This is because they are 'features', 'qualities', or 'characteristics' like eye colors, skin complexion, honesty, good or bad, etc. These are also referred to as attributes. In this case, however, it is possible to count the number of individuals (or items) possessing a particular attribute.

5.5 FREQUENCY DISTRIBUTION

Frequency Distributions When observations, discrete or continuous, are available on a single characteristic of a large number of individuals, often it becomes necessary to condense the data as far as possible without losing any information of interest. For condensing the data, it is represented using either discrete or continuous frequency distribution tables.

Terms used in a frequency distribution

Class Interval: The whole range of variable values is classified into some groups in the form of intervals. Each interval is called a class interval.

Class Frequency: The number of observations in a class is termed the frequency of the class or class frequency.

Class limits and Class boundaries: Class limits are the two endpoints of a class interval that are used for the construction of a frequency distribution. The lowest value of the variable that can be included in a class interval is called the lower-class limit of that class interval. The highest value of the variable that can be included in a class interval is called the upper-class limit of that class interval.

5.5.1 Simple Array (Individual observations): when raw data is arranged in ascending or descending order of magnitude is called arraying of data.

For example, the weekly wages paid to the workers are given below:

300, 240, 150, 160, 145, 120, 320, 140, 130, 175, 143

A simple array in ascending order is:

120, 130, 140, 143, 145, 150, 160, 175, 240, 300, 320

A simple array in descending order is:

320, 300, 240, 175, 160, 150, 145, 143, 140, 130, 120

5.5.2 Discrete or ungrouped frequency distribution: In a discrete frequency distribution, values of the variable are arranged individually. The frequencies of the various values are the number of times each value occurs. For example, the weekly wages paid to the workers are given below.

300, 240, 240, 150, 120, 240, 120, 120, 150, 150, 150, 240, 150, 150, 120, 300, 120, 150, 240, 150, 150, 120, 240, 150, 240, 150, 120, 120, 240, 150.

There are various ways to form a frequency distribution for this data. In the first case, let us assume that data is represented in terms of tally marks in a tabular manner as shown below in Table 2.1:

Table 2.1: Representation of Data using Tally Marks

Wages Months	**Marks**	**Marks**
120	𝍸 III	8
150	𝍸 𝍸 II	12
240	IIII III	8
300	II	2

This data can also be represented without using marks i.e. using frequency only as shown in Table 2.2 which is known as the frequency table.

Table 2.2: Frequency Table for the Data in Table 2.1

Weekly Wages (x)	120	150	240	300	**Total**
No. of Workers (f)	8	12	8	2	**30**

The frequency table 2.1 is an ungrouped frequency table.

5.5.3 Grouped Frequency Distribution:

A grouped frequency distribution is a method used to organize and present quantitative data in a wider range of datasets. The process involves dividing the data into intervals or groups, each representing a specific range of values. By doing so, the distribution of data becomes clearer, and patterns or trends are easier to identify.

To create a grouped frequency distribution, the first step is to determine the range of the data, which is the difference between the highest and lowest values. Next, the number of intervals is chosen based on the desired level of detail. The interval width is calculated by dividing the range by the number of intervals. A starting point is selected, often slightly below the lowest data value, and intervals are defined accordingly. The frequency of data points within each interval is then counted and recorded in a table.

This table typically includes columns for intervals, frequencies, and sometimes additional columns for cumulative frequencies and midpoints. Cumulative frequencies provide a running total of data points up to a certain interval, while midpoints offer an average value within each interval.

A grouped frequency distribution simplifies the data representation, making it more accessible for analysis, comparison, and interpretation. It condenses the information while retaining the essential characteristics of the data's distribution. When interpreting the table, researchers and analysts can quickly grasp the overall distribution patterns and draw meaningful insights. Careful consideration should be given to the choice of interval width and the number of intervals to ensure that the distribution accurately reflects the data's nature.

We can also draw a grouped frequency table depending on the data we have. For designing a grouped frequency table, let us consider the following example regarding daily maximum temperatures in a city for 50 days.

28, 28, 31, 29, 35, 33, 28, 31, 34, 29, 25, 27, 29, 33, 30, 31, 32, 26, 26, 21, 21, 20, 22, 24, 28, 30, 34, 33, 35, 29, 23, 21, 20, 19, 19, 18, 19, 17, 20, 19, 18, 18, 19, 27, 17, 18, 20, 21, 18, 19.

Table 2.3 Temperatures in a city for 50 days

Class Interval	Frequency
17-21	17
22-26	9
27-31	13
32-36	11
Total	50

The classes of type 17-21 and 22-26 are inclusive i.e. both the lower bound and upper bound are included in the limit.

Although there are no hard and fast rules that have been laid down for it The following points may be kept in mind for classification:

(i) The classes should be clearly defined and should not lead to ambiguity.

(ii) The classes should be exhaustive, i.e., each of the given values should be included in one of the classes.

(iii) The classes should be mutually exclusive and non-overlapping.

(iv) The classes should be of equal width. The principle, however, cannot be rigidly followed.

(v) Indeterminate classes, e.g., open-ended classes such as less than 'a' or greater than 'b' should be avoided as far as possible since they create difficulty in analysis and interpretation.

(vi) The number of classes should neither be too large nor too small. It should preferably lie between 5 and 15. However, the number of classes may be more than 15 depending upon the total frequency and the details required. But it is 15 desirables that it is not less than 5 since in that case, the classification may not reveal the essential characteristics of the population.

In Table 2.3, the class intervals are 17-21, 22-26, 27-31 and 32-36. Here, say for the class 17-21, the lower-class limit is 17 and the upper-class limit is 21. Both 17 and 21 are part of this class. This is called an inclusive class. Another type of class is an exclusive class as shown below in

5.3.4 Continuous Frequency Distribution

A continuous frequency distribution is used to present and analyze quantitative data that takes on a wide range of possible values within a continuous range. Unlike discrete data, which comprises distinct, separate values, continuous data includes values that can take any real number within a given interval. The concept of a continuous frequency distribution is particularly relevant when dealing with measurements such as height, weight, temperature, or time.

Creating a continuous frequency distribution involves dividing the entire range of data into intervals, often referred to as "class intervals". These intervals are constructed in a way that ensures they are non-overlapping and collectively cover the entire range of data. The frequency of data points falling within each interval is then counted, and this information is typically represented in a table or histogram.

Continuous frequency distributions are essential tools in data analysis, enabling us to explore and summarize large datasets while preserving the integrity of the data's continuous nature. The accuracy of the distribution heavily relies on the choice of class intervals and the visualization method used to represent the data, making thoughtful consideration of these factors crucial for meaningful interpretation.

Forms of Continuous Frequency Distribution

1. **Inclusive Class Interval**

The inclusive type of data has the class interval 20-29, 30-39, 40-49, and 50-59, in which the upper limit and the lower limit are included in the class. The frictional values between 29 to 30 cannot be accounted for in such a classification. Therefore, inclusive type classification is used in grouped frequency distribution for discrete values like no. of students in the class, no. of road accidents, etc., here the value takes an only integer value.

2. **Exclusive Class Intervals**

let us consider the temperature in exclusive form.

Temperature	Frequency
17 but less than 21	17
21 but less than 25	7
25 but less than 29	10
29 but less than 33	9
33 but less than 37	7
Total	50

In Table 2.4 upper values are excluded from the class i.e., in the class 17-21 only values from 17 to 20 are taken and the values of 21 are considered in the next class. Such a type of distribution is known as an exclusive class.

3. Open-ended Frequency distribution

It may be the case that some values in the data set are extremely small compared to the other values of the data set and similarly some values are extremely large in comparison. Then what we do is we do not use the lower limit of the first class and the upper limit of the last class. Such classes are called open-end classes. A distribution of open-ended frequencies with at least one end open is known as an open-end distribution. The first class's lower limit, the last class's upper limit, or both are not specified. "Below" or "less than" and "above' or "greater than" are used.

Table 2.5: Open-end Class Grouped Frequency Table

Marks	No. of students
Less than 20	5
20-40	14
40-60	27
60-80	30
80 & more	35

Size of the Class: The length of the class is called the class width. It is also known as class size. size of the class = Upper Limit-Lower Limit

Mid-point of the Class: The midpoint of a class interval is called the Mid-point of the Class. It is the representative value of the entire class.

Mid-point of the class = (Upper Limit + Lower Limit) / 2

Continuous Frequency Distribution: If we deal with a continuous variable, it is not possible to arrange the data in the class intervals of the above type.

4. Unequal Class Frequency Distribution

The classes of a frequency distribution may or may not be of equal width. A frequency distribution with unequal class width is reproduced in the Table below. Here, the width of the Ist, 2nd, and 5th classes is 10, while that of the 3rd is 20 and that of the 4th is 25.

Table of Unequal Class Interval

Class Interval	Frequency
0-10	6
10-20	10
20-40	14
40-65	25
65-75	30

5. Cumulative Frequency Distribution

A cumulative frequency distribution is a statistical representation of quantitative data that shows the total number of observations that fall below or within a certain value or interval. It provides a way to understand the distribution of data in terms of cumulative frequencies, allowing for insights into the overall spread and concentration of the data. This cumulative frequency distribution is of two types:

- less than type cumulative frequency
- more than type cumulative frequency

a) **Less Than Type Cumulative Frequency**: less than type cumulative frequency for any value of the variable is obtained by adding successively the frequencies of all the previous values, including the frequency of the variable against which the totals are written, provided the values are arranged in ascending order.

For example,

Marks Class Interval	Frequency	Cumulative Frequency
Less than 50	15	15
Less than 100	18	15+18 = 33
Less than 150	40	33+40 = 73
Less than 200	45	73 +45 = 118
Less than 250	30	118+30 = 148
Less than 300	25	148+25 = 173
Less than 350	20	173+20 = 193

b) More Than Type Cumulative Frequency: this is obtained similarly by finding the cumulative totals of frequencies starting from the highest value of the variable to the lowest values. For example:

Marks Class Interval	Frequency	Cumulative Frequency
More than 50	5	56
More than 100	10	51
More than 150	11	41
More than 200	13	31
More than 250	11	18
More than 300	5	7
More than 350	2	2

5.6 Tabulation Of Data: Meaning

Tabulation of data means the systematic presentation of the information contained in the data i.e., in the form of rows as well as columns as per the required objective or features. In a table, a row denotes a horizontal arrangement of data, whereas a column denotes a vertical arrangement. A table's rows and columns are indicated by the proper stubs and captions (or headers or subheadings), respectively, to describe the type of information provided. Data should be presented logically, simply, and unambiguously in tabular form.

Professor Bowley in his manual of Statistics refers to tabulation as "the intermediate process between the accumulation of data in whatever form they are obtained, and the final reasoned account of the result shown by the statistics".

Tabulation is a midway process between the collection of the data and statistical analysis. Rather, tabulation is the final stage in the collection and compilation of the data and forms the gateway for further statistical analysis and interpretations. Tabulation makes the data understandable and facilitates comparisons, and the work of further statistical analysis, averaging, correlation, etc different tools of statistics. It makes the data suitable for further representation in the form of diagrams as well as graphics.

There are no rigid rules for tabulating the statistical data. To construct an excellent table, one has to have a clear understanding of the information to be presented the points that should be emphasized, and familiarity with the table process of creation. To ensure that the relationship between the data of one or more series, as well as the significance of all the figures given in the classification adopted, the organization of data tabulation requires careful consideration. data represented in the table shows comparisons and contrasts. only the ability, knowledge, experience, and common sense of the tabulator, while keeping in mind the nature, scope, and aims of the inquiry can produce a good table.

5.8 Parts of a Table

Based on the type of data and the goal of the study, the various components of a table vary from problem to problem. But the following elements must be present in a decent statistical table:

- Table number
- Title
- Head notes or Prefatory notes
- Captions and Stubs
- Body of the table
- Foot-note
- Source note

1) **Table number**: It is required for the identification of a table mainly when there is more than one table in a particular analysis. The table number is always mentioned in the center at the top or left side depending upon the researcher's choice.

2) **Title of the table**: It indicates the type of information contained in the body of the table. The title of the table provides a brief description of the contents of the table. It exactly describes the nature of the data, place (region or variable), time, and source of the data. It should be brief and complete to describe the nature of the table.

3) **Head notes**: It is also called prefactory notes are written just below the title. It shows contents and units of measurement like (rupees lakh) or (lakh quintals) or (thousand rupees). It should be written in brackets and should appear on the right-side top just below the title. However, every table does not need a head note, like the number of students in each class.

4) **Stubs and Captions**: Stubs are used to designate rows. They appear on the left-hand column of the table. Stubs consist of two parts: a) Stub head describes the nature of stub entry. b) Stub entry is the description of row entries. while captions are called box heads, designate the data presented in the columns of the table. It may contain more than one column head, and each column head may be subdivided into more than one sub-head. For example, we can divide the students of a college into Section A and Section B and then again into males and females.

5) Main body of the table: It is also called the field of the table, and is its most important and immense part. It contains the relevant numerical information which is already contained in the title of the table. For creating as useful it contains row and column totals separately and then a total.

6) Foot Note: footnotes are used to take further elaboration or some additional information. is a qualifying statement put just below the table (at the bottom). Its purpose is to caution about the limitations of the data or certain omissions. It will use the symbols *, **, etc.

7) Source of data: It may be the last part of a table, yet it is important. It speaks about the authenticity of the data taken into the table. It should be below the footnotes. It is generally required for the secondary data collection, here it explains from where the data has been taken. It contains table no. volume, issue no. page number.

Format of a Blank Table

Title

(Head Note or Prefatory Note)

<table>
<tr><th rowspan="3">Sub Heading ↓</th><th colspan="4">Caption</th><th rowspan="3">Total</th></tr>
<tr><th colspan="2">Sub-Heading 1</th><th colspan="2">Sub-Heading 2</th></tr>
<tr><td>C1</td><td>C2</td><td>C1</td><td>C2</td></tr>
<tr><td>R1</td><td colspan="4" rowspan="3">Body of Table</td><td>Total R1</td></tr>
<tr><td>R2</td><td>Total R2</td></tr>
<tr><td>R3</td><td>Total R3</td></tr>
<tr><td>Total</td><td>Total C1</td><td>Total C2</td><td>Total C1</td><td>Total C2</td><td>Grand Total</td></tr>
</table>

Foot Note:

Source:

5.9 Questions for Practice

Q1. What do you mean by the classification of data?

Q2. What are the functions of classification?

Q3. Explain the basis of classification.

Q4. Explain frequency distribution.

Q5. Explain discrete or ungrouped frequency distribution.

Q6. Discuss continuous or grouped frequency distribution with an example.

Q7. What are the various forms of frequency distributions?

Q8. What is a tabulation of data?

Q9. What are the parts of a table? Explain with an example.

5.10 Suggested Readings

- A. Abebe, J. Daniels, J.W. Mckean, "Statistics and Data Analysis".
- Clarke, G.M. & Cooke, D., "A Basic Course in Statistics", Arnold.
- David M. Lane, "Introduction to Statistics".
- S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", SultanChand & Sons, New Delhi.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 6: DIAGRAMMATIC AND GRAPHICAL PRESENTATION OF DATA

STRUCTURE

6.0 Learning Objectives

After studying the Unit, students will be able to:

- Processing the data after collection
- Passes through various stages of Processing
- Plan the data analysis
- Classify the data

- Tabulate the data
- Analyse the data
- Data Preservation, storage and Retention

6.1 Introduction

After data collection, the researcher turns his focus of attention to the processing and preserving of the data. "Diagrammatic" and "Graphical" presentation of data both refer to methods of visually representing information to enhance understanding and analysis. While the terms are often used interchangeably, they can have slightly different connotations.

A diagrammatic presentation involves using diagrams or visual aids to represent data. Diagrams are usually simplified and symbolic representations that convey information clearly and straightforwardly. Common types of diagrammatic presentations include bar charts, pie charts, line graphs, histograms, scatter plots, and pictograms. These visual representations help illustrate patterns, trends, comparisons, and relationships within the data.

Graphical presentation refers to the use of graphs, charts, and visual elements to display data in a way that makes it easier to interpret and analyze. Graphs can include various types of charts, plots, and diagrams that present data in a visual format. This type of presentation is particularly useful when dealing with complex data sets or when you want to emphasize relationships and trends. Graphical presentations can include more detailed and sophisticated visualizations, such as 3D graphs, heat maps, area charts, and more.

6.2 Data Processing

Data processing refers to certain operations such as editing, coding, computing of the scores, preparation of master charts, etc. A researcher has to make a plan for every stage of the research process. As such, a good researcher makes a perfect plan for processing and analysis of data. To some researchers' data processing and analysis is not a very serious activity.

Data processing occurs when data is collected and translated into usable information. Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly so as not to negatively affect the end product or data output.

Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

Stages of Data Processing

1) **Data Collection:** Collecting data is the first step in data processing. Data is pulled from available sources. The data sources available must be trustworthy and well-built so the data collected (and later used as information) is of the highest possible quality.
2) **Data Preparation:** Once the data is collected, it then enters the data preparation stage. Data preparation, often referred to as "pre-processing" is the stage at which raw data is cleaned up and organized for the following stage of data processing. During preparation, raw data is diligently checked for any errors. The purpose of this step is to eliminate bad redundant, incomplete, or incorrect data and begin to create high-quality data for the best business intelligence.
3) **Data Input:** The data is then entered into its destination and translated into a language that it can understand. Data input is the first stage in which raw data begins to take the form of usable information.
4) **Processing:** During this stage, the data inputted to the computer in the previous stage is processed for interpretation. Processing is done using machine learning algorithms, though the process itself may vary slightly depending on the source of data being processed.
5) **Data output/interpretation:** The output/interpretation stage is the stage at which data is finally usable to non-data scientists. It is translated, readable, and often in the form of graphs, videos, images, plain text, etc.).
6) **Data storage:** The final stage of data processing is storage. After all of the data is processed, it is then stored for future use. While some information may be put to use immediately, much of it will serve a purpose later on. When data is properly stored, it can be quickly and easily accessed by members of the organization when needed.

6.3 Diagrammatic Presentation Of Data Processing

As you know, diagrammatic presentation is one of the techniques of visual presentation of data. It is a fact that diagrams do not add new meaning to the statistical facts but they reveal the facts of the data more quickly and clearly. Because examining the figures from tables become laborious and uninteresting to the eye and also confusing. Here, it is appropriate to state the words of M. J. Moroney, "cold figures are uninspiring to most people. Diagrams help us to see the pattern and shape of any complex situation." Thus, the data presented through diagrams are the best way of appealing to the mind visually. Hence, diagrams are widely used in practice to display the structure of the data in research work.

- **Rules for Preparing Diagrams**

The prime objective of the diagrammatic presentation of data is to highlight their basic hidden facts and relationships. To ensure that the presentation of numerical data is more attractive and effective, therefore, it is essential to keep the following general rules in mind while adapting diagrams in research work. Now, let us discuss them one by one.

1. You must have noted that the diagrams must be geometrically accurate. Therefore, they should be drawn on the graphic axis i.e., the 'X' axis (horizontal line) and the 'Y' axis (vertical line). However, the diagrams are generally drawn on plain paper after considering the scale.
2. While taking the scale on the 'X' axis and 'Y' axis, you must ensure that the scale showing the values should be in multiples of 2, 5, 10, 20, 50, etc.
3. The scale should be set up, e.g., millions of tons, persons in Lakhs, value in thousands, etc. On the 'Y' axis the scale starts from zero, as the vertical scale is not broken.
4. Every diagram must have a concise and self-explanatory title, which may be written at the top or bottom of the diagram.
5. To draw the readers' attention, diagrams must be attractive and well-proportioned.
6. Different colors or shades should be used to exhibit various components of diagrams and also an index must be provided for identification.
7. It is essential to choose a suitable type of diagram. The selection will depend upon the number of variables, minimum and maximum values, and objects of presentation.

6.4 Graphical Representation Of Data Processing Using Excel

Excel charts are graphical representations of numeric data. Graphs make it easier for users to compare and understand numbers, so charts have become a popular way to present numerical data. Every chart tells a story. Stories can be simple: "See how our sales have increased" or complex: "This is how our overhead costs relate to the price of our product." Whether simple or complex, the story should be readily understandable. If you can't immediately understand what a chart means, then it isn't a good chart.

Graphs are constructed with data points, which are the individual number in a worksheet, and data series, which are the groups of related data points within a column or row. Charts and graphs in Microsoft Excel provide a method to visualize numeric data. While both graphs and charts display sets of data points about one another, charts tend to be more complex, varied, and dynamic. People often use charts and graphs in presentations to give management, client, or team members a quick snapshot of progress or results. You can create a chart or graph to represent nearly any kind of quantitative data — doing so will save you the time and frustration of poring through spreadsheets to find relationships and trends. It's easy to create charts and graphs in Excel, especially since you can also store your data directly in an Excel Workbook, rather than importing data from another program. Excel also has a variety of preset chart and graph types so you can select one that best represents the data relationship(s) you want to highlight. Excel comes with a wide variety of charts capable of graphically representing most standard types of data analysis and even some more exotic numeric interpolations. The type of data you are using and presenting determines the type of chart you will plot the data on.

6.5 Types Of Charts

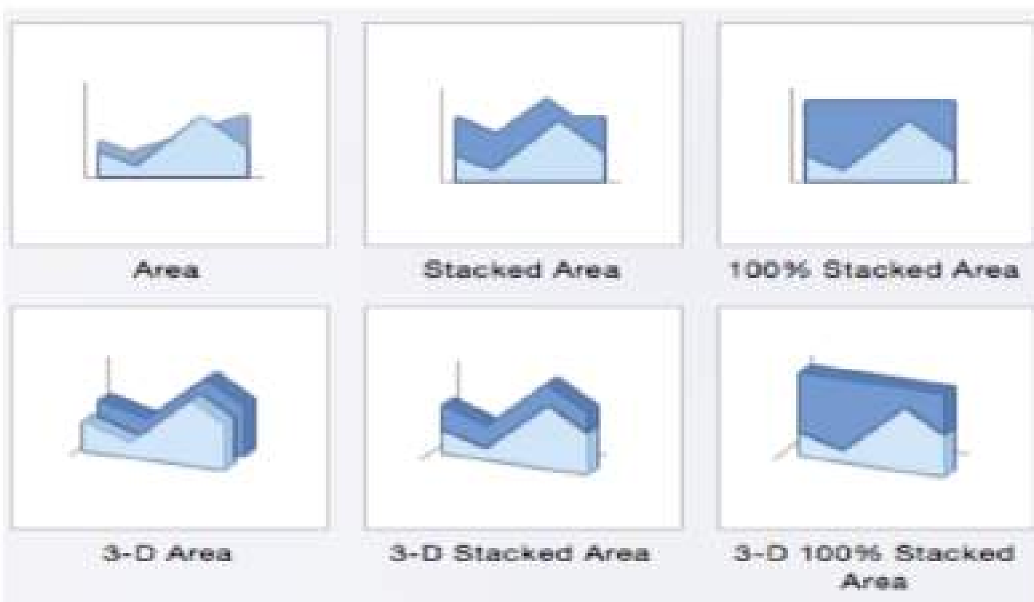
1. Pie Charts: These work best for displaying how much each part contributes to a total value. Pie charts can be exploded for greater visual clarity, or turned into doughnut charts, which can represent more than just one set of data.

- Use pie charts to compare percentages of a whole ("whole" is the total of the values in your data). Each value is represented as a piece of the pie so you can identify the proportions. There are five pie chart types: pie, pie of pie (this breaks out one piece of the pie into another pie to show its sub-category proportions), bar of pie, 3-D pie, and doughnut.

2. Line and area charts: These show data points connected with lines, indicating upward or downward trends in value. Area charts show the area below a line filled in. Both types can be combined with column charts to show more data.

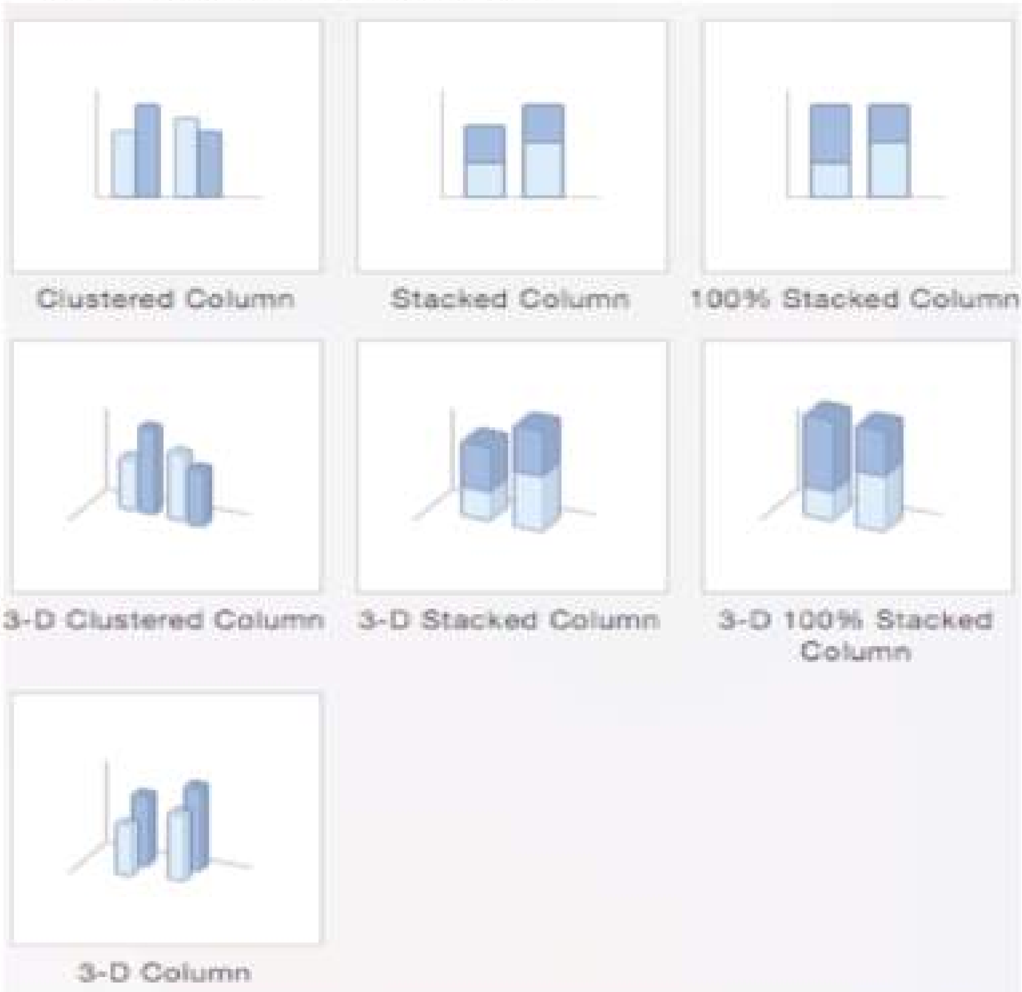
- A line chart is most useful for showing trends over time, rather than static data points. The lines connect each data point so that you can see how the value(s) increased or decreased over some time. The seven-line chart options are line, stacked line, 100% stacked line, line with markers, stacked line with markers, 100% stacked line with markers, and 3-D line.

- **Area:** Like line charts, area charts show changes in values over time. However, because the area beneath each line is solid, area charts are useful to call attention to the differences in change among multiple variables. There are six area charts: area, stacked area, 100% stacked area, 3-D area, 3-D stacked area, and 3-D 100% stacked area.

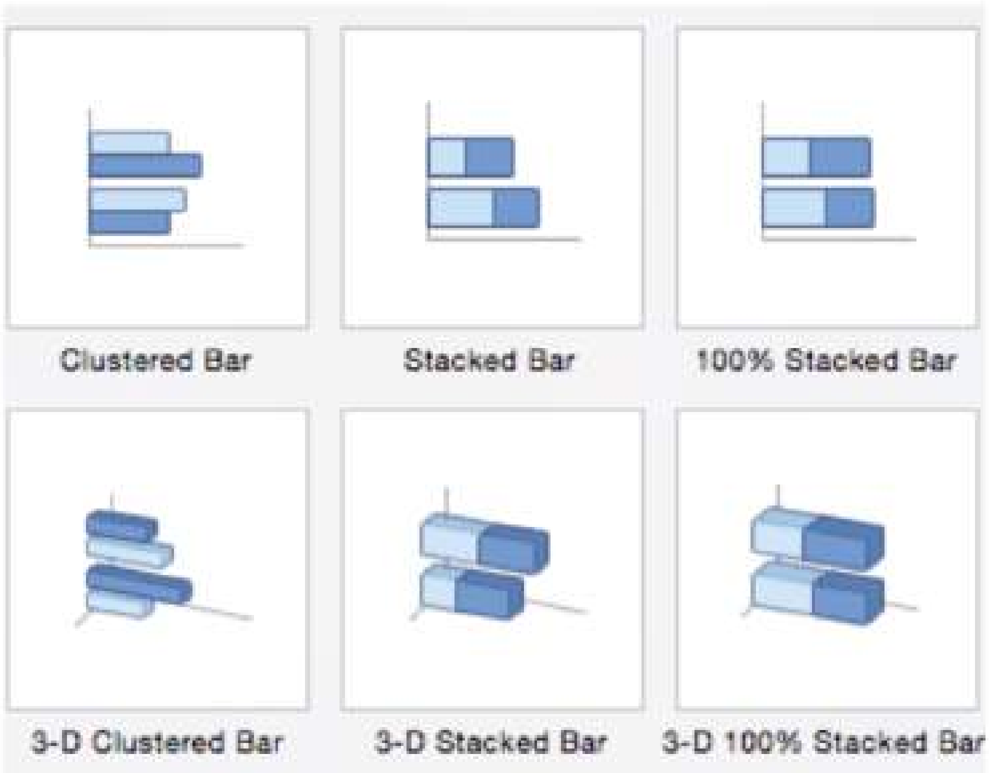


1. **Column and bar charts:** These compare values across categories, with results presented vertically in column charts and horizontally in bar charts, The composition of the column or bar can be stacked in more than one color to represent the contribution of each portion of a category's data to the total for that category.

- **Column Charts:** Some of the most commonly used charts, column charts, are best used to compare information or if you have multiple categories of one variable (for example, multiple products or genres). Excel offers seven different column chart types: clustered, stacked, 100% stacked, 3-D clustered, 3-D stacked, 3-D 100% stacked, and 3-D, pictured below. Pick the visualization that will best tell your data's story.

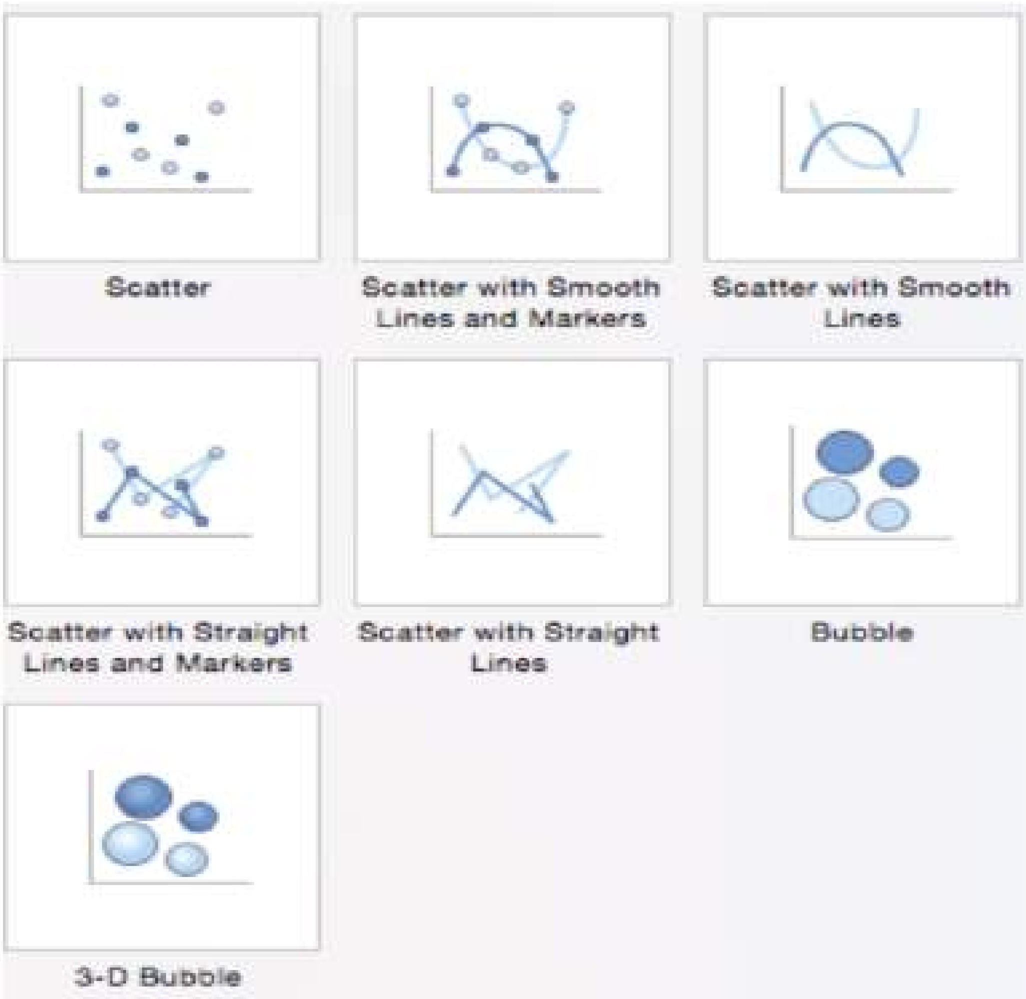


- **Bar Charts:** The main difference between bar charts and column charts is that the bars are horizontal instead of vertical. You can often use bar charts interchangeably with column charts, although some prefer column charts when working with negative values because it is easier to visualize negatives vertically, on a y-axis.

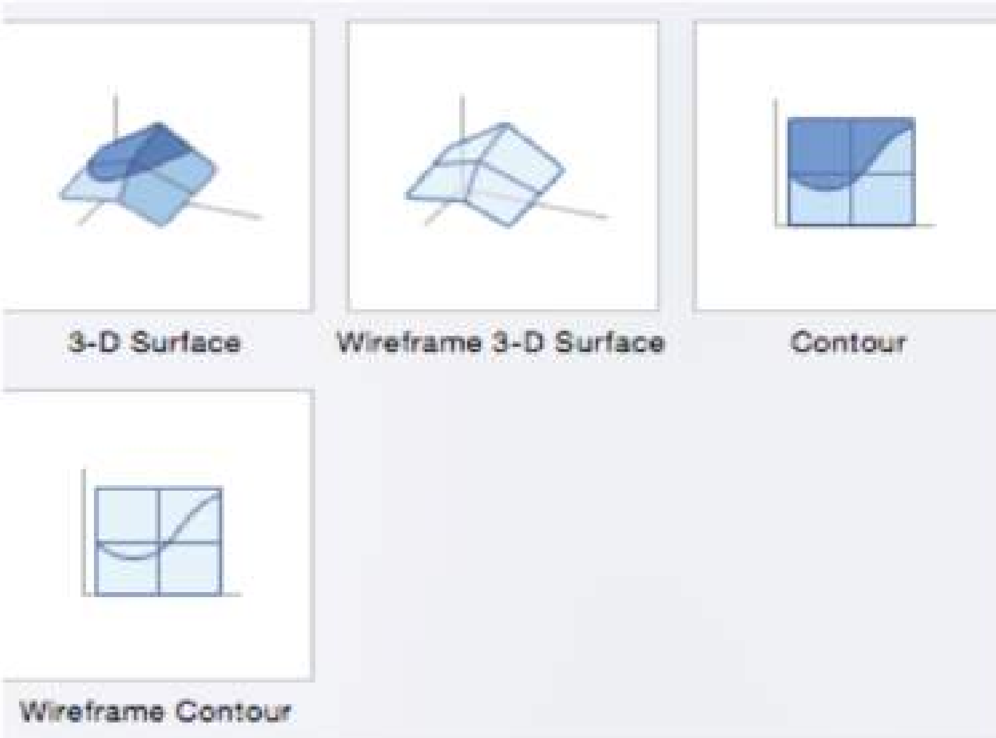


2. **Special charts:** Excel includes several charts suitable for presenting scientific statistical, and financial data. Scatter charts are used to present experimental results. Surface and cone charts are good for presenting 3-D and 2-D changes in data. Radar charts show data values about a single metric. Stock charts present values for between three and five series of data, including open, high, low, close, and volume trading information.

- **Scatter Charts:** Similar to line graphs, because they are useful for showing change in variables over time, scatter charts are used specifically to show how one variable affects another. (This is called correlation.) Note that bubble charts, a popular chart type, are categorized under scatter. There are seven scatter chart options: scatter, scatter with smooth lines and markers, scatter with smooth lines, scatter with straight lines and markers, scatter with straight lines, bubble, and 3-D bubble.



- **Surface:** Use a surface chart to represent data across a 3-D landscape. This additional plane makes them ideal for large data sets, those with more than two variables, or those with categories within a single variable. However, surface charts can be difficult to read, so make sure your audience is familiar with them. You can choose from 3-D surface, wireframe 3-D surface, contour, and wireframe contour.



- **Radar:** When you want to display data from multiple variables about each other use a radar chart. All variables begin from the central point. The key with radar charts is that you are comparing all individual variables about each other — they are often used for comparing the strengths and weaknesses of different products or employees. There are three radar chart types: radar, radar with markers, and filled radar.

1. Pie Charts

Use pie charts to show the relationships between pieces of an entity. The implication is that the pie includes all or something. The pie chart isn't appropriate for illustrating some of anything, so if there's not an obvious "all" in the data you're charting, don't use a pie.

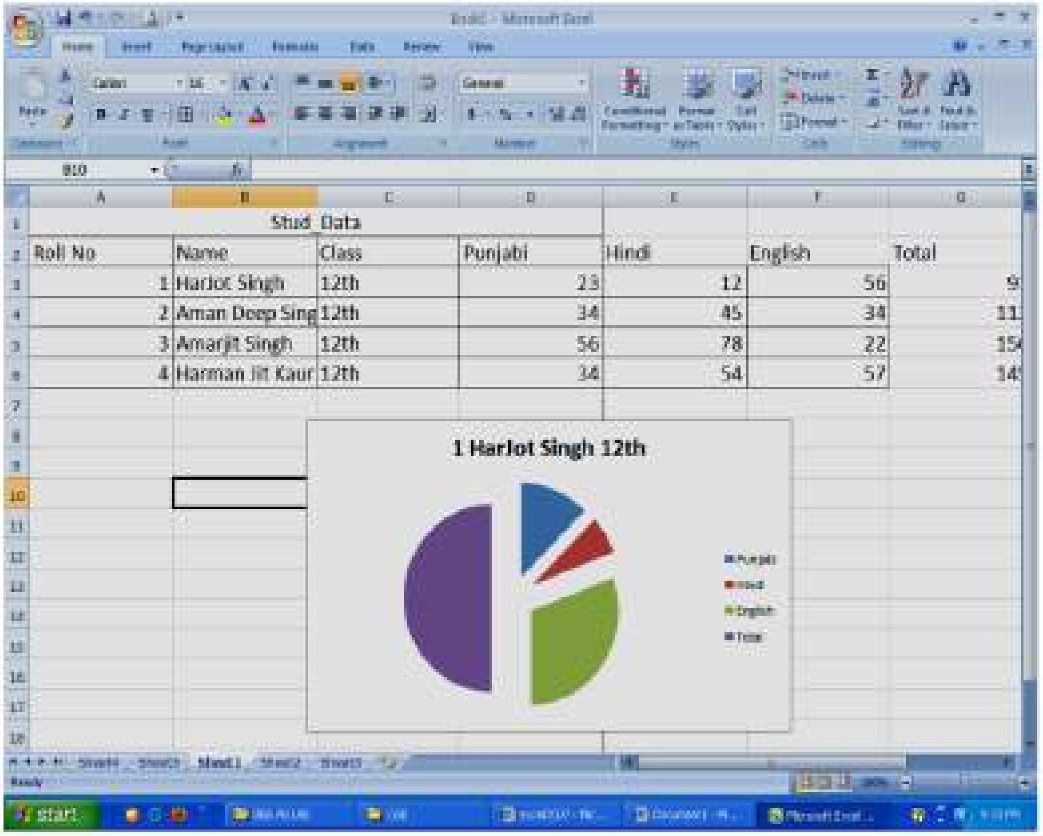


Fig.

A pie chart can only include one data series. If you select more than one data, series, Excel uses the first series and ignores all others. No error message appears, so you won't necessarily know that the chart doesn't show the data you intended to include unless you examine the chart carefully.

When you create a pie chart, Excel totals the data points in the series and then divides the value of each data point into the series total to determine how large each data point's pie slice should be. Don't include a total from the worksheet as a data point; this doubles the total Excei calculates, resulting in a pie chart with one large slice (50 percent of the pie).

2. Line and Area Charts

The series chart shown in the figure is a line chart. In a 2-D version (as shown) or in a 3-D version that is sometimes called a ribbon chart. An area chart is a line chart with the area below the line filled. Line charts and area charts are typically used to show one or more variables (such as sales, income, or price) changing over time.

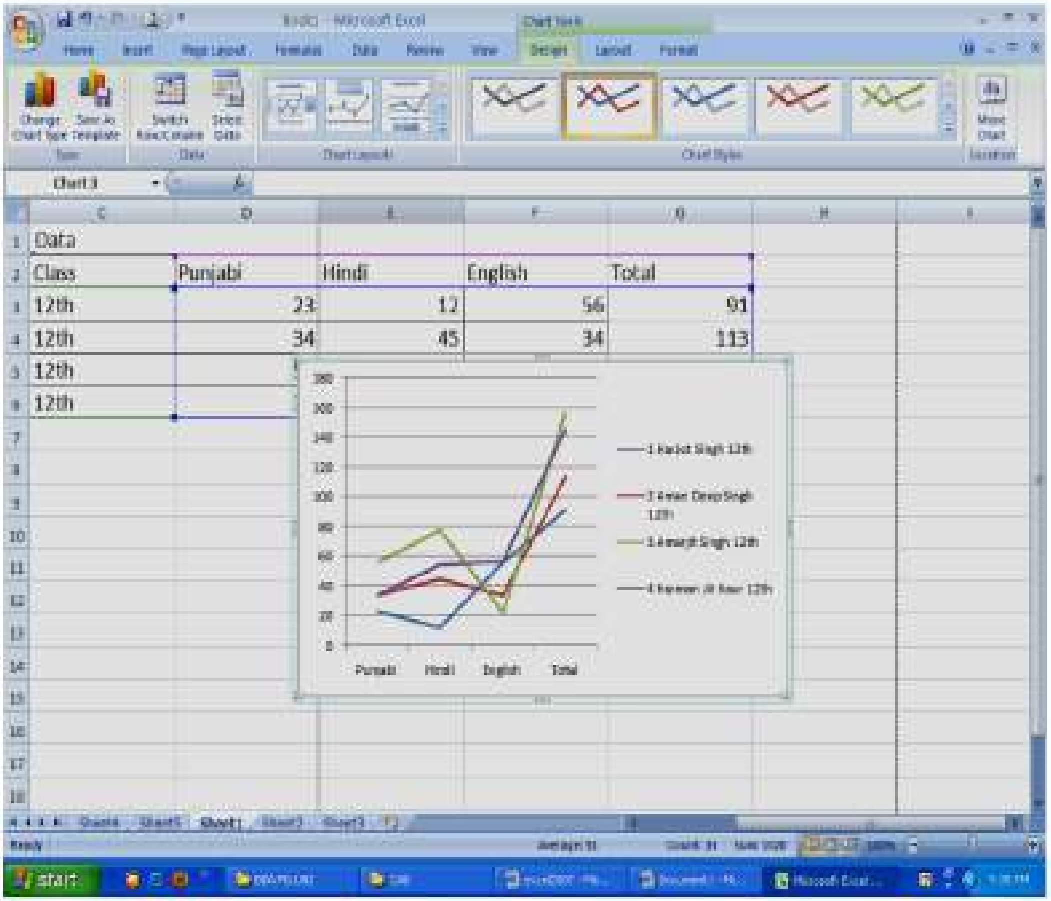


Fig.

3. Column Chart

The figure shows the same information presented as a bar chart. The bars give added substance to the chart. In the line chart, what the reader notices is the trend up or down in each line and the gaps between the lines.

Line and area charts share a common layout. The horizontal line is called the X-axis, and the vertical line is the Y-axis (the same x- and y-axis you may have learned about in algebra or geometry class when plotting data points). In a bar chart, however, the axis is turned 90 degrees so that the x-axis is on the left side.

Excel can also combine columns with line or area charts and embellish line or column charts with 3-D effects. You can make the columns and lines on your charts into tubes, pyramids, cones, or cylinders; or transform regular bars into floating 3-D bars. Plotting data on two axes is also possible with column charts.

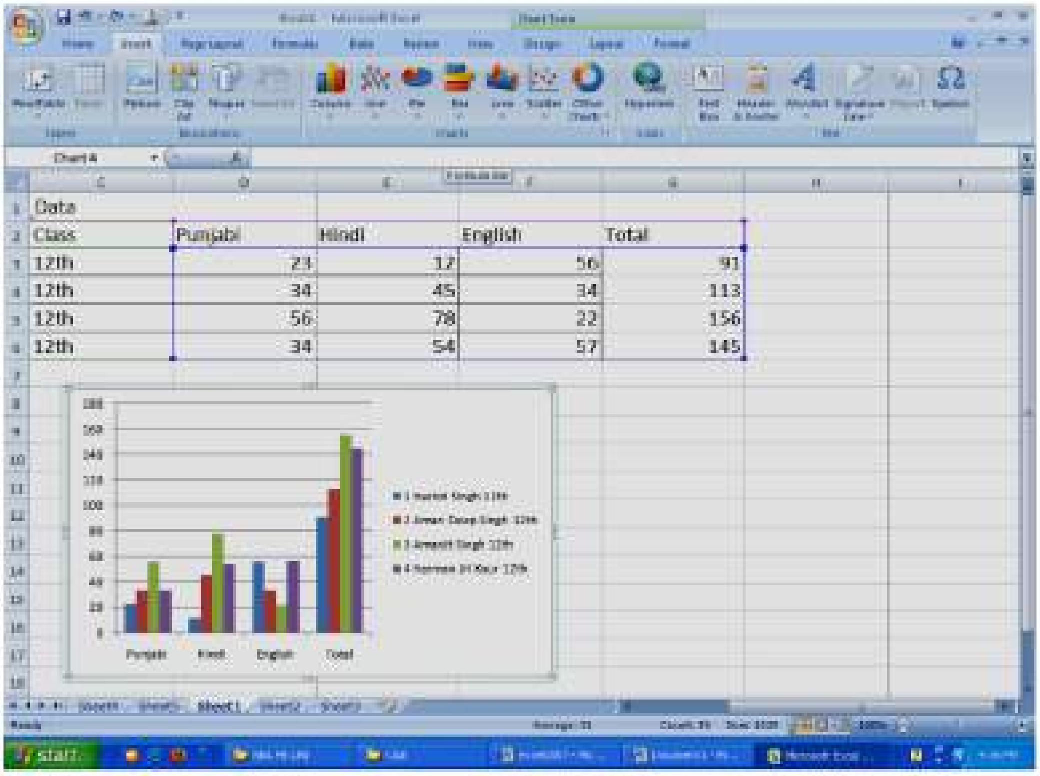


Fig.

4. Bar Chart Variations

Column charts are the same as bar charts but with the X-axis at the bottom. There are three-dimensional varieties of bar and column charts, which add depth to the regular chart. Cylinders, cones, and pyramids are variations of a column chart.

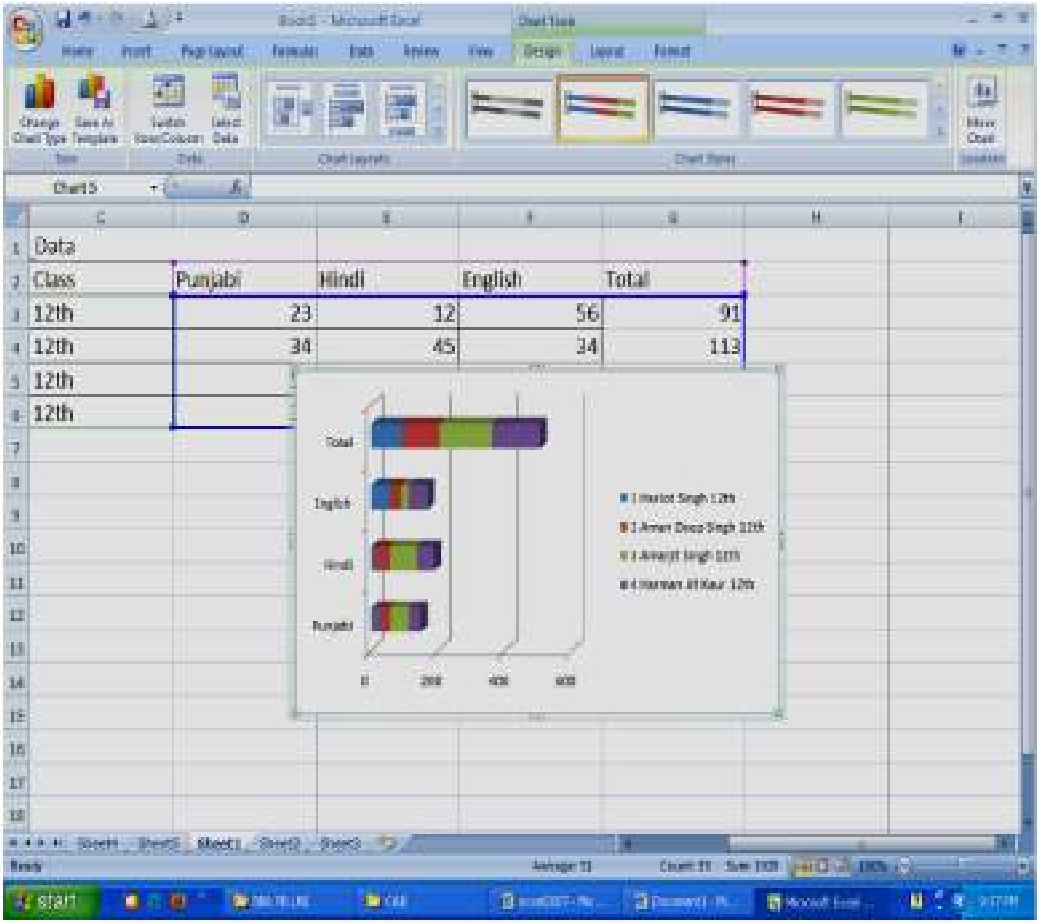


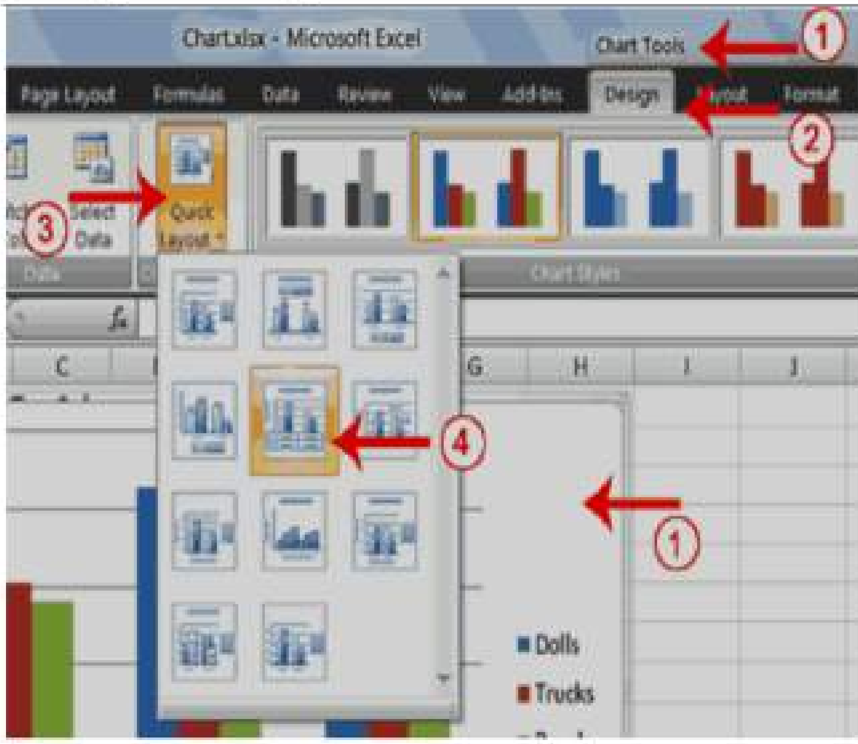
Fig.

Excel also offers another style of bar and column chart–the stacked chart. A stacked 3-D column chart, using the same data as Figure. In a stacked chart, parallel data points in each data series are stacked on top or to the right of each other. Stacking adds another dimension to the chart since it allows the user to compare sales between as well as within time periods-like providing a column chart and a pie chart for each period.

The 3-D charts have three axes. In a 3-D column chart, the X-axis is on the bottom. The vertical axis is the Z-axis; the Y-axis goes from front to back, providing the "third dimension" of depth in the chart. Don't worry about memorizing which axis is which in each chart type; there are ways to know which is which when you're creating or editing the chart.

6.6 Apply Chart Layout

Context tabs are tabs that only appear when you need them. Called Chart Tools, there are three chart context tabs: Design, Layout, and Format. The tabs become available when you create a new chart or when you click on a chart. You can use these tabs to customize your chart. You can determine what your chart displays by choosing a layout. For example, the layout you choose determines whether your chart displays a title, where the title displays, whether your chart has a legend, where the legend displays, whether the chart has axis labels, and so on. Excel provides several layouts from which you can choose.



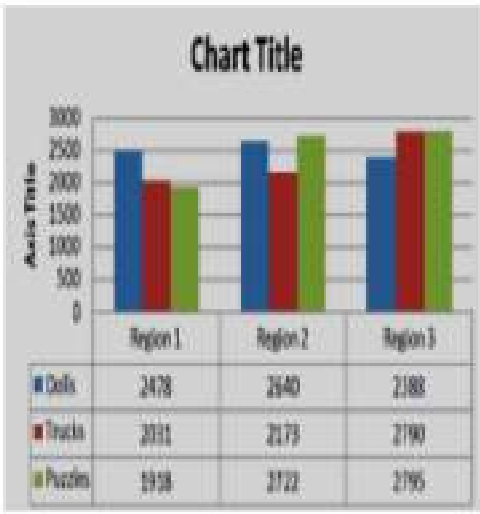
Fig

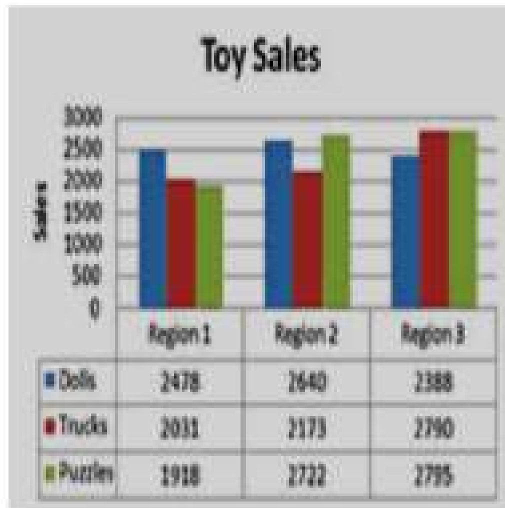
Steps to Apply a Chart Layout

1. Click your chart. The Chart Tools become available.
2. Choose the Design tab.
3. Click the Quick Layout button in the Chart Layout group. A list of chart layouts appears.
4. Click Layout. Excel applies the layout to your chart.

6.7 Add Labels

When you apply a layout, Excel may create areas where you can insert labels. You use labels to give your chart a title or to label your axes. When you applied layout, Excel created label areas for a title and the vertical axis





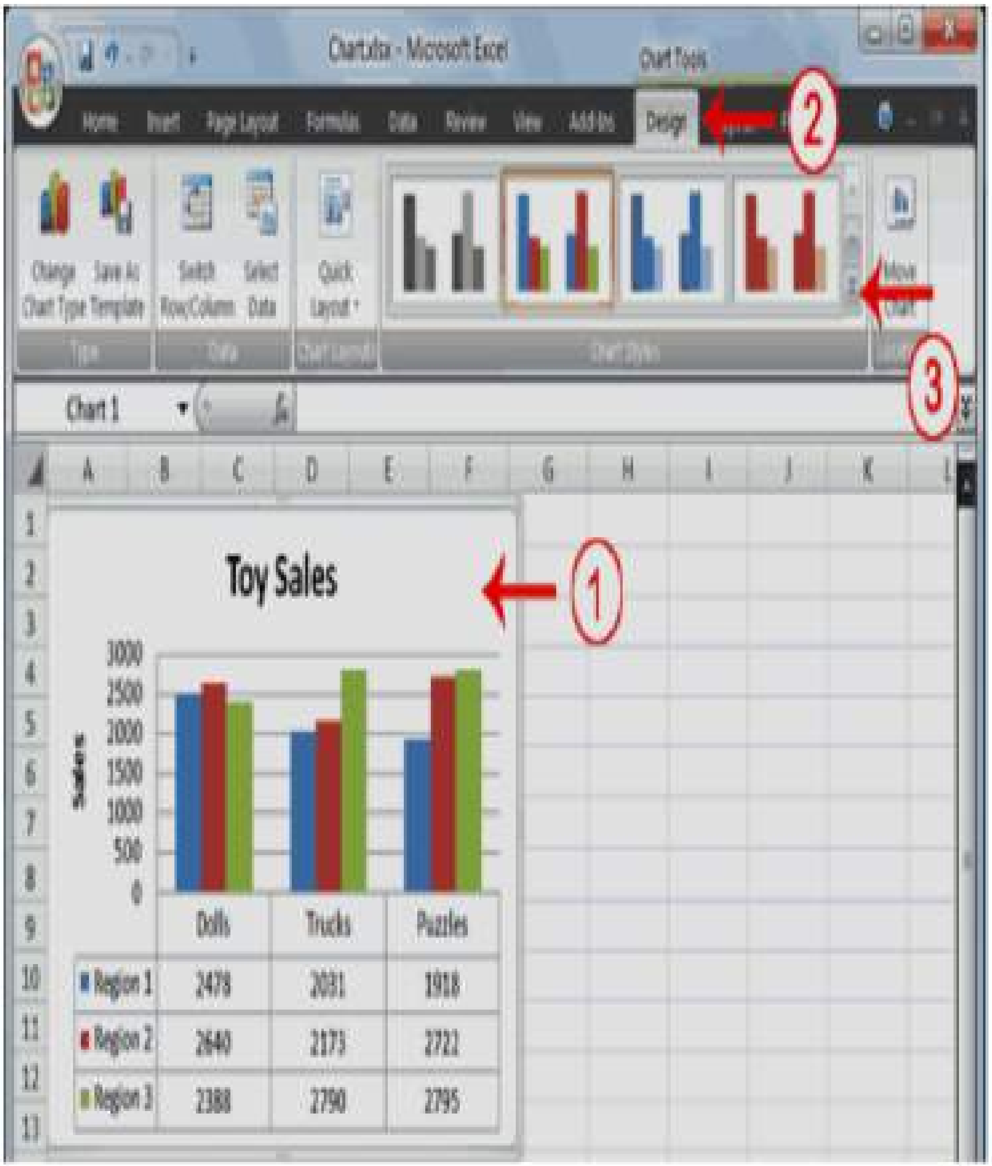
Before **Fig After**

Steps to add labels

1. Select Chart Title. Click on Chart Title and then place your cursor before the C in Chart and hold down the Shift key while you use the right arrow key to highlight the words Chart Title.
2. Type **Toy Sales**. Excel adds your title.
3. Select Axis Title. Click on Axis Title. Place your cursor before the A in Axis. Hold down the Shift key while you use the right arrow key to highlight the words, Axis Title.
4. Type Sales**.** Excel labels the axis.
5. Click anywhere on the chart to end your entry.

6.8 Change The Style Of A Chart

A style is a set of formatting options. You can use a style to change the color and format of your chart. Excel has several predefined styles that you can use. They are numbered from left to right, starting with 1, which is located in the upper-left corner.



	Dolls	Trucks	Puzzles
Region 1	2478	2031	1918
Region 2	2640	2173	2722
Region 3	2388	2790	2795

Fig.

Steps to Change the Style of a Chart

1. Click your chart. The Chart Tools become available.
2. Choose the Design tab.
3. Click the More button in the Chart Styles group. The chart styles appear.

Fig

4. Click Style. Excel applies the style to your chart.

6.9 Data Preservation

Data preservation is the act of conserving and maintaining both the safety and integrity of data. Preservation is done through formal activities that are governed by policies, regulations, and strategies directed towards protecting and prolonging the existence and authenticity of data. Data preservation refers to maintaining access to data and files over time. For data to be preserved, at minimum, it must be stored in a secure location, stored across multiple locations, and saved in file formats that will likely have the greatest utility in the future. Data preservation provides the usability of data beyond the lifetime of the research activity that generated them.

Definition

Data preservation consists of a series of managed activities necessary to ensure continued stability and access to data for as long as necessary. For data to be preserved, at minimum, it must be stored in a secure location, stored across multiple locations, and saved in open file formats that will likely have the greatest utility in the future. Part of the preservation process can include depositing data in an institutional, discipline-specific, or generalist data repository, all of which allow for publication and preservation.

- The new NIH Data Management and Sharing Policy requires data to be preserved and shared, so medical researcher submits their COVID data to the National COVID Cohort Collaborative (N3C), as listed in the Open Domain-Specific Data Sharing Repository.
- There are many different ways to preserve digital information, including text, photos, audio, and video. Some strategies to preserve digital information include: Refreshing: transferring data to the same format. An example would be the transfer of music from an old CD-ROM to a new CD-ROM.

Digital information is an important source in our knowledge economy, valuable for research and education, science and the humanities, creative and cultural activities, and public policy. New high-throughput instruments, telescopes, satellites, accelerators, supercomputers, sensor networks, and running simulations are generating massive amounts of data. These data are used by decision-makers to improve the quality of life of citizens. Moreover, researchers are employing sophisticated technologies to analyze these data to address questions that were unapproachable just a few years ago. Digital technologies have fostered a new world of research characterized by immense datasets, unprecedented levels of openness among researchers, and new connections among researchers, policymakers, and the public domain. Different types of threats are:

1. Users may be unable to understand or use the data,

2. Lack of sustainable hardware, software, or support of the computer environment may make the information inaccessible,

3. Evidence may be lost because the origin and authenticity of the data may be uncertain,

4. Access and use restrictions (e.g., Digital Rights Management) may not be respected in the future,

5. Loss of ability to identify the location of data,

6. The current custodian of the data, whether an organization or project, may cease to exist at some point in the future.

- **Importance of Data Preservation**

Preservation helps protect you from hardware obsolescence. You never want to find yourself in a situation where all of your data is saved on unsupported hardware! Always migrate to new hardware formats so that your data will be available long-term. **The most responsible way to preserve your data is to turn it over to a responsible custodian such as a data repository.** When possible, try to preserve research data in a repository that provides *data curation services*, not just preservation services. Curated data is more valuable, easier to reuse, easier to locate, and more highly cited. Many data repositories have requirements for deposit - they may only accept certain types of data and have file size limits.

6.10 Data Preserving Vs. Storing Data

Preserving is different from storing and backing up data files while your research is still ongoing. The latter typically involves mutable data; the former concerns data (or milestone versions of data) that are 'frozen' and not in active use. Long-term preservation requires appropriate actions to prevent data from becoming unavailable and unusable over time, for example, because of:

- Outdated software or hardware
- Storage media degradation
- A lack of sufficient descriptive and contextual information to keep data understandable

In other words, data preservation involves more than simply not deleting the data files created and stored. Maintaining data in a usable form for the longer term takes effort and has a considerable cost. Selecting which (parts of) data to keep, and for how long, is, therefore, an essential component of data preservation.

As a researcher, you have a key role in deciding what to retain and what not, as you know your data best. Such decisions may depend on factors such as:

- The type of data involved

- The norms in your discipline
- Whether you are keeping data for potential future reuse, verification, or other purposes. Depending on the purpose, you may need to keep the raw data or data in a more processed form.

6.11 DATA PRESERVATION VS. RETENTION OF DATA

- Data retention is a central component of records management and information governance. Retention refers to the storing of data to meet regulatory and recordkeeping obligations
- The preservation is related to the safekeeping of electronically stored information (ESI) for some anticipated legal matter. In other words, data retention is a proactive ongoing process.
- Retention is usually a mandated requirement for researchers - it's the task that ensures that a bare minimum of data will remain available in some format.
- Preservation refers to having an active plan to ensure that when you do need to access your old data, it's readily available and can be easily accessed and manipulated by whoever needs it. When making a plan for data preservation you should include activities such as:
- Transferring data from older storage formats to newer ones. This will ensure that the technology required to access your data is still available.
- Transferring data from older file formats to newer ones. This will ensure that your data can still be opened by current software applications.
- Having multiple copies of the data in different locations. This will ensure that your data is not lost in an unexpected event, such as theft, flood, or fire.
- Ensuring your data is well documented, such as making notes on software used for creating codebooks as outlined in the Documentation and Description section of this guide. This will ensure that when you come back to access your data, you'll be able to remember what it all means.

6.11 Questions For Practice

1. Define data processing. What are the various stages of data processing?
2. What is the purpose of data processing?
3. What is the significance of data processing in the research?
4. Explain the various rules for data processing.
5. What is the role of Excel in the presentation of data processing?
6. What is chart layout? How it can be applied in Excel?

7. What is the utility of the style of the chart in data processing? How it can be applied?
8. Define data preservation. What are the various challenges and threats to the digital preservation of data?
9. Explain data preservation Vs. storage of data.
10. Explain in which situation various types of charts are used in data processing:
    1. Pie chart
    2. Bar chart
    3. Line and area Chart
    4. Surface chart
    5. Radar chart

6.12 Suggested Readings

- A. Abebe, J. Daniels, J.W. Mckean, "Statistics and Data Analysis".
- Clarke, G.M. & Cooke, D., "A Basic Course in Statistics", Arnold.
- David M. Lane, "Introduction to Statistics".
- S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", SultanChand & Sons, New Delhi.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 7: SAMPLE, POPULATION, CHARACTERISTICS OF GOOD SAMPLE, TYPE OF SAMPLING TECHNIQUES, SAMPLING ERRORS

STRUCTURE

7.0 Objective Learnings

After studying this unit, you should be able to:

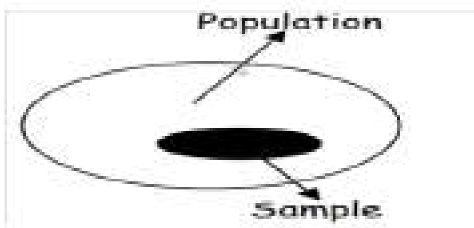
- explain the concepts of sample and population

- define the characteristics of a good sample
- about types of sampling
- explain different sampling techniques

7.1 INTRODUCTION SAMPLE AND POPULATION

The total of items about which information is desired is called population or Universe. Which is further classified into two categories i.e., finite and infinite. A finite population consists of a fixed number of elements so that it is possible to count in its totality. A suitable example of a finite population is the population of a city, the number of students in the class, the number of workers in a factory, etc. While an infinite population is a population in which it is theoretically impossible to observe or count all the elements. In the case of an infinite population, the number of items is infinite. A suitable example of an infinite population is the number of stars in the sky and some hairs in the head. The use of the term infinite population for a population that cannot be counted in a reasonable period of time.

In statistics, population is the aggregate of objects, animate or inanimate under study in any statistical investigation.



A sample is a finite subset of the population that is selected from it to analyze its properties, and the number of units in the sample is known as the sample size. By examining only, the objects or items that are part of the sample, sampling is a method that enables us to make inferences about the features of the population. So, the sample is a part of the population that represents the characteristics of the population, while Sampling is the process of selecting the sample for estimating the population characteristics. It is the process of obtaining information about an entire population by examining only a part of it. The main objectives of the sampling theory are

1. To obtain the optimum results, i.e., the maximum information about the characteristics of the population with the available sources at our disposal in terms of time, money and manpower by studying the sample values only.
2. To get the most accurate population parameter estimates.

7.2 PURPOSE OF SAMPLING

Providing an estimate of the population parameter and testing the hypothesis are the two main purposes of sampling. In order to infer information about the entire population and draw conclusions about it, we are gathering data from a smaller subset of a larger population.

- Through sampling a researcher can able to collect data more efficiently and cost-effectively compared to studying the entire population. It is often unfeasible, expensive, and time-consuming to survey or watch every person in a population. Researchers can save money by choosing a representative sample while still obtaining useful information.
- Enable collection of comprehensive data.
- Enable more accurate measurement as it is conducted by trained and experienced investigators.
- Sampling remains the only way when the population contains infinitely many members.
- In certain situations, sampling is the only way of data collection. For example, in testing the pathological status of blood, boiling status of rice, etc.
- Sampling enables statistical analysis and population parameter forecast. Researchers can use statistical methods to analyse data from a representative sample, estimate population characteristics, and compute confidence intervals or margins of error. This offers a degree of accuracy and permits drawing conclusions about the population.
- To improve methodologies, processes, or protocols, sampling is frequently employed during pilot testing or pre-testing phases of the study. Before beginning the larger study, researchers can find any problems or difficulties by gathering data from a smaller sample first. By doing this, the research's validity and quality are enhanced.
- It provides a valid estimation of sampling error.

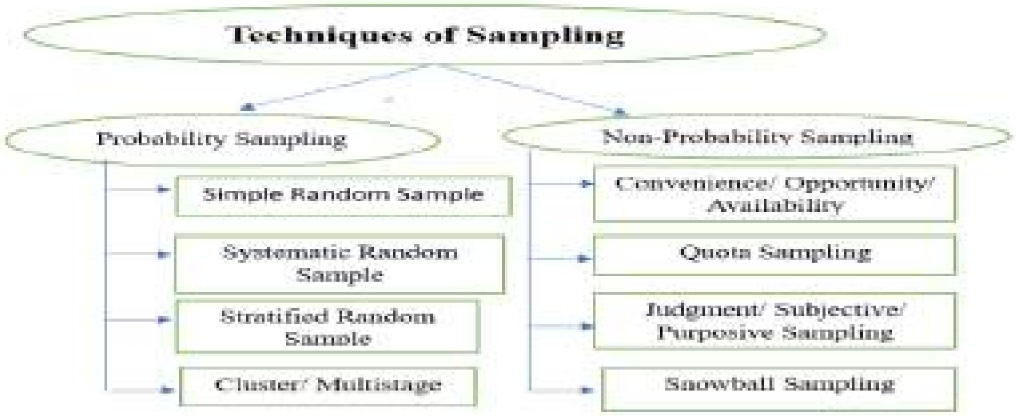
7.3 Characteristics of Good Sampling

The process of gathering data from a selected group or sample of the population that is relevant to a research study is known as sample size. Researchers get data from a manageable group that is representative of the whole population rather than the complete population and solve the objective or purpose of sample collection. The characteristics of the good sample, which is chosen from the population, are as:

- **Good Representativeness:** The sample should represent the population that serves the purpose. It should represent a large number of the population. This makes it possible to apply the sample's results to the entire population.
- **Sampling Method:** Common sampling techniques include random sampling, stratified sampling, cluster sampling, and convenience sampling. The choice of sampling method depends on the research objectives, available resources, and the desired level of representativeness.
- **Sample Size**: The number of observations that make up the sample is referred to the sample size. Choosing the right sample size includes the research objectives, desired level of accuracy, anticipated population variability, and required statistical power. Even though a large sample size typically yields accurate and reliable results, but it might require more resources.
- **Sampling Frame**: A list or framework known as a sampling frame identifies the population from which the sample will be taken. It serves as the basis to selecting the sample and, from the population. To ensure the sample is representative, the sampling frame needs to be comprehensive and relevant.
- **Data Collection from the Sample**: Data is gathered from the individuals or observations that have been chosen after the sample. Various collecting information techniques, including surveys, interviews, observations, and research, may be used in this. The study objectives and the characteristics of the variables being measured should be consistent with the method that is to be chosen.
- **Sampling Bias:** Sampling bias is the systematic inaccuracy or misrepresentation in the sample that happens when some people or groups are either overrepresented or represented. here, the validity and generalizability of the research findings may be affected. To minimize the biases during the sampling process, researchers need to be aware of them.

Sample measurement plays a crucial role in research as it allows researchers to collect data efficiently and make inferences about the larger population. Careful consideration of the sampling method, sample size, and representativeness is important to ensure the reliability and validity of the research findings.

7.4 Types/ Procedures/ Methods/ Techniques Of Sampling



There are two basic approaches to sampling:

- Probability Sampling
- Non-probability Sampling

7.4.1 Probability Sampling

Probability sampling is also known as random sampling or chance sampling. In this, sample is taken in such a manner that every unit of the population has an equal and positive chance of being selected.

In this way, it is ensured that the sample would truly represent the overall population. Probability sampling can be achieved by random selection of the sample among all the units of the population. Major random sampling procedures are:

- Simple Random Sample
- Systematic Random Sample
- Stratified Random Sample
- Cluster/ Multistage Sample

7.4.1.1 Simple Random Sample: For this, each member of the population is numbered. Then, a given size of the sample is drawn with the help of a random number chart. The other way is to do a lottery. Write all the numbers on small, uniform pieces of paper, fold the papers, put them in a container, and take out the required lot in a random manner from the container as is done in the kitty parties. It is relatively simple to implement but the final sample may miss out on small sub groups.

7.4.1.2 Systematic Random Sample: It also requires numbering the entire population. Then every nth number (say every 5th or 10th number, as the case may be) is selected to constitute the sample. It is easier and more likely to represent different subgroups.

The sampling interval (k) is calculated by dividing the total population size (N) by the desired sample size (n). Mathematically, k = N / n, where "N" is the population size, and "n" is the sample size. To introduce randomness into the sample selection process, you randomly select a starting point within the population. You can use random number tables, computer-generated random numbers, or other randomization techniques to pick the initial element. After that select the Sample, once you have the sampling interval (k) and a random starting point, you select every "k^th" element in the population as part of your sample until you reach the desired sample size (n). If you reach the end of the population, you wrap around and continue the selection process until you complete the sample. The primary advantage of systematic random sampling is that it can generate a representative sample from a large population and is usually simple to apply. To prevent biases in the sample, it is essential to be sure the population is sufficiently randomized before choosing its starting point.

5.4.1.3 Stratified Random Sample: At first, the population is divided into groups or strata each of which is homogeneous concerning the given characteristic feature. From each strata, then, samples are drawn at random. This is called stratified random sampling. For example, for the level of socio-economic status, the population may first be grouped in such strata as high, middle, low and very low socio-economic levels as per pre-determined criteria, and a random sample drawn from each group. The sample size for each sub-group can be fixed to get a representative sample. This way, it is possible that different categories in the population are fairly represented in the sample, which could have been left out otherwise in a simple random sample.

As with stratified samples, the population is broken down into different categories. However, the size of the sample of each category does not reflect the population as a whole. The Quota sampling technique can be used where an unrepresentative sample is desirable (e.g., you might want to interview more children than adults for a survey on computer games), or where it would be too difficult to undertake a stratified sample.

7.4.1.4 Cluster/ Multistage Sample: In some cases, the selection of units may pass through various stages, before you finally reach your sample of study. For this, a State, for example, may be divided into districts, districts into blocks, blocks into villages, and villages into identifiable groups of people, and then the random or quota sample from each group. For example, taking a random selection of 3 out of 15 districts of a State, 6 blocks from each selected district, 10 villages from each selected block and 20 households from each selected village, totaling 3600 respondents. This design is used for large-scale surveys spread over large areas. The advantage is that it needs a detailed sampling frame for selected clusters only rather than for the entire target area. There are savings in travel costs and time as well. However, there is a risk of missing important sub-groups and not having a complete representation of the target population.

CHECK YOUR PROGRESS (A)

Q1. What is sampling?

Ans:____

Q2. Define Simple random sample with an example.

Ans:____

Q3. Explain systematic random sample.

Ans:____

Q4. What is stratified random sampling?

Ans:____

Q5. Define Cluster/ Multistage sampling.

Ans:____

7.4.2 Non-Probability Sampling

Non-probability sampling is any sampling technique where certain population categories have no possibility of selection, i.e., are not covered, or when it is difficult to calculate the probability of selection. It entails choosing components based on presumptions about the population of interest, which serves as the selection criteria. Non-probability sampling prevents the estimate of sampling errors since the selection of elements is not random. Non-probability sampling is a non-random and subjective sampling technique in which the sampler's discretion or personal judgment is used to choose the population elements that will make up the sample.

Non-probability sampling includes:

- Convenience/ Opportunity/ Availability sampling
- Quota Sampling
- Judgment/ Subjective/ Purposive Sampling
- Snowball Sampling

7.4.2.1 Convenience/ Opportunity/ Availability Sampling: Convenience sampling (or opportunity sampling) is a type of non-probability sampling that involves the sample being drawn from that part of the population that is approachable. That is, a sample population is selected because it is readily available and convenient. Such a sample would not be sufficiently representative for the researcher to draw any scientific conclusions about the entire population from it. As an example, if the interviewer were to conduct such a survey at a shopping center early in the morning on a particular day, the people that could be interviewed would be restricted to those present there at that time, which would not represent the views of other members of society in such an area. Pilot testing benefits the most from this kind of sample. This type of sampling is most useful for pilot testing. In convenience sampling, the major problem is that one can never be certain what population the participants in the study represent. The population is unknown, the method for selecting cases is random, and the cases studied probably don't represent any population you could come up with. However, there are some situations in which this kind of design has advantages - for example, survey designers often want to have some people respond to their survey before it is given out in the 'real' research setting as a way of making certain the questions make sense to respondents. For this purpose, availability sampling is not a bad way to get a group to take a survey, though in this case researchers care less about the specific responses given than whether the instrument is confusing or makes people feel bad.

7.4.2.2 Quota Sampling: A non-random sampling, which includes dividing the population into predetermined categories or quotas according to specific criteria, and then choosing those who fit those quotas. When obtaining a random sample is challenging or when access to a sampling frame is constrained, quota sampling is frequently used. In quota sampling, the population is first segmented into mutually exclusive sub-groups, just as similar to stratified sampling. These categorizations can be the demographic basis (e.g., age, gender, education level) or behavioral (e.g., purchasing habits, brand preference). Then judgment is used to select the subjects or units from each segment based on a specified proportion. For instance, a sample of 300 men and 200 women between the ages of 45 and 60 may be asked of the interviewer. The selection of the sample in quota sampling is not random. Interviewers might be drawn to individuals who appear to be most helpful, for instance. Because not everyone is given the opportunity to be chosen, the issue is that these samples may be biased. Its biggest drawback is this random component, and the relative merits of quota and probability have long been a source of debate.

7.4.2.3 Subjective or Purposive or Judgment Sampling: In this technique, the sample is selected with a pre-definite purpose in view and the choice of the sampling units depends entirely on the judgment of the researcher. This sampling suffers from drawbacks of favoritism and partiality depending upon the beliefs and prejudgments of the researcher and does not give a representative sample of the population. This method is rarely used and cannot be recommended for general as based on bias due to an element of subjectivity on the part of the researcher. However, judgment samples may produce useful results if the investigator is knowledgeable and talented and this sampling is used based on judgment.

5.4.2.4 Snowball Sampling: snowball sampling is a non-probability sampling used in research to identify and participate people by way of suggestions from original participants. When investigating difficult-to-reach population or when there isn't a complete list of the target population easily accessible, it is frequently used. Studying social networks, secret populations, or sensitive topics makes good use of snowball sampling.

This sampling technique is used against rare populations. Sampling is a big problem in this case, as the defined population from which the sample can be drawn is not available.

The process of snowball sampling typically first, identifies an initial participant. This individual is usually well-connected within the target population and can help in the recruitment process. After that, a researcher conducts an interview or data collection with the initial participant. Afterward, the participant is asked to provide referrals to others who meet the study's criteria or have relevant experiences or characteristics. Therefore, it creates a referral chain, where newly identified participants become part of the sample. The snowball sampling process continues iteratively until the desired sample size or data fullness is reached. Therefore, the process sampling depends on the chain system of referrals. Although small sample sizes are the clear advantages of snowball sampling, bias is one of its disadvantages.

7.5 Sampling And Non-Sampling Errors

Statistically, "error" refers to the discrepancy between the true value and the estimated or approximate value. In other words, error refers to the difference between the true value of a population parameter and its estimate provided by an appropriate sample statistic computed by some statistical device. As a result, the term "error" has a very specific and different meaning in the field of statistics. The existence of these errors the various factors like approximations in measurements, rounding, biases because of incorrect data collection and analysis and personal biases of the researcher. Errors are of two types:

- Sampling Errors
- Non-sampling Errors

Sampling error: This type of error arises due to the variability in the process of selecting a sample from a larger population. As only a small population is measured, so the results are different from the census. Errors attributed to fluctuations of sampling are called sampling errors. Reasons for sampling errors are:

- Incorrect sample selection
- Biasness in the estimation method
- Heterogeneity of the population
- Sample size, smaller samples are more susceptible to larger sampling errors, while larger samples tend to have smaller sampling errors, assuming random sampling
- Method of sampling
- Non-Response Bias, when individuals from the sample do not respond to the survey
- Measurement Error, arises from inaccuracies in data collection or measurement

Non-sampling Errors: Non-sampling errors are errors that occur in the data collection methods other than those caused by the sampling procedure. These can be because of various sources and can impact the quality as well as accuracy of the collected data. Non-sampling errors can be attributed to factors such as data collection, data processing, measurement, and analysis.

Reasons for non-sampling errors are:

- Respondent Errors, because of misunderstanding on the part of respondent.
- Non-response errors, bias happens when certain individuals selected for the sample do not participate or respond to the survey.
- Processing Errors, occur during data entry, coding, or data cleaning. Human errors as well as software bugs can lead to data inaccuracies during these stages.

- Data Editing Errors, occur when data is modified during the data cleaning process. While data cleaning is essential to ensure data quality, some of the errors introduce inaccuracies.
- Compiling and publishing errors.

Check Your Progress (B)

Q1. What are the types of non-probability sampling?

Ans:____

____

Q2. Define convenience sampling.

Ans:____

____

Q3. What is Quota Sampling?

Ans:____

____

Q4: Define Judgment Sampling.

Ans:____

____

Q5. What is Snowball Sampling?

Ans:____

____

Q6. Errors in sampling.

Ans:____

____

7.6 Sum Up

A sample is a part of the population that represents the characteristics of the population, while Sampling is the process of selecting the sample for estimating the population characteristics. It is the process of obtaining information about an entire population by examining only a part of it. There are different sampling techniques available in statistics depending on the research. These are probability and non-probability sampling techniques. Probability sampling techniques are Simple Random Sample, Systematic Random Sample, Stratified Random Sample, and Cluster/ Multistage Sample. While non-probability sampling is Convenience/ Opportunity/ Availability sampling, Quota Sampling, Judgment/ Subjective/ Purposive Sampling, and Snowball Sampling. There are two types of error in the sample which are sampling and non-sampling error. Sampling error is the error that arises due to the variability in the process of selecting a sample from a larger population. As only a small population is measured, so the results are obvious. The non-sampling error is because of various sources and can impact the quality as well as the accuracy of the collected data. Non-sampling errors can be attributed to factors such as data collection, data processing, measurement, and analysis.

7.7 Questions For Practice

Q1. What do you mean by sample and population? Explain with an example.

Q2. What are the techniques for the collection of data available in statistics?

Q3. What do you mean by primary data? Give the sources of primary data.

Q4. What is the questioner? What are the points to keep in mind before drafting questioner?

Q5. Explain the term secondary data with its sources.

Q6. Give limitations of primary data and secondary data.

Q7. what are the precautions to collect secondary data?

Q8. Explain sampling and non-sampling errors.

7.8 Suggested Readings

- A. Abebe, J. Daniels, J.W. Mckean, "Statistics and Data Analysis".
- Clarke, G.M. & Cooke, D., "A Basic Course in Statistics", Arnold.
- David M. Lane, "Introduction to Statistics".
- S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", SultanChand & Sons, New Delhi.

M.A (ECONOMICS)

SEMESTER I

QUANTITATIVE METHODS I

UNIT 8: INDEX NUMBERS: MEANING AND USES AND TYPES OF INDEX NUMBERS, PROBLEMS IN THE CONSTRUCTION, METHODS OF INDEX NUMBERS

STRUCTURE

8.0 Learning Objectives

After studying the Unit, the learner will be able to learn about:

- Meaning of Index Numbers
- Uses of Index Numbers
- Understand how index numbers are prepared
- Problems to construct Index Numbers
- Different methods to construct Index Number

7.1 Introduction And Meaning Of Index Number

Human life is dynamic and hardly there is anything that remains the same over a period of time, whether it is the price of goods, Population of the country, Industrial Production, imports and Exports of the country, everything changes with the passage of time. It is the tendency of humans that he wants to measure the changes that are taking place over a period of time. Now questions arise about how we can measure these changes that are taking place. Index number is one such statistical tool that can help us in measuring these changes.

An index number is a statistical tool that measures the changes in the data over a period of time. Index number is not a new tool used in statistics, rather the use of index numbers is very old. As per available records, the index number was first time constructed in the year 1764 by an Italian named Carli. In his index number, Carli compared the prices of the Year 1750 with the price level of the year 1500. Though normally index numbers are used for measuring the change in price over a period of time, hardly there is any area in Economics or Commerce where Index numbers are not used. There are different types of index numbers that are used in economics such as Industrial Production Index, Agricultural Production Index and Population Index, etc.

An index number is a device with the help of which we can measure the relative change in one variable over some time. Normally while preparing the index number, we compare the current prices of a product with the price of some past period known as the base year. The index number of the base year is mostly taken as 100. A few definitions of index numbers given by different experts are as follows:

According to Croxton and Cowden, "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

According to A.L. Bowley, "A series of index numbers reflects in its trend and fluctuations the movements of some quantity to which it is related."

According to Spiegel, "An Index number is a statistical measure designed to show changes in a variable, or a group of related variables with respect to time, geographic location, or other characteristics such as income, profession etc."

8.2 Features Of Index Number

1. Index numbers are specialized type of average. Normally used measures of average like Mean Median and Mode can be used for two or more different series, if their units are same. In case units of two series are different, these cannot be represented by normal average, However, Index numbers can help in this situation.
2. Normally index numbers are represented in percentages. However, the % sign is not used while showing index numbers.
3. Index numbers give the effect of change over some time or the change that is taking place in two different locations.
4. Index numbers measure those changes that are not capable of measurement normally in quantitative figures. For example, we cannot measure the change in the cost of living directly, but Index numbers can help us in this situation.

8.3 Uses Of Index Numbers

1. Index number is a very powerful tool for economic and business analysis. We often call index number 'Barometer of the Economy'. With the help of Index Numbers, we can see pulse of the economy.
2. Index number is a very helpful tool in planning activities and formulation of business policy.
3. With the help of index numbers, economists try to find out trends in prices, production, import and exports, etc.
4. Index number shows the cost of living over a period of time. This also helps the government in fixing the wage rate of labour.
5. Index number also helps us in the calculation of Real National Income of the country.

8.5 Problems In Construction Of Index Numbers

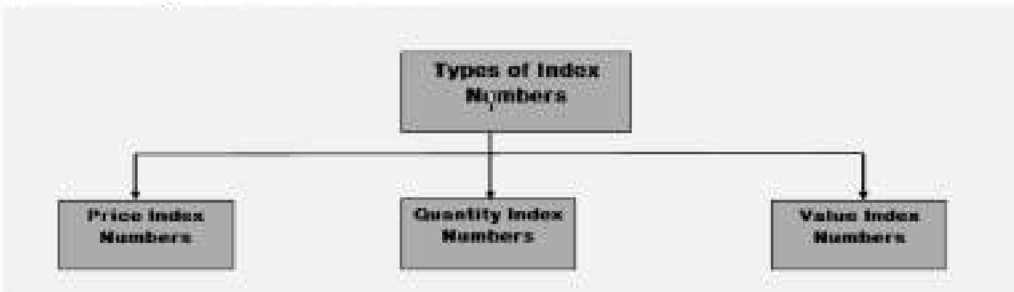
1. **Purpose of Index Numbers**: The first step in construction of index numbers is to decide the purpose of preparing the Index Number. As there is no single purpose index, so we must decide the objective of index very carefully.

2. **Selection Of Base Year**: The selection of base period is most important step in preparation of index number. The best base period is the period with which we can find accurate change in the variable. Following are some of the guidelines that must be kept in mind while selecting the base period.
   - The period selected as base period should be normal one. There must not be any problems like War, Flood, Earthquake, Economic Depression etc. in the base period.
   - The difference between base period and the current period should not be very large
   - Only that period should be taken as base period full data is available.
3. **Selection of Number of Items or Commodities**: The next major problem in preparation for index number is to select the number of items that will form the Index number. The following points must be kept in mind while deciding the number of items in the index numbers.
   - Only those items should be selected that represent the habits and tastes of majority of customers.
   - The number of items selected should not be very large or very small.
   - Only those items should be selected that are available in standard quantity.
   - Only those items must be selected that were available in base period as well as current period.
4. **Selection of Source of Data:** As in index numbers, we compare current variables with the variables of past periods, the source from which data is collected must be authentic. In case of non-authentic data, it will give wrong picture.
5. **Price Quotations:** The prices of the commodities differ from place to place. It is very important to select the price which represents majority of places. Further while preparing index numbers, we may take wholesale prices or retail prices in consideration.
6. **Selection of the Average**: There are different types of averages, like Arithmetic Mean, Geometric Mean, Harmonic Mean, Median and Mode that can be used in preparation for index numbers. One must select appropriate averages in preparation of index numbers based on our objective.
7. **Selection of Appropriate Weight:** The next major problem in preparation of index numbers is to assign weight to the different items. All the items of the data under consideration are not equally important, some items may be more important and some items may be less important. So, more weight must be assigned to important items while preparing the index number. Now

the problem is how to assign weights to the items. Normally we take quantity of the items consumed as weight in Index Number.

8. **Selection of appropriate formula:** There are several formulas that can be used for preparing index numbers. for example, Laspeyer's method. Bowle Method and Fisher Method etc. Each method has its advantages and limitations. so must be selected very carefully.

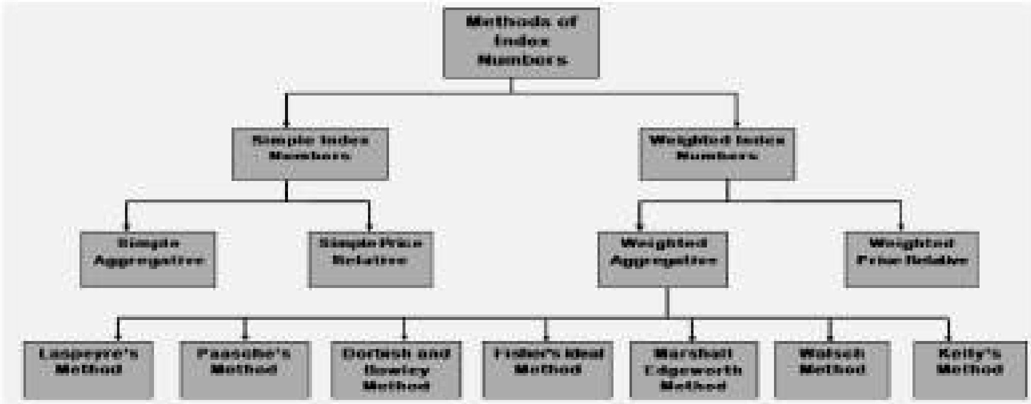
8.6 Different Types Of Index Numbers



1. **Price Index Numbers**: These index numbers are used for measuring the change in prices of the commodities over a period of time. In other words, we can say that these index numbers find the change in value of money over a period of time. These index numbers are most popular index numbers. These Index numbers may be based on Wholesale Price Index or Retail Price Index.
2. **Quantity Index Numbers**: The Quantity or Volume Index Numbers measure the change in quantities used by people over a period of time. under these index numbers, we calculate change in physical quantity of goods produced, consumed or sold over a period of time. There are different types of quantity index numbers such as Agricultural Production Index Number, Industrial Production Index Number, Export Import Index Number etc.
3. **Value Index Numbers**: Value Index Numbers compare the change in total value over period of time. These index numbers take into consideration both prices and quantity of the product while finding the change over a period of time. These Index Numbers are very useful in finding consumption habits of the consumers.

8.7 Different Methods Of Index Numbers

As we have already discussed, Index number is a device that shows changes in price over a period of time. Now a question arises that how to calculate the index number. There are several methods for preparing the index numbers. The following chart shows various methods of preparing index numbers.



8.7.1 Simple Index Number

This further divided into the simple aggregative and simple price relative

8.7.1.1 Simple Aggregative Method

This is one of the old and simple methods of finding the index number. Under this method we calculate the index number of a given period by dividing the aggregate of all the prices of the current year by the aggregate of all the prices of the base year. After that we multiply the resultant figure with 100 to find the index number. The following are the steps:

1. Decide the base year.
2. Add all the prices of base year for all available commodities, it is denoted by $\sum P_0$.
3. Add all the prices of base year for all available commodities, it is denoted by $\sum P_1$.
4. Use following the formula for calculating index number under this method:

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Where,

P_{01} – Price Index Number of Current Year

$\sum P_1$ – Aggregate of Prices of Current Year

$\sum P_0$ – Aggregate of Prices of Base Year

Example 1. Construct Simple Aggregative Index number of the year 2020 by taking the base as prices of 2015.

Commodity	Price of the Year 2015	Price of the Year 2020

Wheat	**20**	**26**
Sugar	**40**	**34**
Oil	**60**	**120**
Pulses	**80**	**140**

Solution: Price Index (The year 2015 taken as the base year)

Commodity	Price of the Year 2015 P_0	Price of the Year 2020 P_1
Wheat	20	26
Sugar	40	34
Oil	60	120
Pulses	80	140
	$\sum P_0 = 200$	$\sum P_1 = 320$

Price Index (P_{01}) = $\frac{\sum P_1}{\sum P_0} \times 100$ = $\frac{320}{200} \times 100$ = 160

Price index shows that prices have increased by 60% in 2020 than 2015.

Merits of Simple Aggregative Method

1. This method is simple to calculate.
2. This method is very simple to understand.
3. This method does not need many mathematical calculations

Limitations of Simple Aggregative Method

1. This method does not give change in price over a period of time.
2. Prices of different commodities are measured in different units some are measured in Kilograms where as others in Meters etc. It creates problems in calculation.
3. This method is influenced by unit of measurement.
4. This method ignores the relative importance of the item.
5. This method uses only Arithmetic mean as a tool for calculating index number. Other measures of average like Geometric mean or median etc. cannot be used in this method.
6. Index number in this method is influenced by magnitude of the price.

8.7.1.2 Simple Price Relative Method

This method is a bit improved over the simple aggregative method. The simple aggregative method is affected by the magnitude of the price of the item. However, this method is not affected by magnitude of the price of item. Further, in this method it is not necessary to use Arithmetic mean as average rather we can use any method of finding average, such as Arithmetic mean, Geometric Mean, Median, Mode etc. However, normally we prefer to use Arithmetic mean in this case. Following are the steps of this method:

1. Decide the base year.
2. Calculate the price relative to current year for each commodity by dividing current Prices (P_1) with base year price (P_0) using the following formula $\frac{P_1}{P_0} \times 100$
3. Find sum of all the price relatives so calculated.
4. Divide the sum or price relatives by number of items to get index number by using the following formula:

$$P_{01} = \frac{\sum \frac{P_1}{P_0} \times 100}{N}$$

Merits of Simple Price Relative Method:

1. This method is very simple to calculate and understand.
2. This method is not affected by the magnitude of price of a particular item.
3. This method is not affected by unit of measurement of the item.
4. This method is not necessarily based on Arithmetic Mean, we can use other averages like Geometric Mean, median etc also.
5. Equal weights are provided for each item.

Limitations of Simple Price Relative Method:

1. Selection of average is a difficult task in this method.
2. If it is to be calculated using Geometric Mean, then a calculation is very difficult.
3. It does not consider which item is used more and provides equal weights to all items.

Example 2. Construct Simple Price Relative Index number of the year 2020 by taking the base as prices of 2015.

Commodity	**Price of the Year 2015**	**Price of the Year 2020**
Wheat	**20**	**26**
Sugar	**40**	**34**
Oil	**60**	**120**
Pulses	**80**	**140**

Solution: Price Index (Year 2015 taken as the base year)

Commodity	Price of the Year 2015 P_0	Price of the Year 2020 P_1	Price Relative $\frac{P_1}{P_0} \times 100$

Wheat	20	26	$\frac{26}{20} \times 100 = 130$
Sugar	40	34	$\frac{34}{40} \times 100 = 85$
Oil	60	120	$\frac{120}{60} \times 100 = 200$
Pulses	80	140	$\frac{140}{80} \times 100 = 175$
			$\sum\frac{P_1}{P_0} \times 100 = 590$

Price Index (P_{01}) = $\frac{\Sigma\frac{P_1}{P_0}\times 100}{N} = \frac{590}{4} = 147.50$

Price index shows that prices have increased by 47.5% in 2020 than 2015.

TEST YOUR PROGRESS (A)

1. Calculate Index number for 2015 taking 209 as base using Simple Aggregative Method and Simple Average of Relatives Method:

Items	Price 2011	Price 2015
A	350	510
B	45	40
C	77	156
D	37	47
E	10	12

2. Find index using simple average of price relative using 2017 as base.

Items	Price 2017	Price 2019
A	15	30
B	18	24
C	16	20
D	14	21
E	25	35
F	40	30

3. Find simple aggregative index

Items	P_0	P_1
Oil	60	70
Pulses	70	60
Rice	50	40
Sugar	40	40

Answers

1) 147.4, 132.84, 2) 137.22, 3) 95.45

8.7.2 Weighted Index Number

Which is further divided into weighted aggregative and weighted price relative method

8.7.2.1 Weighted Aggregative Price Index

Simple Aggregative methods of Index Numbers assume that all the items of Index Number are equally important. There is no item that is more important than others. So, this method provides equal weightage to all items. However, in practical life it is not true. Some items carry more importance than other items, for example in human's life expenditure on food carries more importance than expenditure on entertainment. So, we have weighted method of index numbers which considers relative importance of the item also.

The weighted Aggregative Method is one such method. This method is more or less same as Simple Aggregative Method but main difference is that it also considers relative weights of the items also. Generally, the quantity of the item consumed is considered as weight in this case. There are many methods of calculating Weighted Aggregative Price Index which are discussed as follows:

a) Laspeyre's Method:

This method was suggested by Mr. Laspeyre in 1871. Under this method base year quantities of the various products are assumed as weight for preparing the index numbers. The following steps may be used:

1. Multiply Prices of the base year (P_0) with the quantities of the base year (Q_0) for every commodity.
2. Add the values calculated in step 1, the sum is denoted as $\sum P_0 Q_0$
3. Multiply Prices of the current year (P_1) with the quantities of the base year (Q_0) for every commodity.
4. Add the values calculated in step 3, the sum is denoted as $\sum P_1 Q_0$
5. Use following formula for calculating index number:

$$P_{01} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

b) Paasche's Method:

This method was suggested by Mr. Paasche in 1874. Under this method current year quantities of the various products are assumed as weight for preparing the index numbers. The following steps may be used:

1. Multiply Prices of the base year (P_0) with the quantities of the current year (Q_1) for every commodity.
2. Add the values calculated in step 1, the sum is denoted as $\sum P_0 Q_1$

3. Multiply Prices of the current year (P_1) with the quantities of the current year (Q_1) for every commodity.
4. Add the values calculated in step 3, the sum is denoted as $\sum P_1 Q_1$.
5. Use the following formula for calculating index number:

$$P_{01} = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

c) Dorbish and Bowley's Method:

This method is based on both Lasypeyre's Method and Paasche's Method, that's why this method is also known as L-P formula. Under this method, we calculate the index number by taking the arithmetic mean of the formula given by Laspeyre and Paasche. So, following formula is used in case of the Dorbish and Bowley Method:

$$P_{01} = \frac{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} + \frac{\sum P_1 Q_1}{\sum P_0 Q_1}}{2} \times 100$$

d) Fisher's Ideal Index Method:

This method was suggested by Prof Irving Fisher and it is assumed as one of the best methods of constructing the Index Number. That's why this method is also called Ideal Index Number. This method is based on both Lasypeyre's Method and Paasche's Method, but instead of taking arithmetic mean of both formulas, Fisher used the geometric mean of the formula given by Laspeyre and Paasche. So, following formula for calculating Fisher's ideal Index number:

$$\sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} \times 100$$

Fisher's Method is called ideal index number due to following reasons:

1. This method uses geometric mean as base which is perhaps best average for constructing Index numbers.
2. This method considers both quantities of base year as well as current year as weight.
3. This method satisfies both the time reversal and factor reversal tests.
4. It is comprehensive method and covers all values of data i.e., P_0, Q_0, P_1, Q_1 etc.

e) Marshal Edgeworth Index Method:

Like Fisher's method, this method also uses the quantities of base as well as current year as weight. Under this method arithmetic mean of the quantity of base and current year is assumed as weight. This method is comparatively simple than Fisher's method as it does not use complex concept of Geometric mean. Following is the formula of this method.

$$P_{01} = \frac{\sum P_1(Q_0 + Q_1)}{\sum P_0(Q_0 + Q_1)} \times 100 \quad \text{or}$$

$$\frac{\sum P_1Q_0 + \sum P_1Q_1}{\sum P_0Q_0 + \sum P_0Q_1} \times 100$$

Example 3. Construct Weighted Aggregative Index number of the year 2020 by taking the base as prices of 2015 using Laspeyre, Paasche, Dorbish & Bowley, Fisher, Marshal Edgeworth and Kelly's method.

Item	Price of the The year 2015	Quantity of the The year 2015	Price of the Year 2020	Quantity of the Year 2020
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

Solution:

Item	P_0	Q_0	P_1	Q_1	P_0Q_0	P_0Q_1	P_1Q_0	P_1Q_1
A	6	50	10	56	300	336	500	560
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
E	8	40	12	36	320	288	480	432
					$\sum P_0Q_0 = 1360$	$\sum P_0Q_1 = 1344$	$\sum P_1Q_0 = 1900$	$\sum P_1Q_1 = 1880$

1. Laspeyre's Method:

$$P_{01} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100 = \frac{1900}{1360} \times 100 = 139.71$$

2. Paasche's Method:

$$P_{01} \quad \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100 = \frac{1880}{1344} \times 100 = 139.88$$

3. Dorbish and Bowley's Method:

$$P_{01} = \frac{\frac{\sum P_1Q_0}{\sum P_0Q_0} + \frac{\sum P_1Q_1}{\sum P_0Q_1}}{2} \times 100 = \frac{\frac{1900}{1360} + \frac{1880}{1344}}{2} \times 100 = \frac{2.796}{2} = 139.79$$

4. Fisher's Ideal Index Method:

$$\sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1}} \times 100 = \sqrt{\frac{1900}{1360} \times \frac{1880}{1344}} \times 100 = \sqrt{1.9543} \times 100 = 139.79$$

5. Marshal Edgeworth Index Method: $\frac{\sum P_1Q_0+\sum P_1Q_1}{\sum P_0Q_0+\sum P_0Q_1} \times 100$

$$= \frac{1900 + 1880}{1360 + 1344} \times 100 = \frac{3780}{2704} \times 100 = 139.79$$

Example 4. Construct Weighted Aggregative Index number using Laspeyre, Paasche, Dorbish & Bowley and Fisher, methods.

Item	Price of the Base Year	Expenditure of the Base Year	Price of the Current Year	Expenditure of the Current Year
A	2	40	5	75
B	4	16	8	40
C	1	10	2	24
D	5	25	10	60

Solution:

We know that Expenditure = Price × Quantity

So, Quantity = $\frac{Expenditure}{Price}$

Item	P_0	Q_0	P_1	Q_1	P_0Q_0	P_0Q_1	P_1Q_0	P_1Q_1
A	2	20	5	15	40	30	100	75
B	4	4	8	5	16	20	32	40
C	1	10	2	12	10	12	20	24
D	5	5	10	6	25	30	50	60
					$\sum P_0Q_0 = 91$	$\sum P_0Q_1 = 92$	$\sum P_1Q_0 = 202$	$\sum P_1Q_1 = 199$

1. Laspeyre's Method:

$$P_{01} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100 = \frac{202}{91} \times 100 = 221.98$$

2. Paasche's Method:

$$P_{01} \quad \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100 = \frac{199}{92} \times 100 = 216.39$$

3. Dorbish and Bowley's Method:

$$P_{01} = \frac{\frac{\sum P_1Q_0}{\sum P_0Q_0} + \frac{\sum P_1Q_1}{\sum P_0Q_1}}{2} \times 100$$

$$= \frac{\frac{202}{91} + \frac{199}{92}}{2} \times 100 = \frac{4.3828}{2} = 219.14$$

4. Fisher's Ideal Index Method:

$$\sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} \times \frac{\sum P_1Q_1}{\sum P_0Q_1}} \times 100$$

$$= \sqrt{\frac{202}{191} \times \frac{199}{92}} \times 100 = \sqrt{4.8015} \times 100 = 219.12$$

8.7.2.2 Weighted Price Relative Method

This method almost similar to simple price-relative method. However, simple price relative gives equal importance to all items under consideration. But in our life, all items do not carry equal importance. Some items are more important or on some items we spend more amount. Changes in price of some items affect us more than changes in price of some other items. So, we have weighted price relative method. This method is similar to simple price relative method but also assigns weights to the items. Further, in this method it is not necessary to use Arithmetic mean as average rather we can use any method of finding average, such as Arithmetic mean, Geometric mean, etc. However, normally we prefer to use Arithmetic mean in this case. Following are the steps of this method:

1. Decide the base year.
2. Calculate the price relative to current year for each commodity by dividing current Prices (P_1) with base year price (P_0) using the following formula $\frac{P_1}{P_0} \times 100$.
3. Find the weights of the items to be assigned.
4. Multiply price relative so calculated with the weights and find out the product of both.
5. Find sum of product so calculated.
6. Find sum of the weights assigned.
7. Divide the sum of the weighted price relatives by sum of weights to get index number by using the following formula:

$$P_{01} = \frac{\sum W \frac{P_1}{P_0} \times 100}{\sum W}$$

Merits of Weighted Price Relative Method:

1. This method is very simple to calculate and understand.
2. This method is not affected by the magnitude of price of a particular item.
3. This method is not affected by unit of measurement of the item.
4. This method is not necessarily based on Arithmetic Mean, we can use other averages like Geometric Mean, median etc also.

5. Weights are assigned according to importance of the items.

Limitations of Weighted Price Relative Method:

1. Selection of average is a difficult task in this method.
2. If it is to be calculated using Geometric Mean, then calculation is very difficult.
3. Selection of weights is a difficult task.

Example 5. Construct Weighted Price Relative Index number of the year 2020 by taking the base as prices of 2015.

Commodity	Price of the Year 2015	Price of the Year 2020	Weights
Wheat	**20**	**26**	**40**
Sugar	**40**	**34**	**5**
Oil	**60**	**120**	**3**
Pulses	**80**	**140**	**2**

Solution: Price Index (Year 2015 taken as the base year)

Commodity	Price of the Year 2015 P_0	Price of the Year 2020 P_1	Price Relative $\frac{P_1}{P_0} \times 100$	Weights (W)	Weighted Price Relatives $W\frac{P_1}{P_0} \times 100$
Wheat	20	26	$\frac{26}{20} \times 100 = 130$	40	5200
Sugar	40	34	$\frac{34}{40} \times 100 = 85$	5	424
Oil	60	120	$\frac{120}{60} \times 100 = 200$	3	600
Pulses	80	140	$\frac{140}{80} \times 100 = 175$	2	350
				$\sum W = 50$	$\sum W\frac{P_1}{P_0} \times 100 = 6575$

Price Index (P_{01}) $= \frac{\sum W\frac{P_1}{P_0} \times 100}{\sum W} = \frac{6575}{50} = 131.50$

Price index shows that prices have increased by 31.5% in 2020 than 2015.

8.9 Sum Up

- Index number shows change in variable over a period of time.
- Price index shows change in price in current year in comparison to base year.
- Normally the base of index is taken as 100.
- There are different types of indexes like price index, quantity index, value index.
- Index number can be prepared without assigning weights or after assigning weights.

- Popular weighted aggregative index is Laspeyre, Paasche, Bowley, Fisher, Marshal Edgeworth and Kelly.

8.10 Key Terms

- **Index Numbers:** An index number is a device with the help of which we can measure the relative change in one variable over a period of time. Normally while preparing the index number, we compare the current year variable with the variable of as base year. The index number of the base year is mostly taken as 100
- **Price Index Numbers**: These index numbers are used for measuring the change in prices of the commodities over a period of time. In other words, we can say that these index numbers find the change in value of money over a period of time. These index numbers are most popular index number. These Index numbers may be based on Wholesale Price Index or Retail Price Index.
- **Quantity Index Numbers**: The Quantity or Volume Index Numbers measure the change in quantities used by people over a period of time. Under these index numbers, we calculate change in physical quantity of goods produced, consumed or sold over a period of time. There are different types of quantity index numbers such as Agricultural Production Index Number, Industrial Production Index Number, Export Import Index Number etc.
- **Value Index Numbers**: Value Index Numbers compare the change in total value over period of time. These index numbers take into consideration both prices and quantity of the product while finding the change over a period of time. These Index Numbers are very useful in finding consumption habits of the consumers.

8.11 QUESTIONS FOR PRACTICE

Q1.What are index numbers? What are its uses?

Q2.Explain problems faced in construction of index numbers.

Q3.What are different types of Index numbers.

Q4.Explain different steps in construction of index numbers.

Q5.What are different methods of construction of index numbers?

Q6.Explain Simple Aggregative Index numbers. What are its Merits and Limitations?

Q7.What is Simple Price Relative Index numbers? What are its Merits and Limitations?

Q8.Explain Weighted Aggregative Index numbers. What are its Merits and Limitations?

Q9.What is Weighted Price Relative Index numbers? What are its Merits and Limitations?

8.12 SUGGESTED READINGS

- J. K. Sharma, *Business Statistics*, Pearson Education.
- S.C. Gupta, *Fundamentals of Statistics*, Himalaya Publishing House.
- S.P. Gupta and Archana Gupta, *Elementary Statistics*, Sultan Chand and Sons, New Delhi.
- Richard Levin and David S. Rubin, *Statistics for Management*, Prentice Hall of India, New Delhi.