Business Mathematics and Statistics

DEQTT201

Edited by:
Dr. Pooja Kansra

LOVELY PROFESSIONAL UNIVERSITY

Business Mathematics and Statistics

Edited By
Dr. Pooja Kansra

Content

Unit 01: Matrices

CONTENTS

Objectives

After studying this unit, you will be able to,

- define and give examples of various types of matrices;
- evaluate the sum, difference, product and scalar multiples of matrices;
- evaluate determinants find minors and cofactors of square matrices of different orders; and apply properties of determinants.

Introduction

The knowledge of matrices has become necessary for the individuals working in different branches of science, technology, commerce, management and social sciences. In many economic analyses, variables are assumed to be related by sets of linear equations. Matrix algebra provides a clear and concise notation for the formulation and solution of such problems, many of which would be complicated in conventional algebraic notation. The concept of determinant and is based on that of matrix. In this unit, we introduce the concept of matrices and its elementary properties. Further, discusses the determinant, and a number associated with a square matrix and its properties.

Definition

A rectangular array of numbers is called a matrix.

The horizontal arrays of a matrix are called its ROWS and the vertical arrays are called its COLUMNS. A matrix having mrows and n columns is said to have the order m*n.

Let

$$A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 5 & 6 \end{bmatrix}.$$

Then,

$a_{11} = 1,\ a_{12} = 3,\ a_{13} = 7,\ a_{21} = 4,\ a_{22} = 5,\ a_{23} = 6.$

1.1 Matrices

A set of mn numbers (real or complex), arranged in a rectangular formation (array or table) having m rows and n columns and enclosed by a square bracket [ ] is called m×n matrix (read "m by n matrix") .

An m×n matrix is expressed as

$$A = \begin{bmatrix} a_{11} & a_{12} & ----a_{1n} \\ a_{21} & a_{22} & ----a_{2n} \\ --- & ---- & ------ \\ --- & ---- & ------ \\ a_{m1} & a_{m2} & ----a_{mn} \end{bmatrix}$$

i. A matrix is denoted by capital letters A, B, C, etc. of the English alphabets.
ii. First suffix of an element of the matrix indicates the position of row and second suffix of the element of the matrix indicates position of column. e.g. 23 a means it is an element in the second row and the third column.
iii. The order of a matrix is written as "number of rows ×number of columns".

1.2 Order of a Matrix

The order or dimension of a matrix is the ordered pair having as first component the number of rows and as second component the number of columns in the matrix. If there are 3 rows and 2 columns in a matrix, then its order is written as (3, 2) or (3 x 2) read as three by two. In general, if m are rows and n are columns of a matrix, then its order is (m x n).

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}$$

are matrices of orders (2 x 3), (3 x 1) and (4 x 4) respectively.

Example 1: Write the order of the matrix

$$A = \begin{bmatrix} 9 & 7 & 8 & -3 & -8 \\ 4 & 3 & 6 & 1 & -10 \\ 10 & 12 & 15 & 2 & 5 \end{bmatrix}$$

Also write the elements a_{23}, a_{14}, a_{35}, a_{22}, a_{31}, a_{32}.

Solution: Order of the matrix A is 3×5 and the desired elements are: Matrices and Determinants

a_{23}=6, a_{22}=3,

a_{14}= -3, a_{31}=10,

a_{35}= 5, a_{32}=12

Example 2: Write all the possible orders of the matrix having following elements.

i. 8
ii. 13

Solution: (i) All the 8 elements can be arranged in single row, i.e. 1 row and 8 columns.

Or

They can be arranged in two rows with 4 elements in each row, i.e. 2 rows and 4 columns.

Or

in four rows with 2 elements in each row, i.e. 4 rows and 2 columns.

Or

in eight rows with 1 element in each row, i.e. 8 rows and 1 column.

∴the possible orders are 1 *8, 2* 4, 4 *2, 8 *1.

iii. All the 13 elements can be arranged in single row, i.e. 1 row and 13 columns.

Or

in 13 rows with 1 element in each row, i.e. 13 rows and 1 column.

∴ the possible orders are1* 13, 13 *1.

1.3 Types of Matrices

On the basis of number of rows and number of columns and depending on the values of elements, the type of a matrix gets changed. Various types of matrix are explained as below:

Row Matrix

A matrix having only one row is called a row matrix.

For example, [2 5 7], [8 9], [1 0 3 2] all are row matrices

Column Matrix

A matrix having only one column is called a column matrix. For example,

$$\begin{bmatrix} 9 \\ 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 9 \\ -3 \\ 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 5 \\ -11 \end{bmatrix}$$

all are column matrices.

Null or Zero Matrixes:

A matrix in which each element is „0" is called a Null or Zero matrix. Zero matrices are generally denoted by the symbol O. This distinguishes zero matrix from the real number 0.

For example $O = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is a zero matrix of order 2 x 4.

The matrix O_{mxn} has the property that for every matrix A_{mxn},

A + O = O + A = A

Square matrix:

A matrix A having same numbers of rows and columns is called a square matrix. A matrix A of order m x n can be written as A_{mxn}. If m = n, then the matrix is said to be a square matrix. A square matrix of order n x n, is simply written as A_n.

$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ and $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

are square matrix of order 2 and 3.

Main or Principal (leading)Diagonal:

The principal diagonal of a square matrix is the ordered set of elements aij, where i = j, extending from the upper left-hand corner to the lower right-hand corner of the matrix. Thus, the principal diagonal contains elements a11, a22, a33 etc.

For example, the principal diagonal of

$$\begin{bmatrix} 1 & 3 & -1 \\ 5 & 2 & 3 \\ 6 & 4 & 0 \end{bmatrix}$$

consists of elements 1, 2 and 0, in that order.

Particular cases of a square matrix:

(a)Diagonal matrix:

A square matrix in which all elements are zero except those in the main or principal diagonal is called a diagonal matrix. Some elements of the principal diagonal may be zero but not all.

A square matrix A = [aij] $_{n \times n}$ is said to be diagonal matrix if aij=0, ∀i≠j

$$\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are diagonal matrices.

Notes:

(i) For a diagonal matrix all non-diagonal elements must be zero.

(ii) In a diagonal matrix some or all the diagonal elements may be zero.

(b) Scalar Matrix:

A diagonal matrix in which all the diagonal elements are same, is called a scalar matrix i.e. Thus

$\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ and $\begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix}$

(c) Identity Matrix or Unit matrix:

A scalar matrix in which each diagonal element is 1(unity) is called a unit matrix. An identity matrix of order n is denoted by In.

Thus,

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

are the identity matrices of order 2 and 3.

is an identity matrix if and only if aij = 0 for i ≠ j and aij = 1 for i = j

Notes: If a matrix A and identity matrix I are conformable for multiplication, then I has the property that AI = IA = A i.e.,

I is the identity matrix for multiplication.

d. Equal Matrices:

Two matrices A and B are said to be equal if and only if they have the same order and each element of matrix A is equal to the corresponding element of matrix B i.e for each i, j, aij= b_{ij}

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

B= $\begin{bmatrix} \frac{4}{2} & 2-1 \\ \sqrt{9} & 0 \end{bmatrix}$

hen A = B because the order of matrices A and B is same and aij = bij for everyi , j.

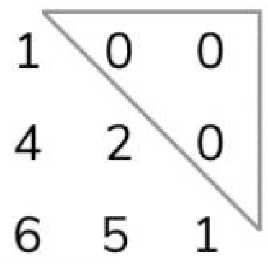
e. Upper Triangular Matrix

A square matrix A = [aij] n×n is said to be upper triangular matrix if all the elements below the principal diagonal are zero.

$$\begin{matrix} 1 & 2 & 4 & 5 \\ 0 & 2 & 3 & 6 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 4 \end{matrix}$$

f. Lower Triangular Matrix

A square matrix A = [aij] n×n is said to be lower triangular matrix if all the elements above the principal diagonal are zero.



2. Write orders and types of the following matrices:

a. $\begin{bmatrix} 2 & 9 \\ 3 & 4 \end{bmatrix}$

b. $\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$

c. $\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$

d. $\begin{matrix} 2 & 5 & 7 \\ 0 & 8 & 0 \\ 0 & 0 & 9 \end{matrix}$

e. $\begin{matrix} 2 \\ 9 \\ 6 \end{matrix}$

Solution: Order	Type
(i) 2× 2	Square matrix [rows and columns are equal in number.]
(ii)2× 2	Diagonal matrix [Θall the non-diagonal elements are zero.]
(iii) 3× 3	Identify matrix [all the diagonal elements are unity and nondiagonal element are zero.]
(iv) 3×3	Lower triangular matrix [all the elements above the principal diagonal are zero]
(v) 3×1	Column matrix [it has only one column.]

1.4 Operations on Matrices

In school times, a child first learns the natural numbers and then learns how these numbers are added, subtracted, multiplied and divided. Similarly, here also we now see as to how such operations (except division) are applied on matrices. These operations are explained by first giving a general formula and then examples followed by some exercises.

a. Addition and subtraction of Matrices

Addition of two matrices A and B make sense only if they are of the same order and obtained by adding their corresponding elements. It is denoted by A + B.

That is, if $A= [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{m\times n}$ then $A + B = [a_{ij}+b_{ij}]_{m\times n}$

Subtraction of two matrices A and B make sense only if they are of the same order, and is given by

$A - B = A + (- B) = A+ (-1) B$, i.e. A – B means addition of two matrices A and – B. So, if $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{m\times n}$, then $A - B = [a_{ij}+ (-1)b_{ij}]_{m\times n} = [a_{ij}-b_{ij}]_{m\times n}$

$$A = \begin{bmatrix} -5 & 1 & -3 \\ 6 & 0 & 2 \\ 2 & 6 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 4 & 5 \\ -8 & 10 & 3 \\ -2 & -3 & -9 \end{bmatrix}$$

$$A + B = \begin{bmatrix} -3 & 5 & 2 \\ -2 & 10 & 5 \\ 0 & 3 & -8 \end{bmatrix} \qquad A - B = \begin{bmatrix} -7 & -3 & -8 \\ 14 & -10 & -1 \\ 4 & 9 & 10 \end{bmatrix}$$

Properties of Addition of Matrices

If A, B, C are of the same orders over R, (i.e. elements of A, B, C are real numbers) then

(i) A + B = B + A (commutative law).

(ii) (A + B) + C = A + (B + C) (associative law).

(iii) A + O = O + A = A, where O is a null matrix. (existence of additive identity) .

(iv) For a given matrix A, there exists a matrix B of the same order such that A + B = O = B + A. Here B is called additive inverse of A. (existence of additive inverse).

a. Product of Matrices:

Two matrices A and B are said to be conformable for the product AB if the number of columns of A is equal to the number of rows of B. Then the product matrix AB has the same number of rows as A and the same number of columns as B.

Thus, the product of the matrices A_{mxp} and B_{pxn} is the matrix $(AB)_{mxn}$. The elements of AB are determined as follows:

The element Cij in the ith row and jth column of (AB)mxn is found by cij = $a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \dots\dots\dots + a_{in}b_{nj}$

for example, consider the matrices

$$A_{2x2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad B_{2x2} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Since the number of columns of A is equal to the number of rows of B, the product AB is defined and is given as

Notes:

1. Multiplication of matrices is not commutative i.e., AB ≠ BA in general.
2. For matrices A and B if AB = BA then A and B commute to each other
3. A matrix A can be multiplied by itself if and only if it is a square matrix. The product A.A in such cases is written as A2. Similarly, we may define higher powers of a square matrix i.e., $A * A^2 = A^3$, $A^2 * A^2 = A4$
4. In the product AB, A is said to be pre multiple of B and B is said to be post multiple of A.

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Thus, c_{11} is obtained by multiplying the elements of the first row of A i.e., a_{11}, a_{12} by the corresponding elements of the first column of B i.e., b_{11}, b_{21} and adding the product.

Similarly, c_{12} is obtained by multiplying the elements of the first row of A i.e., a_{11}, a_{12} by the corresponding elements of the second column of B i.e., b_{12}, b_{22} and adding the product. Similarly, for c_{21}, c_{22}.

Multiplication of Matrices

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{bmatrix}$$

3. Solve

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

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

= $\begin{bmatrix} 5 & 4 \\ 5 & 9 \end{bmatrix}$

4. Solve:

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

Solution:

Since A is a (2 x 3) matrix and B is a (3 x 2) matrix, they are conformable for multiplication. We have

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

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

= $\begin{bmatrix} 11 & 0 \\ 4 & 0 \end{bmatrix}$

Properties of Matrix Multiplication

If A, B, C are three matrices such that corresponding multiplications hold then

(1) A(BC) = (AB)C (associative law)

(2) (i) A(B + C) = AB + AC (left distributive law)

(ii) (A + B)C = AC +BC (right distributive law)

(3) If A is a square matrix of order n, then $I_nA = AI_n = A$, where I_n is the identity matrix of order n.

Remark

Commutative law does not hold, in general, i.e. AB ≠ BA, in general. But for some cases AB may be equal to BA. This has been explained below:

(i) AB may be defined but BA may not be defined and hence AB ≠ BA in this case.

For example, let A be a matrix of order 3×2 and B be a matrix of order 2×4.

Here AB is defined and is of order 3×4

But BA is not defined (∵ number of columns of B ≠ number of rows of A).

(ii) AB and BA both may defined but may not be of same order and hence AB BA. ≠

For example, let A be a matrix of order be 3×2 and B be a matrix of order 2×3.

Here as number of columns of A = number of rows of B.

∴ AB is defined and is of order 3×3.

Also, number of columns of B = number of rows of A.

Hence BA is defined but of order 2×2. ∴ AB ≠ BA.

(iii) AB and BA both may be defined and of same order but even then, they may not be equal.

1.5 Determinant of Square Matrices

Determinant is a number associated with each square matrix. In this section, we will deal with determinant of square matrices of order 1, 2, 3 and 4. Determinants of square matrices of order greater than 4 can be evaluated in a similar fashion.

The determinant of a matrix is a scalar (number), obtained from the elements of a matrix by specified, operations, which is characteristic of the matrix. The determinants are defined only for square matrices. It is denoted by det A or |A| for a square matrix A.

The determinant of the (2 x 2) matrix

$A = \begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix}$

is given by det A = $|A| = \begin{vmatrix} a11 & a12 \\ a21 & a22 \end{vmatrix}$

= a11 a22 – a12 a21

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

Solution: $|A| = \begin{vmatrix} 3 & 1 \\ -2 & 3 \end{vmatrix}$

= 9 – (–2) = 9 + 2 = 11

The determinant of the (3 x 3) matrix

$A = \begin{bmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{bmatrix}$, denoted by

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

is given as, det A = |A|

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

Each determinant in the sum (In the R.H.S) is the determinant of a submatrix of A obtained by deleting a particular row and column of A.

These determinants are called minors. We take the sign + or – , according to (– 1)i+jaij Where i and j represent row and column.

Minor and Cofactor of Element:

The minor Mij of the element aij in a given determinant is the determinant of order (n – 1 x n – 1) obtained by deleting the ith row and jth column of Anxn.

For example, in the determinant

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

The minor of the element a11 is M11= $\begin{vmatrix} a22 & a23 \\ a32 & a33 \end{vmatrix}$

The minor of the element a11 is M12= $\begin{vmatrix} a21 & a23 \\ a31 & a33 \end{vmatrix}$

The minor of the element a11 is M13= $\begin{vmatrix} a21 & a22 \\ a31 & a32 \end{vmatrix}$ and so on

The scalars Cij = $(-1)^{i+j}M_{ij}$are called the cofactor of the element $a_{ij}$of the matrix A.

The value of the determinant in equation (1) can also be found by its minor elements or cofactors, as $a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13}$

Or

Notes:

i. If all the elements of a row (or column) are zeros, then the value of the determinant is zero.

ii. If value of determinant 'Δ' becomes zero by substituting x = α, then x – α is a factor of 'Δ'.

iii. If all the elements of a determinant above or below the main diagonal consist of zeros, then the value of the determinant is equal to the product of diagonal elements.

$a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$

Hence the det A is the sum of the elements of any row or column multiplied by their corresponding cofactors. The value of the determinant can be found by expanding it from any row or column.

1.6 Properties of Determinants

Expanding a determinant in which the elements are large number can be a very tedious task. It is possible, however, by knowing something of the properties of determinants, to simplify the working. So here are some of the main properties.

(i) The value of a determinant remains unchanged if rows are changed to column and column to rows.

$\begin{vmatrix} a1 & a2 \\ b1 & b2 \end{vmatrix} = \begin{vmatrix} a1 & b1 \\ a2 & b2 \end{vmatrix}$

(ii) If two rows (or two columns) are interchanged, the sign of the determinant is changed

$\begin{vmatrix} a2 & b2 \\ a1 & b1 \end{vmatrix} = -\begin{vmatrix} a1 & b1 \\ a2 & b2 \end{vmatrix}$

(iii) If two rows (or two columns) are identical, the value of the determinant is zero.

$\begin{vmatrix} a1 & a2 \\ b1 & b2 \end{vmatrix} = 0$

(iv) If the elements of any one row (or column) are all multiplied by a common factor, the determinant is multiplied by that factors.

$\begin{vmatrix} ka1 & kb1 \\ a2 & b2 \end{vmatrix} = k\begin{vmatrix} a1 & b1 \\ a2 & b2 \end{vmatrix}$

(v) If the elements of any row (or column) are increased (or decreased) by equal multiples of the corresponding elements of any other row (or column), the value of the determinant is unchanged.

$\begin{vmatrix} a1 + kb1 & b1 \\ a2 + kb2 & b2 \end{vmatrix} = \begin{vmatrix} a1 & b1 \\ a2 & b2 \end{vmatrix}$

1. Find the minor of each element of the following matrices:

a. $\begin{bmatrix} 2 & 5 \\ 4 & -7 \end{bmatrix}$

Solution:

a. Let A= $\begin{bmatrix} 2 & 5 \\ 4 & -7 \end{bmatrix}$

Let Mij denotes the minor of (i, j) th element of the matrix A, i, j = 1, 2

∴ $M_{11}=|-7|=-7$

Determinant obtained after deleting first row and first column of matrix A=|– 7|

Similarly, $M_{12}=|4|=4$, $M_{21}=|5|=5$, $M_{22}=|2|=2$

1. Find the cofactor of each element of the following matrices:

$$A = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}$$

Solution: let Mij be the minor of every element

$$M_{11} = \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = 45 - 48 = -3$$

$$M_{12} = \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = 36 - 42 = -6$$

$$M_{13} = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 32 - 35 = -3$$

$$M_{21} = \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} = 18 - 24 = -6$$

$$M_{22} = \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} = 9 - 27 = -12$$

$$M_{23} = \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = 8 -14 = -16$$

$$M_{31} = \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = 12 - 15 = -3$$

$$M_{32} = \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} = 6 - 12 = -6$$

$$M_{33} = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = 5 - 8 = -3$$

The cofactor matrix A is

$$A = \begin{vmatrix} +(-3) & -(-6) & +(-3)) \\ -(-6) & +(-12) & -(-6) \\ +(-3) & -(6) & +(-3) \end{vmatrix}$$

$$A = \begin{vmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{vmatrix}$$

1.7 Adjoint and Inverse of a Matrix

(i) The adjoint of a square matrix A = [aij] n×n is defined as the transpose of the matrix. [aij] n×n , where Aij is the co-factor of the element aij. It is denoted by adj A

$$\text{If } A=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \text{ then } adj\ A=\begin{vmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{vmatrix}, \text{ where } A_{ij} \text{ is co-factor of } a_{ij}.$$

(ii) A (adj A) = (adj A) A = |A| I, where A is square matrix of order n.

(iii) A square matrix A is said to be singular or non-singular according as |A| = 0 or |A| ≠ 0, respectively.

(iv) If A is a square matrix of order n, then |adj A| = $|A|^{n-1}$.

(v) If A and B are non-singular matrices of the same order, then AB and BA are also nonsingular matrices of the same order.

(vi) The determinant of the product of matrices is equal to product of their respective determinants, that is, |AB| = |A| |B|.

(vii) If AB = BA = I, where A and B are square matrices, then B is called inverse of A and is written as B = A–1. Also $B^{-1} = (A^{-1})^{-1} = A$.

(viii) A square matrix A is invertible if and only if A is non-singular matrix. (ix) If A is an invertible matrix, then A^{-1} = (1 / | A |)* (adj A)

2. Find out the inverse of

$$\begin{bmatrix} 1 & -1 & 2 \\ 4 & 0 & 6 \\ 0 & 1 & -1 \end{bmatrix}$$

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

A^{-1}=adj(A)/|A|

To find out the adj(A), first we have to find out cofactor(A).

a11 = -6, a12 = 4, a13 = 4

a21 = 1, a22 = -1, a23 = -1

a13 = -6, a32 = 2, a33 = 4

So, cofactor(A) = $\begin{bmatrix} -6 & 4 & 4 \\ 1 & -1 & -1 \\ -6 & 2 & 4 \end{bmatrix}$

adj(A) = $[\text{cofactor}(A)]^T$

adj(A) = $[\text{cofactor}(A)]^T = \begin{bmatrix} -6 & 4 & 4 \\ 1 & -1 & -1 \\ -6 & 2 & 4 \end{bmatrix}^T$

adj(A) = $\begin{bmatrix} -6 & 1 & -6 \\ 4 & -1 & 2 \\ 4 & -1 & 4 \end{bmatrix}$

Then, | A | = 1(0-6)+1(-4-0)+2(4-0) = -6-4+8 = -2

A^{-1}=adj(A)/|A| = $\dfrac{\begin{bmatrix} -6 & 1 & -6 \\ 4 & -1 & 2 \\ 4 & -1 & 4 \end{bmatrix}}{-2}$

$A^{-1}=\begin{bmatrix} 3 & -1/2 & 3 \\ 2 & 1/2 & -1 \\ -2 & 1/2 & -2 \end{bmatrix}$

1.8 Solution of Equations using Matrices

One of the most important applications of matrices is to the solution of linear simultaneous equations. On this leaflet we explain how this can be done.

Writing simultaneous equations in matrix form

Consider the simultaneous equations

$x + 2y = 4$

$3x - 5y = 1$

Provided you understand how matrices are multiplied together you will realize that these can be written in matrix form as

$$\begin{pmatrix} 1 & 2 \\ 3 & -5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

Writing

$A=\begin{pmatrix} 1 & 2 \\ 3 & -5 \end{pmatrix}$, $X=\begin{pmatrix} x \\ y \end{pmatrix}$ and $B=\begin{pmatrix} 4 \\ 1 \end{pmatrix}$

we have AX = B

This is the matrix form of the simultaneous equations. Here the only unknown is the matrix X, since A and B are already known. A is called the matrix of coefficients.

Solving the simultaneous equations

Given

AX = B

we can multiply both sides by the inverse of A, provided this exists, to give

$A^{-1}AX = A^{-1}B$

But $A^{-1}A = I$, the identity matrix. Furthermore, IX = X, because multiplying any matrix by an identity matrix of the appropriate size leaves the matrix unaltered.

So

$X = A^{-1} B$

2. Example. Solve the simultaneous equations $x + 2y = 4$

$3x - 5y = 1$

Solution. We have already seen these equations in matrix form:

$$\begin{pmatrix} 1 & 2 \\ 3 & -5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

We need to calculate the inverse of $A=\begin{pmatrix} 1 & 2 \\ 3 & -5 \end{pmatrix}$

$A^{-1}=\frac{1}{(1)(-5)-(2)(3)}\begin{pmatrix} -5 & -2 \\ -3 & 1 \end{pmatrix}$

$= -\frac{1}{11}\begin{pmatrix} -5 & -2 \\ -3 & 1 \end{pmatrix}$

Then X is given by

$X = A^{-1}B$

$= -\frac{1}{11}\begin{pmatrix} -5 & -2 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix}$

$= -\frac{1}{11}\begin{pmatrix}-22\\-11\end{pmatrix}$

$= \begin{pmatrix}2\\1\end{pmatrix}$

3. Solve the simultaneous equations

2x + 4y = 2

-3x + y = 11

Solution. In matrix form:

$\begin{pmatrix}2 & 4\\-3 & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}2\\11\end{pmatrix}$

We need to calculate the inverse of $A = \begin{pmatrix}2 & 4\\-3 & 1\end{pmatrix}$

$A^{-1} = \frac{1}{(2)(1)-(4)(-3)}\begin{pmatrix}1 & -4\\3 & 2\end{pmatrix}$

$= \frac{1}{14}\begin{pmatrix}1 & -4\\3 & 2\end{pmatrix}$

Then X is given by

$X = A^{-1}B$

$= \frac{1}{14}\begin{pmatrix}1 & -4\\3 & 2\end{pmatrix}\begin{pmatrix}2\\11\end{pmatrix}$

$= -\frac{1}{14}\begin{pmatrix}-42\\28\end{pmatrix}$

$= \begin{pmatrix}-3\\2\end{pmatrix}$

Hence x = −3, y = 2 is the solution of the simultaneous equations. You should check the solution by substituting x = −3 and y = 2 into both given equations, and verifying in each case that the left-hand side is equal to the right-hand side.

Summary

1. A system of mn numbers arranged in a rectangular formation along m rows and n columns and bounded by the brackets [ ] is called an m by n matrix : when m = n, it is called a square matrix. To locate any particular element of a matrix, the elements are denoted by a letter followed by two suffixes, which respectively specify the rows and the columns. This aij is the element in the ith row and j th column of the matrix A. In the notation, the matrix A is denoted by [aij].
2. Two matrices can be multiplied only when the number of column in the first is equal to the number of row in the second. Thus matrices [aij] = A and B = [b jk] are compatible for multiplication and [c ik] =C is their multiplication is not commutative, that is AB ≠ BA in general.
3. If A be any matrix, then a matrix B, if it exists, such that AB = BA = I is called the inverse of A.
4. Inverse of a square matrix A^{-1} =Adj A/ |A |
5. Product of a square matrix and its inverse AA−1 = A−1A = I
6. Solution of a set of linear equations

   A.X=B is X=A^{-1}.B

Keywords

Matrix: A rectangular array of numbers (elements).

Row Matrix: One row only.

Column Matrix: One column only.

Equal Matrices: Corresponding elements equal.

Diagonal Matrix: All elements zero except those on the leading diagonal.

Null Matrix: All elements zero.

Self Assessment

1. A matrix having one row and many columns is known as?
A. Row matrix
B. Column matrix
C. Diagonal matrix
D. None of the mentioned

2. A square matrix A = [aij]nxn, if aij = 0 for i> j then that matrix is known as _______
A. Upper triangular matrix
B. Lower triangular matrix
C. Unit matrix
D. Null matrix

3. Two matrixes can be added if _______
A. rows of both the matrices are same
B. columns of both the matrices are same
C. both rows and columns of both the matrices are same
D. number of rows of first matrix should be equal to number of columns of second

4. If the order of matrix A is m*p. And the order of B is p*n. Then the order of matrix AB is?
A. m × n
B. n × m
C. n × p
D. m × p

5. A square matrix in which all elements except at least one element in diagonal are zeros is said to be a
A. Identical matrix
B. Null/zero matrix
C. Column matrix
D. Diagonal matrix

6. Two matrices A and B are multiplied to get AB if
A. both are rectangular
B. both have same order
C. no of columns of A is equal to columns of B
D. both are square matrices

7. Which of the following matrix multiplications would not be possible?

A. a 5×6 matrix with a 6×3
B. a 4×3 matrix with a 3×3
C. a 2×2 matrix with a 3×3
D. a 5×6 matrix with a 3×3 matrix

8. A diagonal matrix having equal elements is called a
A. square matrix
B. identical matrix
C. scalar matrix
D. rectangular matrix

9. Generally, the matrices are denoted by
A. capital letters
B. numbers
C. small letters
D. operational signs

10. According to determinant properties, when two rows are interchanged then the signs of determinant
A. Must changes
B. Remains same
C. Multiplied
D. Divided

11. In matrices, the determinant of a matric is denoted by
A. Vertical lines around matrix
B. Horizontal lines around matrix
C. Bracket around matrix
D. None of the above

12. According to determinant properties, the determinant equals to zero if row is
A. Multiplied to row
B. Multiplied to column
C. Divided to row
D. Divided to column

13. According to determinant properties. The determinant of resulting matrix equals to k delta if elements of rows are
A. Multiplied to constant k
B. Added to constant k
C. Multiplied to constant k
D. Divided to constant

14. The matrix which does not have an inverse by solving it, is classified as
A. Unidentified matrix
B. Linear matrix

C. Non-singular matrix

D. Singular matrix

15. Which of the following is not a property of determinant?

A. The value of determinant changes if all of its rows and columns are interchanged

B. The value of determinant changes if any two rows or columns are interchanged

C. The value of determinant is zero if any two rows and columns are identical

D. The value of determinant gets multiplied by k, if each element of row or column is multiplied by k

Answers for Self Assessment

1.	A	2.	A	3.	C	4.	A	5.	D
6.	C	7.	A	8.	C	9.	A	10.	A
11.	D	12.	A	13.	C	14.	D	15.	A

Review Questions

1. Write each sum as a single matrix:

A. A= $\begin{bmatrix} 2 & 1 & 4 \\ 3 & -1 & 0 \end{bmatrix}$ and B= $\begin{bmatrix} 6 & 3 & 0 \\ -2 & 1 & 0 \end{bmatrix}$

B. A= $|1 \quad 3 \quad 5|$ and B= $|0 \quad -2 \quad 1|$

2. Write each product as a single matrix:

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

B. $|3 \quad -2 \quad 5| \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$

3. Solution by matrix inverse method

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

B. A= $\begin{bmatrix} 1 & 3 & 1 \\ 2 & 3 & 3 \\ 3 & 2 & 1 \end{bmatrix}$

4. Solve equations by using matrix

A. -2x + 3y = 8

3x - y = -5

B. 2 x - y + 3 z = - 3

- x - y + 3 z = - 6

x - 2y - z = - 2

Further Readings

- Hoffman, K. H, and Kunze Ray, Linear Algebra, Prentice Hall of India.
- Jain R. K. and Iyengar S. R. K., Advanced Engineering Mathematics, Narosa Publishing House, New Delhi.
- Kreyszing Erwin, Advanced Engineering Mathematics, John Wiley & Sons, Inc.
- Gantmacher, F. R., The Theory of Matrices, 2 Volume, 2nd Edition, New York, Chelsa.
- Stroud K. A., Engineering Mathematics, Macmillan Press Ltd

Web Links

- https://www.intmath.com/matrices-determinants/matrix-determinant-intro.php
- https://www.toppr.com/guides/maths/determinants/determinant-of-a-matrix/
- https://link.springer.com/content/pdf/bbm%3A978-1-4615-5299-4%2F1.pdf

Unit 02: Determinants

CONTENTS

Objectives

After studying this unit, you will be able to,

- To understand the meaning and application of Matrix and Determinant in real life
- Learn the Use of Input Output table in 2 and 3 Sector Models

Introduction

The knowledge of matrices has become necessary for the individuals working in different branches of science, technology, commerce, management and social sciences. In many economic analyses, variables are assumed to be related by sets of linear equations. Matrix algebra provides a clear and concise notation for the formulation and solution of such problems, many of which would be complicated in conventional algebraic notation. The concept of determinant and is based on that of matrix. In this unit, we introduce the concept of matrices and its elementary properties. Further, discusses the determinant, and a number associated with a square matrix and its properties.

Definition

A rectangular array of numbers is called a matrix.

The horizontal arrays of a matrix are called its ROWS and the vertical arrays are called its COLUMNS. A matrix having m rows and n columns is said to have the order m *n.

Let

$$A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 5 & 6 \end{bmatrix}.$$

Then,

$$a_{11} = 1,\ a_{12} = 3,\ a_{13} = 7,\ a_{21} = 4,\ a_{22} = 5,\ a_{23} = 6.$$

2.1 Matrices

A set of m n numbers (real or complex), arranged in a rectangular formation (array or table) having m rows and n columns and enclosed by a square bracket [ ] is called m×n matrix (read "m by n matrix").

An m×n matrix is expressed as

$$A = \begin{bmatrix} a_{11} & a_{12} & ----a_{1n} \\ a_{21} & a_{22} & ----a_{2n} \\ --- & ---- & ------ \\ --- & ---- & ------ \\ a_{m1} & a_{m2} & ----a_{mn} \end{bmatrix}$$

i. A matrix is denoted by capital letters A, B, C, etc. of the English alphabets.
ii. First suffix of an element of the matrix indicates the position of row and second suffix of the element of the matrix indicates position of column. e.g. 23 a means it is an element in the second row and the third column.
iii. The order of a matrix is written as "number of rows ×number of columns".

2.2 Order of a Matrix

The order or dimension of a matrix is the ordered pair having as first component the number of rows and as second component the number of columns in the matrix. If there are 3 rows and 2 columns in a matrix, then its order is written as (3, 2) or (3 x 2) read as three by two. In general, if m are rows and n are columns of a matrix, then its order is (m x n).

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}$$

are matrices of orders (2 x 3), (3 x 1) and (4 x 4) respectively.

Example 1: Write the order of the matrix

$$A = \begin{bmatrix} 9 & 7 & 8 & -3 & -8 \\ 4 & 3 & 6 & 1 & -10 \\ 10 & 12 & 15 & 2 & 5 \end{bmatrix}$$

Also write the elements a_{23}, a_{14}, a_{35}, a_{22}, a_{31}, a_{32}.

Solution: Order of the matrix A is 3×5 and the desired elements are: Matrices and Determinants

$a_{23}=6$, $a_{22}=3$,

$a_{14}= -3$, $a_{31}=10$,

$a_{35}=5$, $a_{32=12}$

Example 2: Write all the possible orders of the matrix having following elements.

i. 8
ii. 13

Solution: (i) All the 8 elements can be arranged in single row, i.e. 1 row and 8 columns.

Or

They can be arranged in two rows with 4 elements in each row, i.e. 2 rows and 4 columns.

Or

in four rows with 2 elements in each row, i.e. 4 rows and 2 columns.

Or

in eight rows with 1 element in each row, i.e. 8 rows and 1 column.

∴ the possible orders are 1 *8, 2* 4, 4 *2, 8 *1.

iii. All the 13 elements can be arranged in single row, i.e. 1 row and 13 columns.

Or

in 13 rows with 1 element in each row, i.e. 13 rows and 1 column.

∴ the possible orders are1* 13, 13 *1.

2.3 Determinants

A staggering paradox hits us in the teeth. For abstract mathematics happens to work. It is the tool that physicists employ in working with the nuts and bolts of the universe! There are many examples from the history of science of a branch of pure mathematics which, decades after its invention, suddenly finds a use in physics. David Peat

A matrix presents a great deal of information in compact, readable form. Finding optimal generalities of solutions to large linear programming problems requires extensive use of matrices. mathematics. The properties and applications of matrices are studied in linear algebra, a discipline that includes much of the material of this chapter. In this section we introduce and solved mathematical the determinant of a square matrix as another tool to help solve systems of linear puzzles supplied by his equations.

Every square matrix A has an associated number called its determinant, denoted by det(A) or |A| . To evaluate determinants, we begin by giving a recursive definition, starting with the determinant of a 2* 2 matrix

A matrix is a rectangular array arranged in horizontal rows and vertical columns. The number of rows and columns give the dimension, or size, of the matrix. A matrix with m rows and n columns is called an m by n (m × n) matrix. Double subscripts provide a convenient system of notation for labeling or locating matrix entries.

Herearesomematricesofvarioussizes:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{33} \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix A is 3 × 3, B is 3 × 1, and C is 2 × 2. A and B show the use of double subscripts: aij is the entry in the ith row and the jth column. The first subscript identifies the row, the second tells the column; virtually all references to matrices are given in the same order, row first and then column. A matrix with the same number of rows and columns is a square matrix.

Determinant of a 2 2 matrix*

For 2 3 2 matrix A, we obtain |A| by multiplying the entries along each diagonal and subtracting.

Definition: determinant of a 2 × 2 matrix

For the 2 × 2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the **determinant of A** is given by

$$\det A = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Minors and Cofactors

The minor of an element of a determinant is the determinant obtained by deleting the row and column which intersect in that element.

For example, in

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

The cofactor of any element in a determinant is its minor with the proper sign. The sign of an element in the ith row and j th column's (− 1)i+j. The co-factor of an element is usually denoted by the corresponding capital letter.

$$\text{minor of } a_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \text{ and minor of } a_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

Thus the cofactor of a_{32} = $(-1)^{3+2}$ minor of a_{32}.

Thus, $C_{32} = - \begin{vmatrix} a11 & a13 \\ a21 & a23 \end{vmatrix}$

Similarly, the cofactor of a23 = $-(1)^{2+3}$ minor of a23.

$C_{23} = - \begin{vmatrix} a11 & a12 \\ a31 & a32 \end{vmatrix}$

Please note that

$C_{ij} = (-1)^{i+j} * M_{ij}$

where Cij is the cofactor of aij, and

Mij is the minor of aij.

Properties of Determinants

Expanding a determinant in which the elements are large number can be a very tedious task. It is possible, however, by knowing something of the properties of determinants, to simplify the working. So here are some of the main properties.

1. The value of a determinant remains unchanged if rows are changed to column and column to rows.

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

2. If two rows (or two columns) are interchanged, the sign of the determinant is changed

$$\begin{vmatrix} a_2 & b_2 \\ a_1 & b_1 \end{vmatrix} = -\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

3. If two rows (or two columns) are identical, the value of the determinant is zero.

$$\begin{vmatrix} a_1 & a_1 \\ a_2 & a_2 \end{vmatrix} = 0$$

4. If the elements of any one row (or column) are all multiplied by a common factor, the determinant is multiplied by that factors

$$\begin{vmatrix} k\,a_1 & k\,b_1 \\ a_2 & b_2 \end{vmatrix} = k\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

5. If the elements of any row (or column) are increased (or decreased) by equal multiples of the corresponding elements of any other row (or column), the value of the determinant is unchanged.

$$\begin{vmatrix} a_1 + k\,b_1 & b_1 \\ a_2 + k\,b_2 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

Cramer's Rule

If AX = B is a system of n linear equations in n unknowns such that det (A) ≠ 0, then the system has a unique solution.

$$(A = [a_{ij}]_{n\times n})$$

This solution is given by

$$X_j = \frac{\det A_j}{\det A},\ j = 1, 2, \dots, n$$

where Aj is the matrix obtained from A by replacing the entries in the j th column of A by this entries in the matrix B = [b1, b2, . . . , bn] T .

Let us consider the following simultaneous equation

$a_{11}x + a_{12}\,y + a_{13}\,z = b_1$

$a_{21}x + a_{22}\,y + a_{23}\,z = b_2$

$a_{31}x + a_{32}\,y + a_{33}\,z = b_3$

Then by Cramer's rule

$$x = \frac{\Delta_x}{\Delta}\ ,\ y = \frac{\Delta_y}{\Delta}\ ,\ z = \frac{\Delta_z}{\Delta}$$

and $$\Delta = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}$$

Example 1:Solve the following system by Cramer's rule :

2 x − y + 3 z = 2

x + 3 y − z = 11

2 x − 2 y + 5 z = 3

Solution:

By Cramer's rule, we get

$$x = \frac{\Delta_x}{\Delta}\ ,\ y = \frac{\Delta_y}{\Delta}\ ,\ z = \frac{\Delta_z}{\Delta}$$

$$\Delta = \begin{matrix} 2 & -1 & 3 \\ 1 & 3 & -1 \\ 2 & -2 & 5 \end{matrix}$$

= (15- 2) -1 (-2- 5)+ 3 (2-- 6) = 9

$$\Delta x = \begin{matrix} 2 & -1 & 3 \\ 11 & 3 & -1 \\ 3 & -2 & 5 \end{matrix}$$

= 2 (15 − 2) − 1 (− 3 − 55) + 3 (− 22 – 9)

= 26 + 58 −93

= − 9

$$\Delta y = \begin{matrix} 2 & 2 & 3 \\ 1 & 11 & -1 \\ 3 & -2 & 5 \end{matrix}$$

= 2 (55 + 3) − 2 (5 + 2) + 3 (3 – 21)

=116 − 14 −57

=45

$$\Delta z = \begin{matrix} 2 & 1 & 2 \\ 1 & 3 & 11 \\ 3 & -2 & 3 \end{matrix}$$

= 2 (9 + 22) + 1 (3 − 22) + 2 (− 2 – 6)

= 62 − 19 − = 27

x= Δx/ Δ= -9/9=-1

y= Δy/ Δ=45/9=5

z== Δz/ Δ=27/9=3

Therefore, x=-1

Y=5

Z=3

2.4 Adjoint and Inverse of a Matrix

i. The adjoint of a square matrix A = [aij] n×n is defined as the transpose of the matrix. [aij] n×n , where Aij is the co-factor of the element aij. It is denoted by adj A

$$\text{If } A=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \text{ then } adj\ A=\begin{vmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{vmatrix}, \text{ where } A_{ij} \text{ is co-factor of } a_{ij}.$$

ii. A (adj A) = (adj A) A = |A| I, where A is square matrix of order n.

iii. A square matrix A is said to be singular or non-singular according as |A| = 0 or |A| ≠ 0, respectively.

iv. If A is a square matrix of order n, then |adj A| = $|A|^{n-1}$.

v. If A and B are non-singular matrices of the same order, then AB and BA are also nonsingular matrices of the same order.

vi. The determinant of the product of matrices is equal to product of their respective determinants, that is, |AB| = |A| |B|.

vii. If AB = BA = I, where A and B are square matrices, then B is called inverse of A and is written as B = A–1. Also $B^{-1} = (A^{-1})^{-1} = A$.

viii. A square matrix A is invertible if and only if A is non-singular matrix. (ix) If A is an invertible matrix, then A^{-1} = (1 / | A |)* (adj A)

Example 03: Find out the inverse of

$$\begin{bmatrix} 1 & -1 & 2 \\ 4 & 0 & 6 \\ 0 & 1 & -1 \end{bmatrix}$$

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

A^{-1}=adj(A)/|A|

To find out the adj(A), first we have to find out cofactor(A).

a11 = -6, a12 = 4, a13 = 4

a21 = 1, a22 = -1, a23 = -1

a13 = -6, a32 = 2, a33 = 4

So, cofactor(A) = $\begin{bmatrix} -6 & 4 & 4 \\ 1 & -1 & -1 \\ -6 & 2 & 4 \end{bmatrix}$

adj(A) = $[\text{cofactor}(A)]^T$

adj(A) = $[\text{cofactor}(A)]^T = \begin{bmatrix} -6 & 4 & 4 \\ 1 & -1 & -1 \\ -6 & 2 & 4 \end{bmatrix}^T$

adj(A) = $\begin{bmatrix} -6 & 1 & -6 \\ 4 & -1 & 2 \\ 4 & -1 & 4 \end{bmatrix}$

Then, | A | = 1(0-6)+1(-4-0)+2(4-0) = -6-4+8 = -2

A^{-1}=adj(A)/|A| = $\dfrac{\begin{bmatrix} -6 & 1 & -6 \\ 4 & -1 & 2 \\ 4 & -1 & 4 \end{bmatrix}}{-2}$

$$A^{-1}=\begin{bmatrix} 3 & -1/2 & 3 \\ 2 & 1/2 & -1 \\ -2 & 1/2 & -2 \end{bmatrix}$$

2.5 Input-Output Analysis

Input- Output analysis explains the interdependence and interrelationship of inputs and outputs of various industries in the economy. It is a method of analyzing how an industry undertakes production by using the output of other industries in the economy, and how the output of the given industry is used up in other industries of the same economy.

In its "static" version, the Input- Output analysis of Professor W. Leontief, a Nobel Prize winner, deals with this particular version: What level of output should each of the n industries in an economy produce, in order that it is just be sufficient to satisfy the total demand for the product? I n view of the inter industry dependence, any set of "correct" output levels for the n industries must be one that is consistent with all the input requirements in the economy, so that no bottlenecks will arise anywhere.

I-0 analysis was an attempt made by Prof. Leontief to take account of 'general equilibrium' phenomena in the 'empirical' analysis of 'production'. These three italicized elements are the main features of I-0 analysis. First, the I-0 analysis deals almost exclusively with production. The problem is essentially technological. Given the quantities of available resources and the state of technology, the analysis is concerned with the use of various inputs by the industries and outputs derived from them.

The second distinctive feature of I-0 analysis is its devotion to empirical investigation. This is primarily what distinguishes it from the work of Walras and later general equilibrium theorists. I-0 employs a model, which is more simplified and also narrower in the sense that it seeks to encompass fewer phenomena than does the usual general equilibrium theory. Its narrowness lies in its exclusive emphasis on the production side of the economy.

The third distinctive feature is its emphasis of general equilibrium phenomena where everything depends on everything else. Thus, in the two-industry model coal is an input for steel industry and steel is an input for coal industry, though both are the output of respective industries. According to I-0 analysis, it is not possible to find some industries as being in the 'earlier' stages of production and some other industries as being in the 'later' stages. For the production of coal, steel is needed; whereas, for the production of steel, coal is required. No one can say whether the coal industry or the steel industry is earlier or later in the hierarchy of production.

The basic problem, then, is to see what can be left over for final consumption and how much of each output will be used up in the course of the productive activities which must be undertaken to obtain these net outputs. Solution of this problem can be used in predicting future production requirements if usable demand estimates can somehow be obtained. Particularly, it can be used for economic planning including problems of economic development in 'backward areas' as well as problem of military mobilization. A more modest purpose that it has already successfully begun to serve is the provision of a very illuminating detailed structure for national income accounting.

2.6 Assumptions

As it was stated earlier, the intransigence of empirical materials and the computational problems have forced on 1-0 analysis to adopt a number of simplifying assumptions even more extreme than those usually employed in our theoretical models.

The economy can be meaningfully divided into finite number of sectors (industries) on the basis of the following assumptions:

i. Each industry produces only one homogeneous output. NO two products Application are produced jointly; but if at all there is such a case, then it is assumed that they (products) are produced in fixed proportions.
ii. Each producing sector satisfies the properties of linear homogeneous production function. In other words, production of each sector is subject to constant returns to scale so that k-fold change in every input will result in an exactly k-fold change in the output.

iii. A far stronger assumption is that each industry uses a fixed input ration for the production of its output; in other words, input requirement per unit of output in each sector remains fixed and constant. The level of output in each sector (industry) uniquely determines the quantity of each input, which is purchased.

iv. The final demand for the commodities is given from outside the system. The total amount of the primary factor is also given. These two are called open end of the system and for this, the model is called 'open model'. In contrast in the 'closed model' all the variables are determined within the system.

v. The model is static in the sense that all variables in it refer to the same period of time. The static model is also known as flow model.

2.7 Structure

For the purpose of simplification, let us consider an economy with only two producing sectors: agriculture and industry. If we concentrate on agriculture, it will produce a given output of goods in a given time period, say, one year. These goods will be used in various ways. In other words, agricultural products will have several destinations. These destinations are broadly classified into two groups: (a) output going as intermediate input to the agricultural sector itself and to the industrial sector and (b) output going to the final demand sector for final consumption. For example, wheat is an output of agricultural sector and it may be demanded (i) as input in the agricultural sector itself (in the form of seed) for further production of wheat; and in industry for the manufacture of bread and (ii) for final consumption by the final demand sector. In the same manner, the output of the industry may be demanded by agricultural or industrial sectors as intermediates input and by final demand sector for final consumption.

In addition to the intermediate (also known as secondary) inputs, each sector requires primary inputs. This primary input is in the form of services of factors like land, labour, capital and entrepreneurship and is supplied by the final demand sector, also known as household sector.

Let us consider an input-output table involving agriculture and industry. Such a table is often called an input-output transactions matrix.

Table 1: Input-Output Transactions Matrix

Producing Sector	Purchasing Sector		Final Demand (Consumption)	Total Output
	1 (Agriculture)	2 (Industry)		
1 (Agriculture) 2 (Industry)	X_{11} X_{21}	X_{12} X_{22}	d_1 d_2	X_1 X_2
Primary Input (labour)	L_1	L_2		

In the above input-output transaction matrix, Xij (i = 1, 2; j = 1, 2) denote the output of ith producing sector that is being used as intermediate input in j th producing sector. In the above table, agricultural sector is denoted by 1 and industrial sector by 2. Thus, we can interpret X11 as the output of agricultural sector that is being used as an intermediate input in agriculture. Similarly, X21 can be interpreted as the output of industrial sector that is being used as input in agriculture etc. In addition, the output for each producing sector goes for the final use also. Thus, d1 is the output of agriculture and d2 is the output of industry going for the final use (consumption). Further, X1 and X12 denote the total output of agriculture and industry respectively. In view of the above introduction, we can write

$X_{11} + X_{12} + d_1 = X_1$ (1)

$X_{21} + X_{22} + d_2 = X_2$ (2)

Finally, the elements in the last row of the above table i.e. , L1 and L2 denote the requirement of primary input (say, labor) by the respective sector. These primary inputs, as mentioned earlier, are presumed t be supplied by the household sector. It is clear that our input-output table is an oversimplified representation. However, the table can readily be made more realistic by introducing more producing sectors, additional kinds of final uses (say, investment, government expenditure and net exports) and other categories of primary inputs (that is, land, capital and entrepreneurship).

Let us denote this amount by a_{ij}, where $a_{ij}=\frac{xij}{xj}$. If xij and xj denote value of outputs, then aij can be interpreted in value terms. Thus a_{ij}, = 0.04 can mean that forty paise worth of the ith commodity is required to produce onerupee worth of the jth commodity. This a_{ij}, is called an input coefficient.

We can arrange the input-coefficients of an economy with a given number of producing sectors in the form of a matrix. This matrix is called an input coefficient matrix.

We should note that the sum of elements in each column gives the requirement of secondary input to produce a rupee worth of output in that producing sector. As a result, the sum of the elements in each column of the input coefficient matrix should be less than 1, since; it does not include the cost of the primary inputs per rupee worth of the output. Assuming that there is pure competition with free entry, the primary input cost per one rupee worth output for a producing sector should be one minus the relevant column sum of the elements of the input coefficient matrix. For the economy considered above, if the cost of primary inputs per rupee worth of output for the two producing sectors are l1 and l2 respectively; then

$l_1 = 1- (a_{11} + a_{21})$

and

$l_2 = 1- (a_{12} + a_{22})$

The total output for each producing sector can be expressed in terms of the input coefficients, by replacing xj aij= xij in equations (1) and (2). Thus, for out two producing sector economy

$X_1=a_{11}X_1+a_{12}X_2+a_{12}X_2+d_1$

or

$X_1-a_{11}X_1-a_{12}X_2=d_1$

or

$(1-a_{11})X_1-a_{12}X_2=d_1$ (3)

and similarly

$X_2=a_{21}X_1+a_{22}X_2 +d_2$

or

$X_2-a_{21}X_1-a_{22}X_2 =d_2$

or

$-a_{21}X_1+(1-a_{22})X_2-a_{12}X_2- a=d_1$ (4)

Writing equations (3) and (4) in the matrix form

$$\begin{pmatrix} 1-a11 & -a12 \\ -a21 & 1-a22 \end{pmatrix} \begin{pmatrix} X1 \\ X2 \end{pmatrix} = \begin{pmatrix} d1 \\ d2 \end{pmatrix}$$

Suppose

$$A= \begin{pmatrix} a11 & a12 \\ a21 & a22 \end{pmatrix}, x= \begin{pmatrix} X1 \\ X2 \end{pmatrix} \text{and } d= \begin{pmatrix} d1 \\ d2 \end{pmatrix} \quad (5)$$

$(I - A) X = d$ (6)

Where I is a (2×2) identity matrix. The matrix (I-A) is called the technology matrix. If (I -A) is non-singular (that is, $|I - A| \neq 0$, equation (6) can be solved for X. Thus

$X= (I - A)^{-1} d$ (7)

It is important to note that for a given technology, as embodied in the technology matrix, equation (7) can be used to determine the total output that is needed to be produced by different producing sectors to satisfy a given final demand for the commodities.

2.8 Closed Model

In the closed model, all the goods are intermediate in nature, for everything that is produced, is produced for the sake of satisfying the input requirements of the other industries of the model.

Let,

aij = Required minimum input of commodity i, per unit of output of commodity j (the first subscript refers to the input, and the second to the output),)

aij = In order to produce each unit of jth commodity, the input need for the ith commodity)

aij = How much of the ith commodity be used for the production of each unit of the jth commodity)

aij = In order to produce each unit of the jth commodity, the input need for the ith commodity must be a fixed amount)

aij = "Input Coefficient" which is assumed to be fixed.

For an n industry economy, the input coefficients can be arranged into a matrix, A = [aij]; as shown in the table 2, in which each column specifies the input requirements for the production of one unit of the output of a particular industry.

It is to note that if no industry uses its own product as an input, then the elements in the principal diagonal of matrix A will all be zero.

Table 2: : Input Coefficient Matrix (closed model)

	Output				
Input ↓	I	II	III	...	N
I	a_{11}	a_{12}	a_{13}	...	a_{1n}
II	a_{21}	a_{22}	a_{23}	...	a_{2n}
III	a_{31}	a_{32}	a_{33}	...	a_{3n}
...	...	...	...	...	...
N	a_{n1}	a_{n2}	a_{n3}	...	a_{nn}

Here, in a static model, it is to observe that each column sum represents the final input cost incurred in producing a Rupee's worth of some commodity. Symbolically,

$$\sum_{i=1}^{n} aij = 1; (j = 1,2,3, \dots \dots . . n)$$

where the summation is over i, that is, over the elements appearing in the various rows of a specific column j. If industry I is to produce an output just sufficient to meet the input requirements of the n industries of the closed sector, its output level x1 must satisfy the following equation:

$x_1 = a_{11}x_1 + a_{12}x_2 + \dots\dots\dots\dots\dots + a_1nxn$

That is in general,

$x_1=a_{11}x_1+a_{12}x_2+\dots\dots\dots\dots\dots+a_1nxn=\sum_{i=1}^{n} aij.xj; = 1; (j = 1,2,3, \dots \dots . . n)$

where aijxj represents the input demand from the jth industry; and $\sum_{i=1}^{n} aij.xj$ represents the total amount of xi needed as input for the n industries.

2.9 Coefficient Matrix and Open Model

Our open model in matrix notation is given by, X = AX + F, where A is the input coefficient matrix, F is the final demand vector and X is the total output matrix. The input coefficient matrix or 'technology matrix' represented by [a,,] is of great importance. Each element must be non-negative, i.e., we rule out the possibility of negative inputs. But to maintain complete interdependence among the industries each element of [a,,] matrix must be positive and no element can exceed unity, i.e., we rule out the possibility of negative outputs. Each column of this matrix specifies the input requirements for the production of 1 unit of a particular commodity; thus, sum of the elements in each column must be less than unity. Symbolically, this fact may be stated as: $\sum_{i=1}^{n} aij < 1; (j = 1,2,3,\dots\dots..n)$ and each a,, is non-negative, i.e., either zero or 1=1 greater than zero. The cost of the primary inputs (which is also termed as 'value added') needed in producing a unit of jth commodity would be (1-$\sum_{i=1}^{n} aij$)

If this wen not true, it would mean that the total value of intermediate products used by an industry exceeded the value of its input. This in turn would mean that the value added by that industry was negative. Now, this is not impossible, but, if we assume that the wage bill cannot be negative, it means that the industry must be making losses (indeed, losses greater in absolute value than its wage bill). An industry in which value added is negative is not covering variable costs (intermediate inputs plus the wage bill), and we know from the elementary micro theory that in such a case, losses will be reduced by closing down. Thus, we do not want to describe such an industry in our technology matrix at all. Given the assumption of CRS, we've described the technology by a constant coefficient matrix. We should notice that we have also built in the assumption that there are no externalities. An externality in production would exists if, for example, a factory discharged waste into a river so that a factory further downstream had to use the resources to clean the water before it could use it. In this case, the resource requirement of the later factory would not depend solely on its output but would, also depend on the activity of the former.

2.10 Solution to Open Model

If the n industries in table 1 constitute the entirety of the economy, then all their products would be for the sole purpose of meeting the input demand of the same n industries (to be used in further production) as against the final demand (such as consumer demand, not for further production). At the same time, all the inputs used in the economy would be in the nature of intermediate inputs (those supplied by the n industries) as against primary inputs (such as labour, not an industrial product).

To allow for the presence of final demand and primary inputs, we must include in the model an open sector outside the n industry network.

Here, in a open model, it is to observe that each column sum represents the partial input cost (not including the cost of primary inputs) incurred in producing a Rupee's worth of some commodity. Symbolically,

$$\sum_{i=1}^{n} aij < 1; (j = 1,2,3,\dots\dots..n)$$

where the summation is over i, that is, over the elements appearing in the various rows of a specific column j. Thus, the value of primary inputs needed in producing a unit of the jth commodity should be (1-$\sum_{i=1}^{n} aij > 0$)

	Output					Final
Input ↓	I	II	III	...	N	Demand
I	a_{11}	a_{12}	a_{13}	...	a_{1n}	d_1
II	a_{21}	a_{22}	a_{23}	...	a_{2n}	d_2
III	a_{31}	a_{32}	a_{33}	...	a_{3n}	d_3
...	...	...	...	...	...	...
N	a_{n1}	a_{n2}	a_{n3}	...	a_{nn}	d_n
Primary Input	a_{01}	a_{02}	a_{03}	...	a_{0n}	–

If industry I is to produce an output just su¢ cient to meet the input requirements of the n industries as well as the Önal demand of the open sector, its output level x1 must satisfy the following equation:

$x_1 = a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots\ldots + a_{1n}x_n + d_1)$

$=d_1 = x_1 - (a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{1n}x_n)$

$=(1- a_{11})$ x1 $-a_{12}x_2 - a_{1n}x_n = d_1$

By the same token, the output levels of all the n industries should satisfy the following system of linear equations:

$-a_{21}x_1 -(1- a_{22})x_2 -\ldots\ldots\ldots\ldots\ldots\ldots\ldots- a_{2n}x_n = d_2$

$-a_{31}x_1 -a_{32}x_2- (1- a_{33})x_3 -\ldots\ldots\ldots\ldots\ldots\ldots\ldots- a_{2n}x_n = d_2$

$-a_{n1}x_1 -a_{n2}x_2 -\ldots\ldots\ldots\ldots\ldots\ldots(1- a_{nn})xn = d_n$

In matrix form, this may be written as:

In the matrix notation this may be written as:

$$\begin{pmatrix} 1-a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & 1-a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix}$$

(I - A) X = F,

i.e X= $(I - A)^{-1}$ F

Here A is the given matrix of input coefficients, while X and F are the vectors of output and final demand of each producing sector. If II-Al* 0, then [I-A]-' exists, we can then estimate for either of the two matrices X and F by assuming one of them to be given exogenously. It is to be observed that, the assumptions made in 1-0 analysis go a long way in making the problem simplified. For example, with the assumption of linear homogenous function, it is possible to write a linear equation of each producing sector, which then can be easily transformed into matrix notation. On the other hand, as long as the input coefficients remain fixed (as assumed), the matrix A will not change or [I-A] will not change. Therefore, in finding the solution of X = [LA]-' F, only one matrix inverse needs to be performed even if we are to consider thousands of different finals, demand vectors according to alternative development targets. Hence, such an assumption of fixed technical coefficient has meant considerable savings in computational effort.

2.11 Hawkin Simon conditions

Many a times I-0 solution may give output expressed by negative numbers. If our solution gives negative outputs, it means that more than one unit of a product is used up in the production of every one unit of that product; it is definitely an unrealistic situation. Such a system is not a viable. Hawkins Simon condition guards against such eventualities. Our basic equation is X = [I-A]-1 F, in order that this does not give negative numbers as a solution, the matrix [I-A], which in fact is

$$\begin{pmatrix} 1-a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & -a_{22} & \cdots & -a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 1-a_{nn} \end{pmatrix}$$ hould be such that

i. the determinant of the matrix must always be positive, and

ii. the diagonal elements: $(I-a_{11})$, $(1-a_{22})$,. ... $(I-a_{nn})$ should all be positive or, in other words, elements, a1 1, a22,...,a_{nn} should all be less than one. Thus, one unit of output of any sector should use not more than one unit of its own output. These are called Hawkins-Simon conditions. Further, the first condition, that implies D > 0, implies that (for 2 industry case) $\begin{pmatrix} 1-a11 & -a12 \\ -a21 & 1-a22 \end{pmatrix} > 0$, or,

(1-a11)(1-a22)-a12a21>0. This condition implies that the direct and indirect requirement of any commodity to produce one unit of that commodity must also be less than one. On the other hand, the interpretation is always that all subgroups of commodities should be 'self-sustaining', directly and indirectly.

Example 1: Suppose [A]= $\begin{pmatrix} 0.2 & -0.2 \\ -0.9 & 0.3 \end{pmatrix}$

Then [I-A]= $\begin{pmatrix} 0.2 & -0.2 \\ -0.9 & 0.3 \end{pmatrix}$

and the value of the determinant [I-A] = (-) 8.12, which is less than zero. As the Hawkins-Simon conditions are not satisfied, no solution will be possible in this case.

2.12 General Equilibrium: 3-Industry case

- Simplest case: Suppose we divide the economy into 3 sectors:

1. Agriculture
2. Manufacturing
3. Services

- The three industries each use inputs from two sources:

1. Domestically produced commodities form the three industries
2. Other inputs, such as imports, labour, and capital.

- The outputs of the industries have two broad uses or destinations:

1. Inputs to production of the three industries (intermediate inputs)
2. Final demand (Consumption, Investment, Government expenditure, Exports)

	Output					
Input		Agriculture	Manufactures	Services	Final demand	Total
	Agriculture	30	40	0	30	100
	Manufactures	10	200	50	140	400
	Services	20	80	200	200	500
	Final demand	40	80	250	230	600
	Total	100	400	500	600	1600

Take for example manufacturing:

- Its output is worth 400 crore, which is allocated as follows:
  - ❖ 10crore is used by the agricultural sector
  - ❖ 200 crore as intermediate goods for the manufacturing sector
  - ❖ 50 crore is used by the services sector.
  - ❖ 140 crore is the final demand (consumption, investment, government expenditure & exports
- In order to produce, the manufacturing sector uses inputs worth of e400, of which
  - ❖ 40crore comes from the agricultural sector,
  - ❖ 200crore from the manufacturing sector (intermediate inputs),
  - ❖ 80 crore from the services sector,
  - ❖ 80 crore from other sources, including imports, labour and capital

Define the following 2 vectors

b= $\begin{bmatrix} 30 \\ 140 \\ 200 \end{bmatrix}$ is the vector of final demands for output of the industry sectors

x= $\begin{bmatrix} 100 \\ 400 \\ 500 \end{bmatrix}$ is the vector of total ouput of the industry sectors

As said above, one (critical) assumption is that each sector produces according to fixed proportion technological coefficients (also called input-output coefficients).

2.13 Application of Leontief Inverse

Example 1: Agriculture uses 20 crore from the services sector. Given that the value of its total inputs is 100 crore, then services represent 20/100 =0:20 of its total inputs.

	Output					
Input		Agriculture	Manufactures	Services	Final demand	Total
	Agriculture	3/10	1/10	0	1/20	1/16
	Manufactures	1/10	1/2	1/10	7/30	1/4
	Services	1/5	1/5	2/5	1/3	5/16
	Final demand	2/5	1/5	1/2	23/60	3/8
	Total	1	1	1	1	1

In matrix form:

A= $\begin{bmatrix} 3/10 & 1/10 & 0 \\ 1/10 & 1/2 & 1/10 \\ 1/5 & 1/5 & 2/5 \end{bmatrix}$ is the matrix of inter-industry coefficients

One important consequence of the input-output analysis is that we can express the vector of total demand (x) as a function of the final demand (b) and the matrix of inter-industry Coefficients (A):

$x = Ax + b$

Then:

$x - Ax = b$

$(I - A)\, x = b$

If (I - A) has an inverse:

$(I - A)^{-1}(I - A)\, x = (I\ \ A)^{-1}b$

$x = (I - A)^{-1}b$

Thus, The matrix (I- A)-1 is known as the input-output inverse, or the Leontieff Inverse

$(I - A)^{-1} = \begin{bmatrix} 0.7 & -.1 & 0 \\ -0.1 & 0.5 & -0.1 \\ -0.2 & -0.2 & 0.6 \end{bmatrix}$

Thus, $x = (I\ \ A)^{-1}b$

X= $\begin{bmatrix} 1.49 & 0.32 & 0.05 \\ 0.42 & 2.23 & 0.37 \\ 0.64 & 0.85 & 1.81 \end{bmatrix} \begin{bmatrix} 1/20 \\ 7/30 \\ 1/3 \end{bmatrix}$

$$X=\begin{bmatrix}1.49*\frac{1}{20}+0.32*\frac{7}{30}+0.05*1/3\\0.42*\frac{1}{20}+2.23*\frac{7}{30}+0.37*1/3\\0.64*\frac{1}{20}+0.85*\frac{7}{30}+1.81*1/3\end{bmatrix}$$

$$X=\begin{bmatrix}0.165\\0.664\\0.833\end{bmatrix}$$

2.14 Determination of Equilibrium Prices

Let the prices of commodities 1, 2, 3,. .. be pl, p2, p3 ,... respectively, and the price of the primary factor inputs be w (here, primary factor is labor; so w represents wage rate), then the technology matrix or transaction matrix in quantity may be converted into that in value terms. The problem can be posed as follows:

Sector	1	2	Final demand
1 2	$a_{11}X_1p_1$ $a_{21}X_1p_2$	$a_{12}X_2p_1$ $a_{22}X_2p_2$	F_1p_1 F_2p_2
Primary Input	$l_1X_1\,w$	$l_2X_2\,w$	
Total Costs	$a_{11}X_1p_1+a_{21}X_1p_2+l_1X_1\,w$	$a_{12}X_2p_1+a_{22}X_2p_2+l_2X_2\,w$	

With pure competition and free entry, profit in each industry must be zero, i.e., revenues equal costs. Hence, for the first industry receipts are (output x I price) and cost is allXlpl + a2lXlp2 + I1x1 w. Same is true for the second industry.

Hence, for equilibrium; $p_1X_1 = a_{11}X_1p_1 + a_{21}X_1p_2 + I_1X_1w$; and $p_2X_2 = a_{12}X_2p_1 + a_{22}X_2p_2 + {}_{12}X_2w$, which simplify to $p_1 = a l_1 p_1 + a_{21}p_2 + l_1 w$; and $p_2X_2 = a_{12}p_1\,I + a_{22}p_2 + 1_2W$, which can be put in matrix form as under.

$$\begin{pmatrix}1-a11 & -a12\\ -a21 & 1-a22\end{pmatrix}\begin{pmatrix}p1\\p2\end{pmatrix}=\begin{pmatrix}I1W\\I2W\end{pmatrix}$$

Notice that the set of coefficients here are transposed, this matrix is transposed of [I-A].

$$\begin{pmatrix}p1\\p2\end{pmatrix}=\begin{pmatrix}I1W\\I2W\end{pmatrix}\begin{pmatrix}1-a11 & -a12\\ -a21 & 1-a22\end{pmatrix}^{-1}$$

Therefore, $p_1 = 1/D(A_{11}I_1+ A_{12}I_2)w$

$P_2 = 1/D(A_{21}I_1+ A_{22}I_2)w$

where A1 I, f D D A12, etc., are the cofactors of the matrix [I-A] as in the preceding cases.

2.15 Uses of Input–Output Analysis

Input–output analysis has a wide range of applications:

(1) Input–output analysis is used to obtain projections of demand, output, employmentand investment for a country or region.

(2) Input–Output analysis is helpful in providing necessary information for formulationof economic policies. It is used in economic development planning, studies of inter-regional and international economic relationships.

(3) Input–Output analysis is useful for national income accounting as it providesdetailed breakdown of the macro aggregates and the money flows.

(4) The inter relations between various sectors, as revealed in the input–output table, provide indication regarding prospective trends in which they are likely to combine with each other.

(5) Given a certain final output target, it can show the production requirements of various sectors.

2.16 Limitations of Input–Output Analysis

Despite its many useful applications, the input–output analysis has many shortcomings

which are given below:

(i) The input–output analysis is based on the assumption of fixed input co-efficients or proportions. Over time, technology and input prices change, and these are likely to greatly affect the proportions in which inputs are combined in the production of many commodities. This necessitates frequent and continuous updatings of input–output tables, which are very costly and time consuming.

(2) Input–output analysis is based on linear equations relating outputs of one industry to inputs of the others. This appears unrealistic

(3) In the input–output analysis labour is the only input which is considered scarce. This is not true in practice.

(4) Final demand (i.e, the purchases by consumers and the government) in the input–output analysis is taken as given and treated as independent of the production sector.

Though input–Output analysis has its short–comings, yet it is considered an important tool for the analysis of economic decisions of the government and in development planning.

Summary

- A system of mnnumbers arranged in a rectangular formation along m rows and n columns and bounded by the brackets [ ] is called an m by n matrix : when m = n, it is called a square matrix. To locate any particular element of a matrix, the elements are denoted by a letter followed by two suffixes, which respectively specify the rows and the columns. This aij is the element in the ith row and j th column of the matrix A. In the notation, the matrix A is denoted by [aij].
- If A be any matrix, then a matrix B, if it exists, such that AB = BA = I is called the inverse of A.Inverse of a square matrix A-1 =Adj A/ | A |
- Ax=λx says that eigenvectors x keep the same direction when multiplied by A.
- Ax=λx also says that det.(A-λI)= 0. This determines n eigenvalues.
- The eigenvalues of A^2 and A^{-1} are λ^2 and λ^{-1}, with the same eigenvectors.
- The sum of theλ 's equals the sum down the main diagonal of A (the trace).The product of the 's equals the determinant.

Keywords

Matrix: A rectangular array of numbers (elements).

Row Matrix: One row only.

Column Matrix: One column only.

Equal Matrices: Corresponding elements equal.

Diagonal Matrix: All elements zero except those on the leading diagonal.

Null Matrix: All elements zero

Characteristic Roots and vector: eigen value and eigen vector

Self Assessment

1. Let A be a square matrix of order 3 × 3, then | kA| is equal to

A. k| A |

B. k2 | A |

C. k3| A |

D. 3k | A |

2. Which of the following is correct?

A. Determinant is a square matrix.

B. Determinant is a number associated to a matrix.

C. Determinant is a number associated to a square matrix.

D. None of these

3. Determinant is a number associated to a matrix.

A. True

B. False

4. A matrix with only one column is called:

A. A Null matrix.

B. A row matrix.

C. Homogeneous matrix.

D. None of the above

5. If Ax = b is a system of n linear equations in n unknowns such that det (A)$\neq$ 0, then the system has:

A. Infinitely many solutions.

B. Unique solution.

C. Both (a) and (b).

D. None of the above.

6. Transpose of a rectangular matrix is a:

A. Rectangular matrix.

B. Diagonal matrix.

C. Square matrix.

D. Scalar matrix

7. If |A|=0, then A is:

A. Zero matrix.

B. Singular matrix.

C. Non-singular matrix.

D. None of the above.

8. For a non-trivial solution |A| is:

A. |A| > 0

B. |A| < 0

C. |A| = 0

D. None of the above.

9. Two vectors are linearly dependent if and only if they lie:

A. On a line parallel to x-axis.

B. On the same line through origin

C. On a line parallel to y-axis.

D. None of the above.

10. A basis is a linearly independent set that is as large as possible:

A. True.

B. False.

C. May be.

D. None of the above.

11. The scalar λ is characteristic root of the matrix A if:

A. λI) is singular

B. λI is non -singular

C. A is singular.

D. None of the above.

12. If eigenvalue of matrix A is λ, then eigenvalue of A2 is:

A. 1

B. 1/λ

C. λ2

D. None of the above.

13. If A is invertible matrix and eigenvalue of A is λ, then eigenvalue of A^{-1} is:

A. 1

B. 1/ λ

C. λ2

D. None of the above.

14. In matrices, the inter-industry demand is summarized as

A. Input-Output Matrix

B. Output-input matrix

C. Linear buying matrix

D. Linear selling matrix

15. In input-output analysis, if the exogenous sectors of the open input output model is absorbed in to the system as just another sector _____

A. The transaction matrix

B. A technology coefficient

C. Leontief closed model

D. None of the above

Answers for Self Assessment

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

11. A 12. C 13. B 14. A 15. C

Review Questions

1. Write each sum as a single matrix:

A. $A=\begin{bmatrix} 5 & 1 & 4 \\ 3 & -1 & 9 \end{bmatrix}$ and $B=\begin{bmatrix} 1 & -1 & 0 \\ -2 & 1 & 0 \end{bmatrix}$

B. $A=\begin{vmatrix} 1 & 3 & 0 \end{vmatrix}$ and $B=\begin{vmatrix} 4 & -3 & 1 \end{vmatrix}$

2. Write each product as a single matrix:

A. $\begin{bmatrix} 2 & 5 & 4 \\ 3 & 6 & 0 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 8 & 4 \\ 3 & 1 \end{bmatrix}$

B. $\begin{vmatrix} 3 & -2 & 5 \end{vmatrix}\begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$

3. Solution by matrix inverse method

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

4. Determine whether the given set of vectors is linearly independent or linearly dependent in Rn. In the case of linear dependence, find a dependency relationship.

A. {(1, −1), (1, 1)}.

B. {(2, −1), (3, 2), (0, 1)}.

C. {(1, −1, 0), (0, 1, −1), (1, 1, 1)}

5. Evaluate the eigenvalues and eigenvectors of the following matrices:

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

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

6. A 3 by 3 matrix B is known to have eigenvalues 0, 1,2. This information is enoughto find three of these (give the answers where possible):

   A. the rank of B

   B. the determinant of B^TB

   C. the eigenvalues of B^TB

   D. the eigenvalues of $(B^2+ I)^{-1}$.

Further Readings

- Mathematics for Economics-Council for economic education
- Essential Mathematics for Economists- Nutt Sedester, peter Hawmond, Prentice Hall Publication
- Mathematics for Economists- Carl P Simone, Lawrence Bloom.
- Chiang Alpha C. (1984), Fundamental Methods of Mathematical Economics, Third edition, McGraw-Hill Book Company.
- Henderson J.M. &Quandt R.E, Microeconomic Theory: A Mathematical Approach, Third edition, McGraw-Hill Book Company.
- Chakravorty, J.G & P.R. Ghosh (1996), Higher Algebra including Modern Algebra, Twelfth Edition, U.N. Dhur& Sons Pvt. Ltd. Calcutta

Unit 03: Function

CONTENTS

Objectives

After studying this unit, you will be able to,

- Know the Functions and Related Quantities.
- Determine the Value of Functions.
- Know the Definition of Functions by Mapping.
- Determine the Domain and Range of Functions.
- Understand the use of Functions in Economics.

Introduction

In many practical situations, the value of one quantity may depend on the value of a second. For example, the consumer demand for beef may depend on the current market price; the amount of air pollution in a metropolitan area may depend on the number of cars on the road; or the value of a rare coin may depend on its age. Thus, volume of a cylinder depends on its radius, area of square and volume of a cuboid depend on the length of its arm. The distance covered by a running train in dynamic velocity depends on the time taken. Velocity of a falling particle depends on the distance covered by it. Atmospheric pressure of a certain place depends on the alleviation of its height from sea-coast etc. Such relationships can often be represented mathematically as functions.

Suppose that average weekly household expenditure on food (C), measured in rs., depends on average net household weekly income (Y) according to the relationship

$C = 42 + 0.2Y$

For any given value of Y, one can evaluate what C will be. For example

if $Y = 500$

then expenditure on food is $C = 42 + 0.2(500) = 42 + 100 = \text{Rs. } 142$

Whatever value of Y is chosen there will be one unique corresponding value of C. This is an example of a function.

A relationship between the values of two or more variables can be defined as a function when a unique value of one of the variables is determined by the value of the other variable or variables.

If the precise mathematical form of the relationship is not actually known then a function may be written in what is called a general form. For example, a general form demand function is

$Q_d = f(P)$

This just tells us that quantity demanded of a good (Q_d) depends on its price (P). The 'f' is not an algebraic symbol in the usual sense and so f(P) means 'is a function of P' and not 'f multiplied by P'. In this case P is what is known as the 'independent variable' because its value is given and is not dependent on the value of Q_d, i.e. it is exogenously determined. On the other hand, Q_d is the 'dependent variable' because its value depends on the value of P.

Loosely speaking, a function consists of two sets and a rule that associates elements in one set with elements in the other. For instance, suppose you want to determine the effect of price on the number of units of a particular commodity that will be sold at that price. To study this relationship, you need to know the set of admissible prices, the set of possible sales levels, and a rule for associating each price with a particular sales level. Here is the definition of function we shall use.

3.1 Definition of Function

A function is a correspondence between a first set, called the domain, and a second set, called the range, such that each member of the domain corresponds to exactly one member of the range.

A (real-valued) function of a real variable x with domain D is a rule that assigns a unique real number to each real number x in D. As x varies over the whole domain, the set of all possible resulting values f (x) is called the range of f.

Functions are given letter names, such as f, g, F, or ϕ. If f is a function and x is a number in its domain D, then f (x) denotes the number that the function f assigns to x. The symbol f (x) is pronounced "f of x", or often just "f x". It is important to note the difference between f, which is a symbol for the function (the rule), and f (x), which denotes the value of f at x.

If f is a function, we sometimes let y denote the value of f at x, so y = f (x) Then we call x the independent variable, or the argument of f, whereas y is called the dependent variable, because the value y (in general) depends on the value of x. The domain of the function f is then the set of all possible values of the independent variable, whereas the range is the set of corresponding values of the dependent variable. In economics, x is often called the exogenous variable, which is supposed to be fixed outside the economic model, whereas for each given x the equation y = f (x) serves to determine the endogenous variable y inside the economic model.

A function is often defined by a formula such as $y = 8x^2 + 3x + 2$. The function is then the rule $x \rightarrow 8x^2 + 3x + 2$ that assigns the number $8x^2 + 3x + 2$ to each value of x.

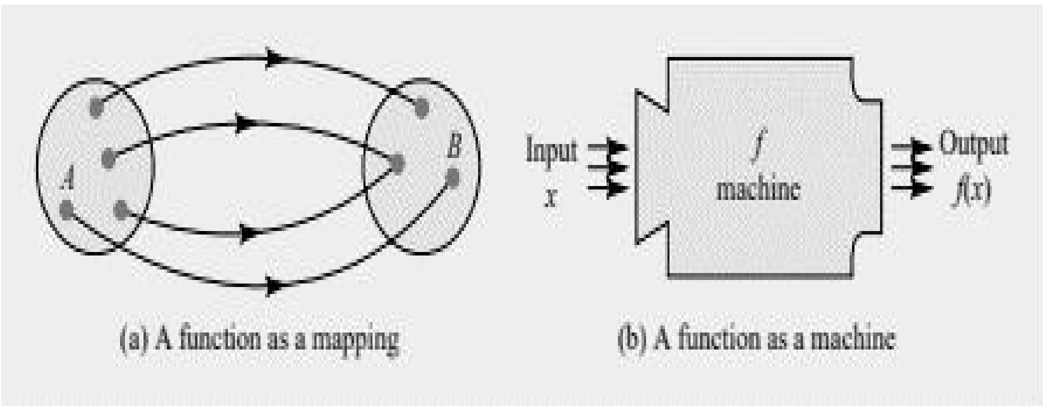


Fig.1: Interpretations of Function

It may help to think of such a function as a "mapping" from numbers in A to numbers in B (Figure 1 a), or as a "machine" that takes a given number from A and converts it into a number in B through a process indicated by the functional rule (Figure 1b). For instance, the function f(x)= x2 + 4 can be thought of as an "f machine" that accepts an input x, then squares it and adds 4 to produce an output y = x 2 + 4. No matter how you choose to think of a functional relationship, it is important to remember that a function assigns one and only one number in the range (output) to each number in the domain (input).

Example 1.The total dollar cost of producing x units of a product is given by

$C(x) = 100x\sqrt{x} + 500$

for each nonnegative integer x. Find the cost of producing 16 units.

Solution: The cost of producing 16 units is found by substituting 16 for x in the formula for C(x): $C(16) = 100 \cdot 16\sqrt{16} + 500 = 100 \cdot 16 \cdot 4 + 500 = 6900$

Example 2.The squaring function is given by

$F(x)=x^2$

Find f(-3), f(1).

Solution: $f(-3)=(-3)^2=9$

$f(1)=1^2=1$

Example 3:A function is given by $f(x)= 3x^2-2x+8$. Find f(0), f(-5) and f(7a).

Solution: One way to find function values when a formula is given is to think of the formula with blanks, or placeholders, as follows:

$f(0)= 3.0^2-2.0+8=8$

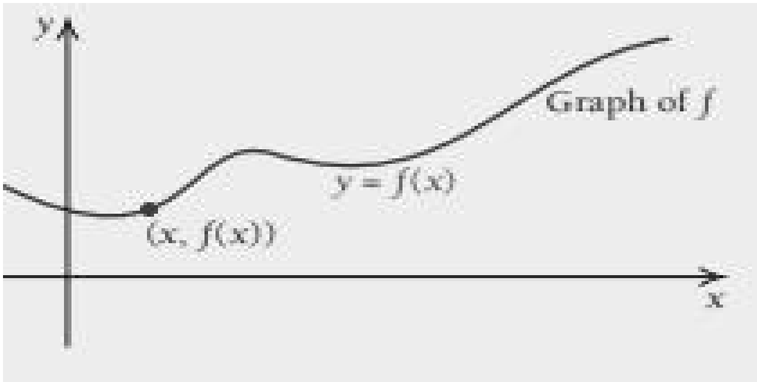
$f(-5)= 3(-5)^2-2(-5)+8= 3.25+10+8=75+10+8=93$

$f(7a)= 3(7a)^2-2(7a)+8= 3.49a^2-14a+8=147a^2-14a+8$

Graphs of Functions

Consider again the squaring function. The input 3 is associated with the output 9. The input–output pair (3, 9) is one point on the graph of this function.

The graph of a function is a drawing that represents all the input–output pairs In, cases where the function is given by an equation, the graph of the function is the graph of the equation y = f(x).



It is customary to locate input values (the domain) on the horizontal axis and output values (the range) on the vertical axis.

3.2 Domain and Range

The definition of a function is not really complete unless its domain is either obvious or specified explicitly. If a function is defined using an algebraic formula, the domain consists of all values of the independent variable for which the formula gives a unique value (unless another domain is explicitly mentioned).

If function f: X→Y is defined, then non-empty set Wis called domain of this mapping or function (f) and non-element set(Y) would be its range or co-domain.

Domain of function (f)= Element of actual values of X for which function is defined.

Range of function (f)**=** values of elements of Y corresponding to domain of all numbers x.

Let f be a function with domain D. The set of all values f (x) that the function assumes is called the range of f. Often, we denote the domain of f by Df, and the range by Rf. These concepts are illustrated in Fig. 3, using the idea of the graph of a function. (Graphs are discussed in the next section.) Alternatively, we can think of any function f as an engine operating so that if x in the domain is an input, the output is f (x). (See Fig.3.) The range of f is then all the numbers we get as output using all numbers in the domain as inputs. If we try to use as an input a number not in the domain, the engine does not work, and there is no output.

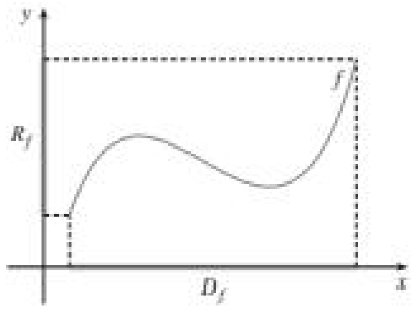


Fig. 3. Domain (D_f) and Range (R_f) of a function

Domain calculation: Steps to be followed

1. Denominator must be non-zero as f(x)=1/(x-2), x≠2
2. Even roots cannot have negative number: f(x)=√x-3, as it makes function undefined.

Example 1:Determine the domain of x−4/ (x^2 −2x−15)

Solution: Set the denominator to zero and solve for x

⟹ x^2 − 2x – 15 = (x − 5) (x + 3) = 0

Hence, x = −3, x = 5

For the denominator not to be zero, we need to avoid the numbers −3 and 5. Therefore, the domain is all real numbers except −3 and 5.

Example 2:Find the domain and range of the following function.

f(x) = 2/ (x + 1)

Solution:Set the denominator equal to zero and solve for x.

x + 1 = 0

x= -1

Since the function is undefined when x = -1, the domain is all real numbers except -1.

Example 3: Find the domain of f(x) = 1/√ (x2 -4)

Solution: By factoring the denominator, we get x ≠ (2, – 2).

Test your answer by plugging -3 into the expression within the radical sign.

⟹ (-3)2 – 4 = 5

also try with zero

⟹ 02 – 4 = -4, therefore number between 2 and -2 are invalid

Try number above 2

$\Rightarrow$ 32 – 4 = 5. This one is valid.

Hence, the domain = $(-\infty, -2) \cup (2, \infty)$

Range calculation: Steps to be followed:

1. Put y=f(x)
2. Express x as a function of y
3. Find possible values for y
4. Eliminate values according to definition of range (Applied only in case of underoot).

Example 4: Find the range of function f defined by:

$F(x)=x^2-2$

Solution: The domain of this function is the set of all real numbers. The range is the set of values that f(x) takes as x varies. If x is a real number, x2 is either positive or zero. Hence, we can write the following:

$x^2 \geq 0$

Subtract - 2 to both sides to obtain

$x^2 - 2 \geq - 2$

The last inequality indicates that $x^2 - 2$ takes all values greater than or equal to - 2. The range of f is given by

$[-2, +\infty)$

Example 5:Find the Range of function f defined by

$f(x) = 4x + 5$

Solution:

Assuming that the domain of the given function is the set of all real numbers R, so that the variable x takes all values in the interval

$(-\infty , +\infty)$

If x takes all values in the interval $(-\infty, +\infty)$ then 4 x + 5 takes all values in the interval $(-\infty, +\infty)$ and the range of the given function is given by the interval $(-\infty, +\infty)$.

Example 6: Find the Range of function f defined by

$f(x) = -2x^2 + 4x - 7$

We first write the given quadratic function in vertex form by completing the square

$f(x) = -2x^2 + 4x - 7 = -2(x 2 - 2x) - 7 = -2((x - 1)^2 - 1) - 7 = -2(x - 1)^2 - 5$

The domain of the given function is R with x taking any value in the interval $(-\infty, +\infty)$ hence $(x - 1)^2$ is either zero or positive. We start by writing the inequality

$(x - 1) 2 \geq 0$

Multiply both sides of the inequality by - 2 and change the symbol of inequality to obtain

$- 2 (x - 1) 2 \leq 0$

Add - 5 to both sides of the inequality to obtain- $2 (x - 1)^2 - 5 \leq - 5$ or $f(x) \leq - 5$ and hence the range of function f is given by the interval $(-\infty - 5]$.

Example 7:Find the Range of function f defined by

$f(x) = (x - 1) / (x + 2)$

Solution: For this rational function, a direct algebraic method similar to those above is not obvious. Let us first find its inverse, the domain of its inverse which give the range of f.

We first prove that f is a one to one function and then find its inverse. For a function to be a one to one, we need to show that

If f(a) = f(b) then a = b.

(a - 1) / (a + 2) = (b - 1) / (b + 2)

Cross multiply, expand and simplify

(a - 1)*(b + 2) = (b - 1)*(a + 2)

a b + 2 a - b - 2 = a b + 2b - a - 2

3a = 3b, which finally gives a = b and proves that f is a one to one function.

Let us find the inverse of f

y = (x - 1) / (x + 2)

Solve for x

x = (2y + 1) / (1 - y)

change y into x and x into y and write the inverse function

f^{-1} (x) = y = (2x + 1) / (1 - x)

The range of f is given by the domain of f -1 and is therefore given by the interval

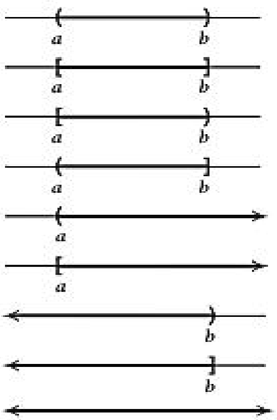
(-∞, 1) U (1, + ∞).

Notes:

Intervals: Notation and Graphs

Interval Notation	Set Notation
(a, b)	$\{x \mid a < x < b\}$
$[a, b]$	$\{x \mid a \le x \le b\}$
$[a, b)$	$\{x \mid a \le x < b\}$
$(a, b]$	$\{x \mid a < x \le b\}$
(a, ∞)	$\{x \mid x > a\}$
$[a, \infty)$	$\{x \mid x \ge a\}$
$(-\infty, b)$	$\{x \mid x < b\}$
$(-\infty, b]$	$\{x \mid x \le b\}$
$(-\infty, \infty)$	$\{x \mid x$ is a real number$\}$

Graph

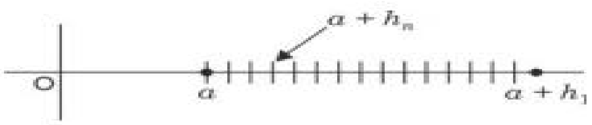


3.3 Limits Basics, Inclination and Slope

Assume y-f(x) is a function and h_1, h_2,………….h_n,………….. is a set of positive numbers, which value is continually decreasing viz

$h_1 > h_2 > h_3 >$…………………..h_n……………>0 (1)

And which, choosing n sufficiently greater, can be made smaller as desired. In this state, as the h_n goes down, the value of function decreases

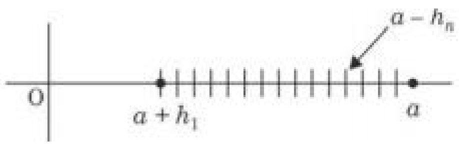


$f(a+h_1)$, $f(a+h_2)$, ………………, $f(a+h_n)$ (2)

if a number tends to A then this number is call right hand limit of function f(x) at x=a or this number A is called the right-hand limit of function f(x), when x tends to a. This can be expressed as:

$$\lim_{x \to a+0} f(x) = A = f(a+0)$$

Here we have considered only those values of x, which is greater than a. (in the figure onlya at the right side)



Now we will consider those values of x, which is smaller (viz in the figure only at the left side of a). As, hn goes down, the value of function f(a-h1), f(a-h2)……………., f(a-hn)………………tends to B. This number B is called the lefthand limit of function f(x), when x=a.

This is expressed as

$$\lim_{x \to a-0} f(x) = b = f(a-0)$$

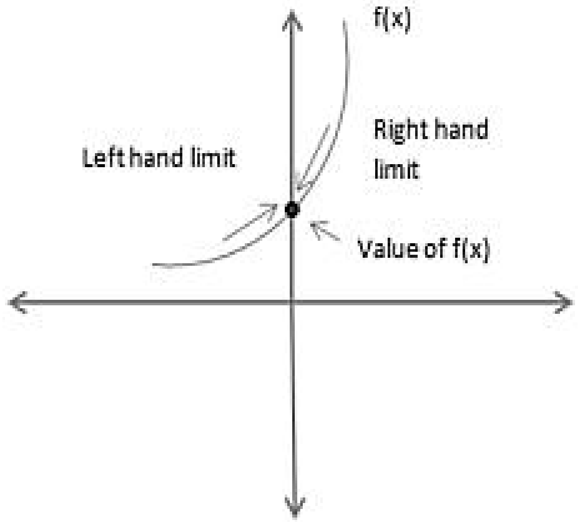
Ifa=b viz

$$\lim_{x \to a+0} f(x) = \lim_{x \to a-0} f(x)$$

Then A is called limit of f(x) at x=A

Set h_1,.h_2,…………….., h_n………… is a sequence for which limit is 0. Similarly, second makes sequence (2).Here, is to be specially noted that for limit to exist, like sequence (1) f(a+h_n) every type of sequence should tend to A. viz the statistical difference of f(a-hn)-A, choosing hn sufficient smaller, can be reduced as desired. Assigning a+h_n=x, we can define the limit as under

Limits of functions at a point are the common and coincidence value of the left and right-hand limits.



The value of a limit of a function f(x) at a point a i.e., f(a) may vary from the value of f(x) at ‘a’.

Left and right limits. When we let “x approacha” we allow x to be both larger or smaller than a, as long as x gets close to a. If we explicitly want to study the behaviour of f(x) as x approaches a through values larger than a, then we write

$\lim_{x \to a+}$ f(x) or $\lim_{x \to a+}$ f(x) or $\lim_{x \to a+0}$ f(x) or $\lim_{x \to a, x>a}$ f(x).

All four notations are in use. Similarly, to designate the value which f(x) approaches as x approaches a through values below a one writes

$\lim_{x\to a+} f(x)$ or $\lim_{x\to a^-} f(x)$ or $\lim_{x\to a^-0} f(x)$ or $\lim_{x\to a,\ x<a} f(x)$.

Definition of right-limits: Let f be a function. Then

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

means that for every ε > 0 one can find a δ > 0 such that

$a < x < a + \delta \Longrightarrow |f(x) - L| < \varepsilon$

holds for all x in the domain of f.

The left-limit, i.e. the one-sided limit in which x approaches a through values less than a is defined in a similar way. The following theorem tells you how to use one-sided limits to decide if a function f(x) has a limit at x = a.

Working Rules for Finding Right Hand Limit and Left-Hand Limit

i. To obtain the limit of right andleft hand, replace x variable with (x+h) and (x-h) respectively in the function.
ii. Thus, obtained function x, should be replaced with point (assume a)
iii. Now at h→0 determine the limit of function.

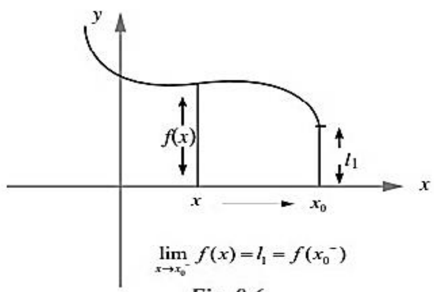
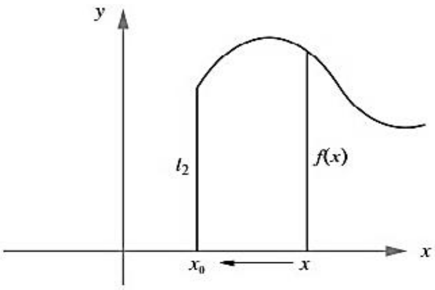
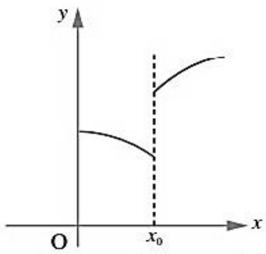


Fig. 9.6



$\lim_{x\to x_0^+} f(x) = l_2 = f(x_0^+)$

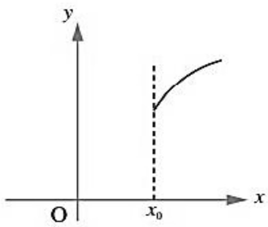
Fig. 9.7



$\lim_{x\to x_0} f(x)$ does not exist

Fig. 9.8

(Different values are obtained as x_0 is approached from the left and from the right)



$\lim_{x\to x_0} f(x)$ does not exist

Fig. 9.9

(Function not defined to the left of x_0)

Example 8: f (x) = {x^3-4 for x<2, 2x for x≥2}

a) What is f (x)?

b) What is f (x)?

c) What is f (x)?

Solution:a) $_{x\to 2-}$f (x) =$_{x\to 2-}$ (x^3 -4) = 23 - 4 = 4.

b) $_{x\to 2+}$f (x) = $_{x\to 2+}$(2x) = 2(2) = 4.

c) Because $\lim_{x\to 2-}$f (x) = $\lim_{x\to 2+}$ f (x) = 4, f (x) exists and is equal to 4.

Example 9:f (x) =

[x^5-12 for x<2 & $(x+1)^3$-8 for x≥2}

a) What is $_{x\to 2-}$ f (x)?

b) What is$_{x\to 2-}$ f (x)?

c) What is$_{x\to 2-}$ f (x)?

a) lim $_{x\to 2-}$f (x) = $_{x\to 2}$+(x^5 -12) = 25 - 12 = 20.

b) lim $_{x\to 2-}$f (x) = $_{x\to 2}$+[$(x + 1)^3$ -8] = $(2 + 1)^3$ - 8 = 19.

c) Because $_{x\to 2}$f (x)≠f$_{x\to 2+}$ (x),$_{x\to 2}$ f (x) does not exist.

Example 10: Find the limit

lim $_{x\to 2^-}$ (x – 1)/ (x^2 − 3x + 2).

Solution: We first evaluate the fraction (x − 1)/(x^2 − 3x + 2) at x = 2. We get 1/0.

This tells us that the limit lim x→2− (x – 1)/ (x2 − 3x + 2) is either +∞, −∞, or it does not exist. We are approaching 2 from the left, which means that x < 2, but really close to 2. In this case x 2 − 3x + 2 is a negative number, since x 2 − 3x + 2 is a parabola that opens up and has x-intercepts at the points 1 and 2. Thus the bottom of the fraction is negative, and the top is close to 1, hence positive. Together the fraction is negative, which gives

lim $_{x\to 2^-}$ (x – 1)/ (x^2 − 3x + 2)= −∞

Example 11: Find the limit: lim$_{x\to -3}$(x + 2) /(x + 3)

Solution: Arguing as we did above we can show that lim $_{x\to -3+}$ (x + 2) /(x + 3)= −∞ and lim $_{x\to -3-}$ (x + 2) /(x + 3) = +∞. Since the limit from the right is different than the limit from the left the full limit lim$_{x\to -3}$ (x + 2) /(x + 3) does not exist.

3.4 Logarithmic and Exponential Function

A quantity that increases (or decreases) by a fixed factor per unit of time is said to increase (or decrease) exponentially. If the fixed factor is a, this leads to the exponential function f (t) = Aat (a and A are positive constants) (1)

(It is obvious how to modify the subsequent discussion for the case when A is negative.) Note that if f (t) = Aat , then f (t + 1) = Aat+1 = Aat · a1 = af (t), so the value of f at time t + 1 is a times the value of f at time t. If a > 1, then f is increasing; if 0 <a< 1, then f is decreasing. (See Figs. 1 and 2.) Because f (0) = Aa0 = A, we can write f (t) = f (0)at .

Exponential functions appear in many important economic, social, and physical models. For instance, economic growth, population growth, continuously accumulated interest, radioactive decay, and decreasing illiteracy have all been described by exponential functions. In addition, the exponential function is one of the most important functions in statistics.

A population Q(t) is said to grow exponentially if whenever it is measured at equally spaced time intervals, the population at the end of any particular interval is a fixed multiple (greater than 1) of the population at the end of the previous interval. For instance, according to the United Nations, in the year 2000, the population of the world was 6.1 billion people and was growing at an annual rate of about 1.4%. If this pattern were to continue, then every year, the population would be 1.014 times the population of the previous year. Thus, if P(t) is the world population (in billions) t years after the base year 2000, the population would grow as follows:

2000 P(0) = 6.1

2001 P(1) = 6.1(1.014) = 6.185

2002 P(2) = 6.185(1.014) = [6.1(1.014)](1.014) = $6.1(1.014)^2$= 6.272

2003 P(3) = 6.272(1.014) = [6.1(1.014)2](1.014) = 6.1(1.014)3 = 6.360

2000 + t P(t) = 6.1(1.014)t

The graph of P(t) is shown in Figure 4.1a. Notice that according to this model, the world population grows gradually at first but doubles after about 50 years (to 12.22 billion in 2050).

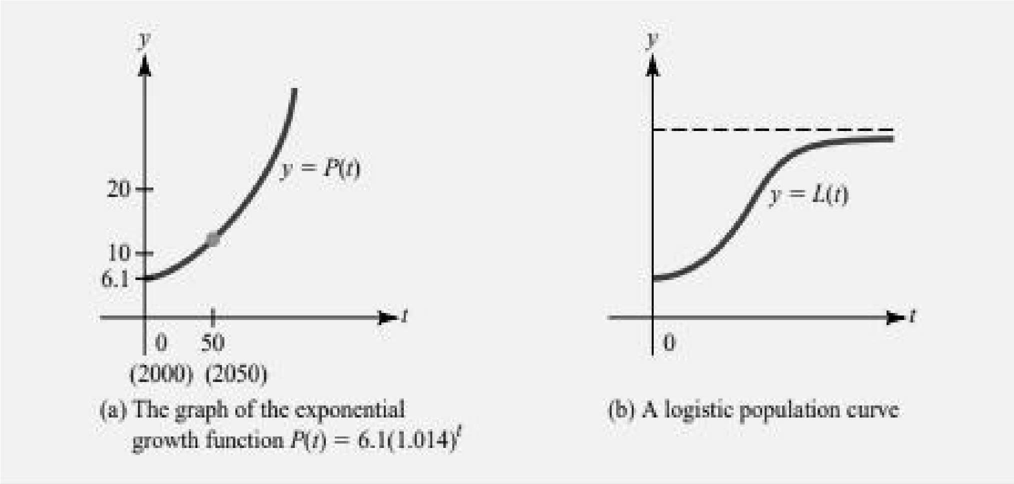


Fig. 1 Two models for population growth.

Exponential population models are sometimes referred to as Malthusian, after Thomas Malthus (1766–1834), a British economist who predicted mass starvation would result if a population grows exponentially while the food supply grows at a constant rate (linearly). Fortunately, world population does not continue to grow exponentially as predicted by Malthus's model, and models that take into account various restrictions on the growth rate actually provide more accurate predictions. The population curve that results from one such model, the so-called logistic model, is shown in Figure 4.1b. Note how the logistic growth curve rises steeply at first, like an exponential curve, but then eventually turns over and flattens out as environmental factors act to break the growth rate.

A function of the general form f(x)= bx, where b is a positive number, is called an exponential function. Such functions can be used to describe exponential and logistic growth and a variety of other important quantities. For instance, exponential functions are used in demography to forecast population size, in finance to calculate the value of investments, in archaeology to date ancient artifacts, in psychology to study learning patterns, and in industry to estimate the reliability of products.

Definition of b^n for Rational Values of n (and $b > 0$) ■ Integer powers: If n is a positive integer,

$$b^n = \underbrace{b \cdot b \cdots b}_{n \text{ factors}}$$

Fractional powers: If n and m are positive integers,

$$b^{n/m} = (\sqrt[m]{b})^n = \sqrt[m]{b^n}$$

where $\sqrt[m]{b}$ denotes the positive mth root.

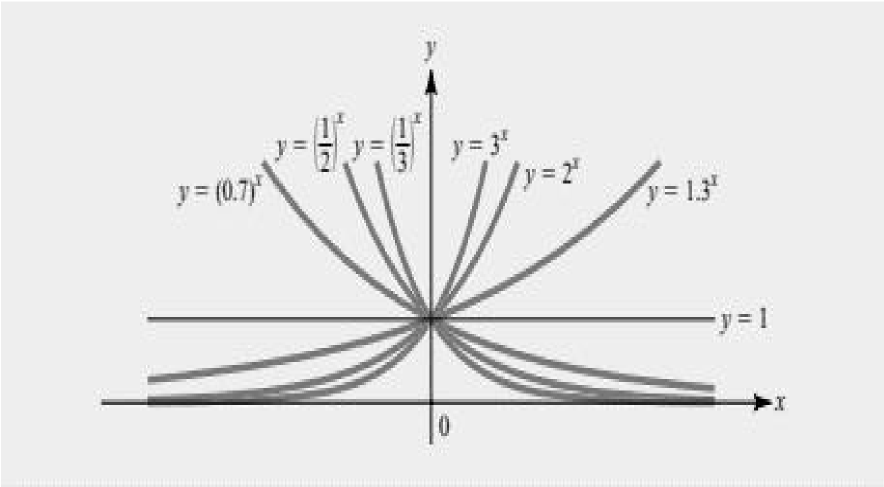
Negative powers: $b^{-n} = \dfrac{1}{b^n}$

Zero power: $b^0 = 1$

Exponential Functions

If b is a positive number other than 1 ($b>0$, $b \neq 1$), there is a unique function called the exponential function with base b that is defined by

$f(x) = b^x$ for every real number x



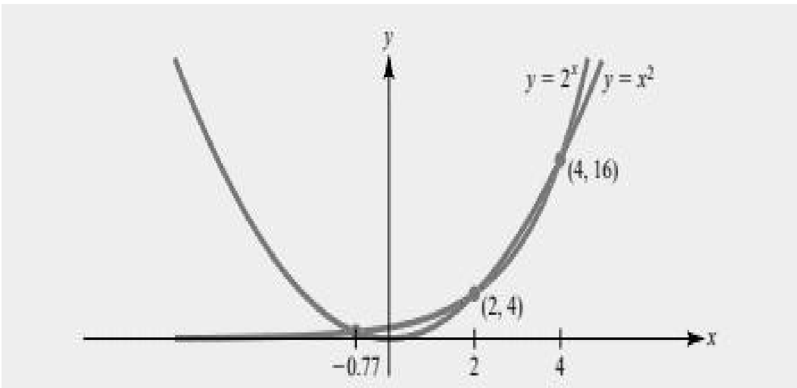
Important graphical and analytical properties of exponential functions are summarized in the following box

Properties of an Exponential Function ■ The exponential function $f(x) = b^x$ for $b > 0$, $b \neq 1$ has these properties:

1. It is defined, continuous, and positive ($b^x > 0$) for all x
2. The x axis is a horizontal asymptote of the graph of f.
3. The y intercept of the graph is (0, 1); there is no x intercept.
4. If $b > 1$, $\lim_{x \to -\infty} b^x = 0$ and $\lim_{x \to +\infty} b^x = +\infty$.

   If $0 < b < 1$, $\lim_{x \to -\infty} b^x = +\infty$ and $\lim_{x \to +\infty} b^x = 0$.
5. For all x, the function is increasing (graph rising) if $b > 1$ and decreasing (graph falling) if $0 < b < 1$.

Notes:Students often confuse the power function $p(x)= x^b$ with the exponential function $f(x)=b^x$. Remember that in x^b, the variable x is the base and the exponent b is constant, while in b^x, the base b is constant and the variable x is the exponent. The graphs of $y= x^2$ and $y = 2^x$ are shown in Figure. Notice that after the crossover point (4, 16), the exponential curve $y= 2^x$ rises much more steeply than the power curve $y = x^2$. For instance, when x = 10, the y value on the power curve is $y = 10^2 = 100$, while the corresponding y value on the exponential curve is $y = 2^{10} = 1{,}024$.



Comparing the power curve $y = x^2$ with the exponential curve $y =2^x$.

Exponential functions obey the same algebraic rules as the rules for exponential These rules are summarized as follows:

Exponential Rules

- For bases a, b (a > 0, b> 0) and any real numbers x, y, we have

1. The equality rule: $b^x = b^y$ if and only if x = y
2. The product rule: $b^x b^y = b^{x+y}$
3. 3.The quotient rule: $b^x/b^y = b^{x-y}$
4. The power rule: $(bx)^y = b^{xy}$
5. The multiplication rule: $(ab)^x = a^x b^x$
6. The division rule: $(a/b)^x = a^x/b^x$

Example 12: Let $f(x)=5(3)^{x+1}$. Evaluate f(2) without using a calculator.

Solution:

Follow the order of operations. Be sure to pay attention to the parentheses.

$f(x)=5(3)^{x+1}$

$f(2)=5(3)^{2+1}$ Substitute x=2.

$=5(3)^3$ Add the exponents.

$=5(27)$ Simplify the power.

$=135$ Multiply.

Example 13: Find dy/dx:

a) $Y=3e^x$

$d/dx(3e^x) = 3\ d/dx*e^x$

$=3e^x$

b) $Y=x^2e^x= x^2 \cdot e^x + e^{x \cdot 2x}$

$= e^x(x^2 +2x)$

c) $Y=g(x)=(1/2)^x$

$= (2^{-1})^x$

$= 2^{-x}$

Logarithmic Function

Suppose you invest 1,000 at 8% compounded continuously and wish to know how much time must pass for your investment to double in value to 2,000. According to the formula derived, the value of your account after t years will be $1{,}000e^{0.08t}$, so to find the doubling time for your account, you must solve for t in the equation

$1{,}000e^{0.08t} = 2{,}000$

or, by dividing both sides by 1,000,

$e^{0.08t} =2$

We will answer the question about doubling time. Solving an exponential equation such as this involves using logarithms, which reverse the process of exponentiation. Logarithms play an important role in a variety of applications, such as measuring the capacity of a transmission channel and in the famous Richter scale for measuring earthquake intensity. In this section, we examine the basic properties of logarithmic functions and a few applications. We begin with a definition

A logarithm is defined as follows:

$Log_a x=y$ means $a^y=x,\ a>0,\ a\neq 1$.

The number is the power y to which we raise a to get x. The number a is called the logarithmic base. We read as "the logarithm, base a, of x.

For logarithms with base 10, is the power y such that Therefore, a logarithm can be thought of as an exponent.

Logarithmic Equation	Exponential Equation
$\log_a M = N$	$a^N = M$
$\log_{10} 100 = 2$ $\log_5 \frac{1}{25} = -2$ $\log_{49} 7 = \frac{1}{2}$	$10^2 = 100$ $5^{-2} = \frac{1}{25}$ $49^{1/2} = 7$

In order to graph a logarithmic equation, we can graph its equivalent exponential equation.

Example 14: Graph: $y=\log_2 x$.

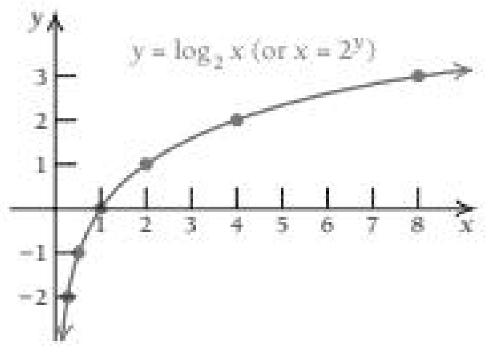
Solution: We first write the equivalent exponential equation:

$2^y=x$

We select values for y and find the corresponding values of 2y. Then we plot points, remembering that x is still the first coordinate, and connect the points with a smooth curve.

x, or 2^y	y
1	0
2	1
4	2
8	3
$\frac{1}{2}$	−1
$\frac{1}{4}$	−2

(1) Select y.
(2) Compute x.



The graphs of and are shown below on the same set of axes. Note that we can obtain the graph of g by reflecting the graph off across the line y=x. Functions whose graphs can be obtained in this manner are known as inverses of each other.

Although we do not develop inverses in detail here, it is important to note that they "undo" each other. For example,

$f(3)=2^3=8$,	The input 3 gives the output 8.
and $g(8)=\text{Log}_2 8=3$	The input 8 gets us back to 3

Homogenous Function, Cobb Douglas Production Function, Cost Functions and Production Functions

Homogeneous production functions consist of a broad array of functions with a special characteristic. A production function is said to be homogeneous of degree n if when each input is multiplied by some number t, output increases by the factor tn . Assuming that the time period is sufficiently long such that all inputs can be treated as variables and are included in the production function, n, the degree of homogeneity refers to the returns to scale. Homogeneous production functions are frequently used by agricultural economists to represent a variety of transformations between agricultural inputs and products. A function homogeneous of degree 1 is said to have constant returns to scale, or neither economies or diseconomies of scale. A function homogeneous of a degree greater than 1 is said to have increasing returns to scale or economies of scale. A function homogeneous of degree less than 1 is said to have diminishing returns to scale or diseconomies of scale. While there are many different production functions, only certain kinds of production functions are homogeneous. In general, they are multiplicative rather than additive although a few exceptions exist. The production function

y = Ax1 0.5x2 0.5 is homogeneous of degree 1. Multiply x1 and x2 by t to get

$A(tx_1)^{0.5}(tx_2)^{0.5} = tAx_1^{0.5}x_2^{0.5} = t^1y$

Thus, the function in equation exhibits constant returns to scale without any economies or diseconomies.

Euler's Theorem

Euler's Theorem states that all factors of production are increased in a given proportion resulting output will also increase in the same proportion each factor of production (input) is paid the value of its marginal product, and the total output is just exhausted. If every means of production is credited equal to its marginal productivity and total production is liquidated completely. In mathematical formula Euler's Theorem can be indicated. If production, P = f (L, K) is Linear Homogeneous Function:

Euler's theorem states that if $f(x, y)$ is an homogeneous function of degree r, then

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = rf(x, y).$$

This follows from a simple application of the chain rule since, using the definition of a function that is homogeneous of degree r, we have

$$f(\lambda x, \lambda y) = \lambda^r f(x, y),$$

for any $\lambda \in \mathbb{R}$. As such, differentiating both sides with respect to λ and using the chain rule from (6.3) on the left-hand side, we have

$$\frac{\partial f}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}\lambda} + \frac{\partial f}{\partial v}\frac{\mathrm{d}v}{\mathrm{d}\lambda} = r\lambda^{r-1}f(x, y),$$

if we think of $f(\lambda x, \lambda y)$ as $f(u, v)$ with $u = \lambda x$ and $v = \lambda y$. This then gives us

$$x\frac{\partial f}{\partial u} + y\frac{\partial f}{\partial v} = r\lambda^{r-1}f(x, y).$$

and, if we now set $\lambda = 1$, we get the desired result as we have $u = x$, $v = y$ and $\lambda^{r-1} = 1$.

Cobb-Douglas Production Function

The Cobb-Douglas (CD) production function is an economic production function with two or more variables (inputs) that describes the output of a firm. Typical inputs include labor (L) and capital (K). It is similarly used to describe utility maximization through the following function [U(x)]. However, in this example, we will learn how to answer a minimization problem subject to (s.t.) the CD production function as a constraint.

The functional form of the CD production function:

$$f(L, K) = Y = AL^{\alpha}K^{\beta},$$

where the output Y is a function of labor (L) and capital (K), A is the total factor productivity and is otherwise a constant, L denotes labor, K denotes capital, alpha represents the output elasticity of labor, beta represents the output elasticity of capital, and (alpha + beta = 1) represents the constant returns to scale (CRS). The partial derivative of the CD function with respect to (w.r.t) labor (L) is:

$$\frac{\partial Y}{\partial L} = A\alpha L^{\alpha-1}K^{\beta}$$

$$\frac{\partial Y}{\partial L} = \alpha A L^{\alpha} L^{-1} K^{\beta}$$

$$\frac{\partial Y}{\partial L} = \frac{\alpha A L^{\alpha} K^{\beta}}{L}$$

Recall that quantity produced is based on the labor and capital; therefore, we can solve for alpha:

$$Y = AL^{\alpha}K^{\beta},$$

$$\frac{\partial Y}{\partial L} = \alpha\frac{Y}{L}$$

$$\alpha = \frac{\partial Y}{\partial L}\frac{L}{Y}$$

This will yield the marginal product of labor (L). If alpha = 2, then a 10% increase in labor (L) will result in a 20% increase in output (Y).

The partial derivative of the CD function with respect to (w.r.t) labor (K) is:

$$\beta = \frac{\partial Y}{\partial K}\frac{K}{Y}$$

This will yield the marginal product of capital (K).

The CD production function can be converted to a linear model by taking the logarithm of both sides of the equation:

$$f(L, K) = Y = AL^{\alpha}K^{\beta}$$

$$\log(Y) = \log(A) + \alpha\log(L) + \beta\log(K)$$

We've shown that the Cobb–Douglas function gives diminishing returns to both labor and capital when each factor is varied in isolation. But what happens if we change both K and N in the same proportion?

Suppose an economy in an initial state has inputs K0 and N0 and produces output

$Y_0 = AK^{\alpha}_0 N^{1-\alpha}_0$

Now suppose we scale the inputs by some common factor λ. (For example, λ = 2 would mean that we double each input.) We'll then have inputs K1 = λK0 and N1 = λN0 and will produce output Y1. The question is, how does Y1 relate to Y0? Let's see:

$Y1 = AK^{\alpha}{}_1 N^{1-\alpha}{}_1$

$= A(\lambda K_0)^{\alpha} (\lambda N_0)^{1-\alpha}$

$= A\lambda^{\alpha} K_0^{\alpha}\lambda^{1-\alpha} N_0^{1-\alpha}$

$= \lambda^{\alpha+1-\alpha}AK_0^{\alpha} N_0^{1-\alpha}$

$= \lambda Y_0$

So, if we scale both inputs by a common factor, the effect is to scale the output by that same factor. This is the defining characteristic of constant returns to scale. From the math above we can see that this occurs in the Cobb–Douglas function because the exponents on capital and labor, α and 1 − α, add up to 1.

We could imagine a generalization of Cobb–Douglas in which the exponents on capital and labor are (say) α and β respectively, preserving the requirement that each exponent be a positive fraction (this is needed to give positive but diminishing marginal products) but dropping the requirement that they sum to 1. In that case we'd get increasing returns to scale if α + β > 1 and decreasing returns to scale if α + β < 1.

Economic significance of Cobb Douglas Production Function

Cobb Douglas Production Function has a very importance role in economic area. At present many economists are using Cobb Douglas Production Function in various economic areas. The use of this function is day-by-day is increasing especially in various industries and agriculture. This bring important information for these sectors. This also helps in framing various policies. With the help of this function, we can also determine the Marginal Productivity and similarly it helps in determining principle of wages. Production function describes production technique. With the help of this function we can also determine whether any factor is paid the value with respect to its equality with the marginal productivity. In a same fashion it helps in agriculture to find the elasticity of economy. By this function we also display elasticity coefficients. These elasticity coefficients help us in comparing the international and internal areas.

As has already be described when function is linear and homogeneous and $a + b = 1$, then production would be under the constant result, when $a + b > 1$, then increase in production happens, and if $a + b < 1$, then decrease in production happens. This way this function helps us in studying the rules of various results. Besides these it also fetches important information related to substitutability of various factors of production.

In short, this function plays an important role especially in agriculture and industries. This is used in determining the labour policies, inter-area comparison, substitutability of factors and degree of homogeneity.

Limitation of Cobb-Douglas Production Functions:

Although Cobb Douglas Production Function is used widely in economic areas and its use is increasing in especially in various industries and agriculture, but some economists criticize this production function. Among them are Prof K.J. Arrow, H.B. Chenery, B.S. Minhas and R.M. Salow. Their main criticizes are:

1. The main demerit of this function is this that it considers only two factors of production i.e. Capital and Labour, whereas in reality other factors also have important role in production. In other words, this function does not apply to more than two factors. Besides it can be used only in construction industries. This way its use becomes narrow.

2. This function works under the constant result of formula. Rule of increase and decrease in result also apply to production function. But this function does not work under these rules.

3. Function is based on the assumptions that technical knowledge remains constant and no change in techniques happen in production. But the same can change in production. This way assumption of constant technique is irrelevant.

4. Cobb-Douglas Production Functions assume that all inputs are homogeneous. In reality all units of a factors are not homogeneous. For example, some people are skilled and others are not in a labour population.

5. This does not determine any maximum level of production. Prof M. Chand says "Since, this does not ascertain the maximum level of P (Production), it would be practical and convenient not to use this function beyond a certain limit for statistical measurement of its values.

6. a and b of the function reflects the proportion of labour and capital in production. This becomes true only when market has a complete competition. But in case economy has an incomplete competition or monopoly, then above relation cannot be obtained.

7. It takes into account only positive marginal productivity of factors and ignores the negative marginal productivity. Whereas marginal productivity of any factor can be zero or negative.

8. Last, the function is unable to produce information related to inter-relation of factors.

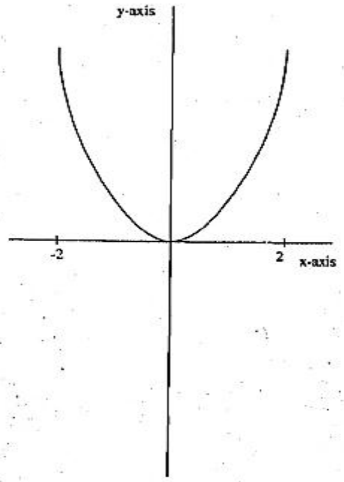
3.5 Continuous Functions, their Optima and their Existence

Suppose that you have functions which are defined on an interval, either opeN; or closed. If you draw the graph of these functions, you will observe that some of these can be sketched down in one smooth 'continuous' sweep of your pen, while others have many breaks or jumps. For example, draw the graph of the following two functions:

A. $f(x) = x^2, x \in [-2, 2]$;

B. . $g(x) = \begin{cases} \frac{1}{x}, x\varepsilon[-2,2] \\ 0, x = 0 \end{cases}$

These are as shown in figures l(a) and l(b).



Pig. 1(a)

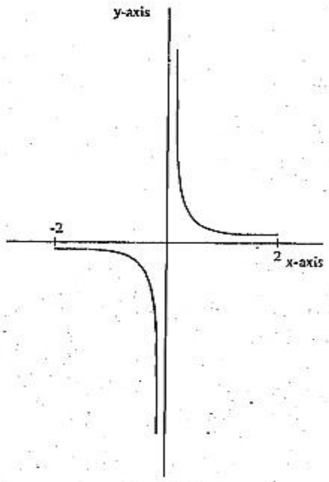


Fig. 1(b)

You can see that while the graph of the first function can be drawn in the 'continuous' motion without lifting the pen from the paper while the graph of the other function cannot lifting &awn in this manner. This is an interesting property of the first function which is nut possessed by the second function. It is, therefore, natural to wonder if it can be given some mathematical meaning. In fact, mathematicians of the past several centuries did confront this question, namely:

"1s there a way to specify those curves which can be drawn with a single stroke of one's pen?"

Continuous Functions

We have seen that the limit of a function f as the variable x approaches a given point a in the domain of'a function f does not depend at all on the value of the function at that point a but it depends only on the values of the function at the points near a. In fact, even if the function f is not defined at a then $\lim_{x\to a} f(x)$ may exist.

For example, $\lim_{x\to a} f(x)$ exists when

$f(x) = \dfrac{x^2-1}{x-1}$ though f is not defined at x = 1.

We have also seen that $\lim_{x\to a} f(x)$ may exist, still it need not be the same as f(a) when it exists. Naturally, we would like to examine the special case when both $\lim_{x\to a} f(x)$ and f(a) exist and are equal. If a function has these properties, then it is called a continuous function at the point a. We give the precise definition as follows:

DEFINITION 1: Continuity of a Function at a Point

A function f defined on a subset S of the set R is said to be continuous at a point

a S, if

i) lim f(x) exists and is finite
ii) $\lim_{x\to a} f(x) = f(a)$

Note that in this definition, we assume that S. contains some open interval containing the point a. If we assume that there exists a half open (semi-open) interval [a, c] contained in S for some c E R, then in the above definition, we can replace $\lim_{x\to a} f(x)$ by $\lim_{x\to a+} f(x)$ and say that the function is continuous from the right of a or f is right continuous at a.

iii) Similarly, you can define left continuity at a, replacing the role of by $\lim_{x\to a} f(x)$. Thus, f is continuous from the right at a if and only if $\lim_{x\to a} f(x)$ by $\lim_{x\to a-} f(x)$. Thus, f is continuous from the right at a if and only if
iv) F(a+)=f(a)

It is continuous from the left at a if and only if f(a-) = f(a)

From the definition of continuity of a function f d a point a and properties of limits it follows that f(a+) = f(a-) = f(a) if and only if, f is continuous at a. If a function is both continuous from the right and continuous from the left at a point a, then it is continuous at a and conversely.

The definition 1 is popularly known as the Limit-Definition of Continuity.

Since $\lim_{x \to a} f(x)$ is also defined in terms of E and 6, we have an equivalent x-a formulation of the definition 1 in terms of ε and 6. Note that whenever we talk of continuity of a function f at a in S, we always assume that S contains a neighborhood containing a. Also remember that if there is orie such neighborhood there are infinitely many such neighborhoods. An equivalent definition of continuity in terms of ε and 6 is given as follows:

Definition 2: Definition of Continuity

A fuction f is continuous nt x = a if f is defined in a neighborhood of a and corresponding of a given number E > 0, there exists some number $\delta > 0$ such that (x – a) <δ implies (f(x) -- f(a)<E.

Example 1: Examine the continuity of the function f(x)=x^3 +3x-4 at x=1

Sol.1: Condition 1: For x=1, f(1)= $(1)^3$ +3(1)-4=1+3-4=0

Therefore, the function is defined for x =1.

Condition 2: $\underset{x \to 1}{Lt}(x^3 + 3x - 4) = (1)^3 + 3(1) - 4 = 0$. That is, the limit exists.

Condition 3: Thus, we find that f(1)= : $\underset{x \to 1}{Lt}\ f(x) = 0$

Since all the conditions are satisfied the function is continuous at x =1.

Example 2: Determine at which values of x , the functions given below are continuous.

a. $f(x) = \frac{x^4+3x^2-1}{(x-1)(x+2)}$

Solution: $f(x) = \frac{x^4+3x^2-1}{(x-1)(x+2)}$. This is a rational function and hence is continuous at all points except when (x-1)(x-

2) i.e. denominator vanishes (is zero). This gives us two cases:

(i) When x– 1 = 0, x = 1 ,

$\underset{x \to 1}{Lt} f(x) = \infty$i.e. undefined.

(ii) When x +2 = 0, x = −2,

$\underset{x \to -2}{Lt} f(x) = \infty$i.e. undefined.

Other conditions will not be needed. Hence, the function is discontinuous being undefined at x =1and x = −2.

Continuity of a Function over an Interval

A function y = f (x) is said to be continuous in an interval (a, b), if it is continuous at every value of x in that interval, i.e., it is continuous

1) at x a = 2) at x b = 3) and at any point between a b and.

Example 3: Show that the function 1/ x − 2 is continuous for values of x from x = −2 to x = −1i.e., in the interval [−2, −1].

Solution: In order to prove that the function is continuous over the range [−2, −1], we will prove that:

1. It is continuous at x = −2

2. It is continuous at x = −1
3. It is continuous for any value between - − 2 and 1, say, 1.5 or 3/2.

Case 1: Let us first find out $\lim_{x \to -2} f(x)$

Assessing Right hand limit

$\lim_{x \to -2+} f(x) = \lim_{h \to 0} f(-2+h) = \lim_{h \to 0} 1/(-2+h-2) = -1/4$

Assessing Left hand limit

$\lim_{x \to -2-} f(x) = \lim_{h \to 0} f(-2-h) = \lim_{h \to 0} 1/(-2-h-2) = -1/4$

Since, $\lim_{x \to -2+} f(x) = \lim_{x \to -2-} f(x) = -\frac{1}{4}$.

so the function f (x) is continuous at x = −2

Case 2: Similarly, we can prove that the function is continuous at x = −1. We leave this exercise for the students to try themselves

Case 3: Let us first find out $\lim_{x \to -3/2} f(x)$

Assessing Right hand limit

$\lim_{x \to -\frac{3}{2}+} f(x) = \lim_{h \to 0} f(-3/2+h) = \lim_{h \to 0} 1/(-3/2+h-2) = -2/7$

Assessing Left hand limit

$\lim_{x \to -3/2-} f(x) = \lim_{h \to 0} f(-3/2-h) = \lim_{h \to 0} 1/(-3/2-h-2) = -2/7$

Since, $\lim_{x \to -3/2+} f(x) = \lim_{x \to -3/2-} f(x) = -2/7$.

Therefore, we find that the function is continuous at x = − 3/2 which is a point lying between -2 and -1.

It can easily be checked that the function is continuous at every point between −2 and−1.

Summary

- If function f: X→Y is defined, then non empty set X is called domain of this mapping or function (f) and non-element set (Y) would be its range or codomain.
- If variable u is dependent on variable x and y, then u=f(x,y) is called functions of many variable.
- Function with different exponential value of x variable, in which factors are certain is called algebraic function.
- Function expressed in the form of fraction, in which numerators and denominators are of algebraic function of exponential value, is called rational numbers.
- When the limit of function is obtained from the right hand of the independent variable, then it is called Right Hand Limit(RHL) and applying positive (+) sign for the right side, this can be expressed as under

  Right Hand Limit = f(a+0)

  $= \text{Lim}_{x \to a^+} f(X) = l_1$
- When the limit of function is obtained from the right hand of the independent variable, then it is called Left Hand Limit(lHL) and applying positive (-) sign for the right side, this can be expressed as under

- Right Hand Limit = f(a-0)
  $=\text{Lim}_{x\to a^-} f(X) = l_2$
- A function f(x) is called continuous in an open interval (a,b) if it is continuous for every values of x in this interval (a,b).
- If any such equation exists between x andy such that cannot be solved for y instantaneously then y is said to be the implicit function of x. In contrast if value of y can be found out in terms of x then y is said to be explicit function of x.

  Log(m*n) = logm +logn

  Log(m/n)= logm-logn

  $\text{Log}(m)^n$ =nlogm

- Maximum use of special production is referred to as Homogeneous function.
- Euler's Theorem states that all factors of production are increased in a given proportionresulting output will also increase in the same proportion each factor of production (input) ispaid the value of its marginal product, and the total output is just exhausted.
- Euler's Theorem has an important place in economic area especially in marketing area.
- Cobb Douglas Production Function has a very important role in economic area. At presentmany economists are using Cobb Douglas Production Function in various economic areas.
- Continuous Variable: If x takes all possible real values from a given number a to another given number b, then x is called a continuous variable.
- Continuity: A function f(x) is continuous provided its graph is continuous, i.e., a continuous function does not have any break at any point of its graph. More formally, a function f(x) is said to be continuous for x = a, provided $\lim_{x\to a} f(x)$ exists, finite and is equal to a.

Keywords

Domain: Affected area

Range: Series, limits of variation.

Sequence: Serial

Continually: Continuously

Infinite: which doesn't have an end

Homogenous: undefined, similar

Continuity: without any break or jerk

Self Assessment

1. A function is said to be _______________ if and only if f(a) = f(b) implies that a = b for all a and b in the domain of f.

A. One-to-many

B. One-to-one

C. Many-to-many

D. Many-to-one

2. If a<0, then the function f(x)=ax2+bx+c

A. Maximum value

B. Minimum value
C. Constant value
D. Positive value
3. A parabola graph which opens downward is classified as
A. Concave right
B. Concave left
C. Concave up
D. Concave down

4. What is the domain of a function?
A. the maximal set of numbers for which a function is defined
B. the maximal set of numbers which a function can take values
C. it is a set of natural numbers for which a function is defined
D. none of the mentioned

5. What is range of function f(x) = x-1 which is defined everywhere on its domain?
A. (-∞, ∞)
B. (-∞, ∞) – {0}
C. [0, ∞)
D. None of the mentioned

6. Codomain is the subset of range.
A. True
B. False
C. May be True or False
D. Can't say

7. The exponential function class f(x)=b^x where 0<b<1, the function is between x and y is classified as
A. x is increasing function of y
B. y is decreasing function of x
C. y is increasing function of x
D. x is decreasing function of y

8. Solve the equation by expressing each side as a power of the same base and then equating exponents: 4logby+6logbz
A. logb y4z6
B. 10logbyz
C. 24logbyz
D. logb(YZ)10

9. Solve the equation by expressing each side as a power of the same base and then equating exponents: e(x+3)= 1/e2
A. {-1}
B. {5}
C. {-5}
D. {1}

10. $Log_a mn$ equals to
A. $Log_a m + log_a n$
B. $Log_a m - log_a n$
C. $Nlog_a m$
D. $Log_b n * log_a b$

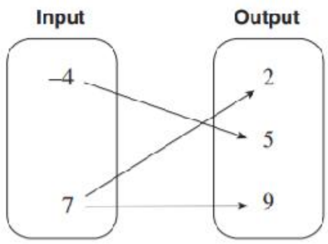
11. A significant property of the Cobb-Douglas production function is that the elasticity of substitution between inputs is
A. Equal to 1
B. More than 1
C. Less than 1
D. 0

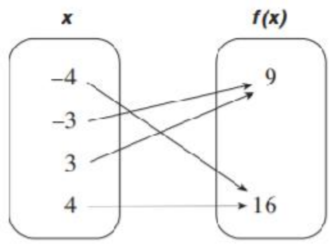
12. Cobb-Douglas production function $Q = AL^{\alpha} K^{1-\alpha}$ does not possess the characteristics of
A. Constant Returns to Scale
B. Unit Elasticity of Substitution
C. Variable Elasticity of Substitution
D. Linear homogeneity

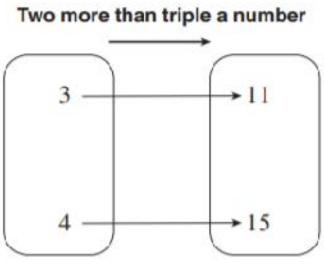
13. f (x) is a continuous function and takes only rational values. If f (0) = 3, then f (2) equals
A. 5
B. 0
C. 1
D. None of these

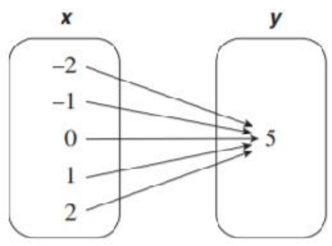
14. If $\lim_{x \to 2} \frac{(x-2)(x+2)}{(x-2)}$, is limit exist at x approaches to 2?
A. True
B. False

15. Which of the following relations is not a function?

A.


B.


C.


D.


A. A
B. B
C. C
D. D

Answers for Self Assessment

1.	B	2.	A	3.	D	4.	A	5.	A
6.	A	7.	B	8.	A	9.	C	10.	C
11.	A	12.	C	13.	D	14.	A	15.	A

Review Questions

1. If function f: R→R is defined in the following way
   F(x)= {3x-1, if x>3, x^2-2, if 2≤x≤3 & 2x+3, if x<-2}
   Then find out the value of f(2) & f(4).
2. Find out the domain and range of function f(x)=x/1+x^2. If the function is single?
3. Solve each of the following equations for x:

a. $Log_4 x=1/2$
b. $Log_{64} 16=x$

4. Evaluate the given expressions

A. $(-128)^{3/7}$
B. $(X^{2/3})^{-3/4}$
C. $(2.14)^{x-1}=(2.14)^{1-x}$
D. $10^{(x2-1)}=10^3$

5. A marketing manager estimates that t days after termination of an advertising campaign, the sales of a new product will be S(t) units, where
   $S(t) = 4000\ e^{-0.015t}$

A. How many units are being sold at the time advertising ends?
B. How many units will be sold 30 days after the advertising ends? After 60 days?

6. The population density x miles from the center of a city is given by a function of the form

$Q(x) = Ae^{-kx}$

Find this function if it is known that the population density at the center of the city is 15,000 people per square mile and the density 10 miles from the center is 9,000 people per square mile.

7. Define homogeneous function with example.
8. Explain Euler's Theorem with realistic example.

Further Readings

- Mathematics for Economics-Council for economic education
- Essential Mathematics for Economists- Nutt Sedester, peter Hawmond, Prentice Hall Publication
- Mathematics for Economists- Carl P Simone, Lawrence Bloom.
- Mathematics for Economist- Simone and Bloom, Viva Publication

Unit 04: Maxima and Minima

CONTENTS

Objectives

- define critical points, stationary points, saddle points, local maxima and local miniina;
- state a necessary condition for functions to have local extrema and apply it;
- state and prove the theorem known as "second derivative test" which gives a sufficient condition for finding local maxima and minima;
- use Hessian for classifying local maxima and local minima; and
- apply Lagrange's multiplier method for finding the stationary points when the variables are subject to some constraints.

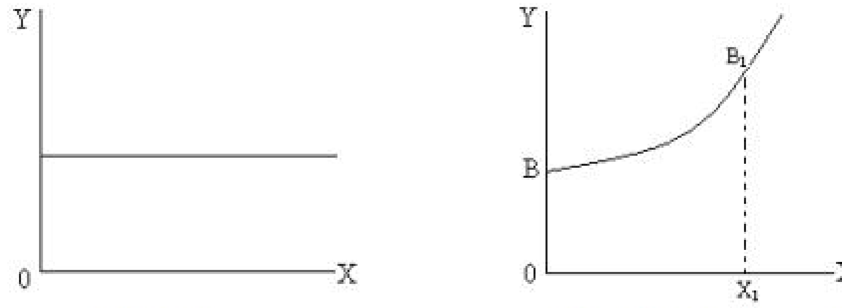
Introduction

The primary objective of this unit is to characterize Optimal points, which are also called ‘extremal points’ and lay down the conditions that need to be satisfied to classify a point as an ‘extremum’, either a maximum or a minimum. Optimum (plural: optima) is the generic term for maximum and minimum. The process of finding an optimum is called optimization. The basic idea is that a decision maker (for example a consumer or a firm) has an objective function that the decision-maker is attempting to optimize (that is, maximize or minimize).

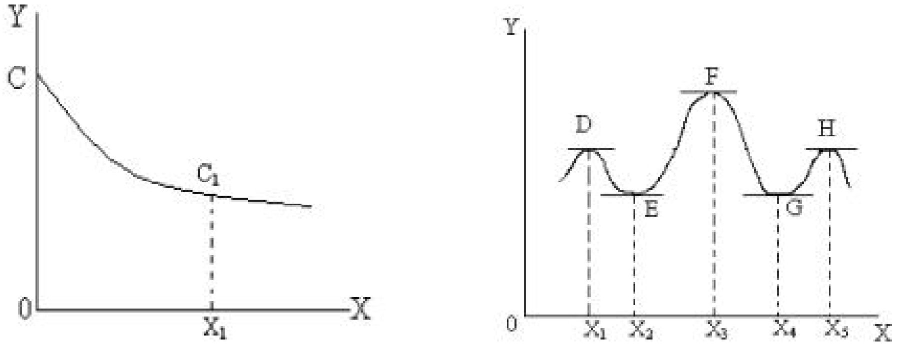
The question that needs to be answered at the outset is as follows: What is the importance of 'locating an extrema' in the context of Economics? The process is important because an economic unit (consumer, producer, etc.) is often faced with various different alternatives. For instance, a consumer has to choose from different commodity bundles, or a producer has to choose amongst various combinations of factor inputs, viz. Labor, capital, etc. The economic agent has to choose one particular alternative, which very often either maximize something (e.g. a producer will maximize profit, a consumer maximizes her utility) or minimize something (e.g. cost of producing a given output). Economically, this process of maximization or minimization is characterized as a 'process of optimization' or 'the quest for the best'. However, from the standpoint of a mathematician, the location of a 'maximum' or a 'minimum' does not carry forth any notion of optimality. To solve an optimizationproblem, the first task of the economic agent is to construct an 'objective function'. The dependent variable of this function is the so-called 'object', which has to be either maximized or minimized. The independent variable(s) of the function are the choice/decision/policy variables that can be manipulated by the agent to achieve the desired goal. The optimization process involves choosing a value of the independent variable that will yield an extreme value ('minimum' or 'maximum' as the case might be) for the requisite dependent variable of the objective function. The process of locating an extremum value (either a maximum or a minimum) is discussed in the following sections. Note, only classical technique for locating extreme positions, using differential calculus, will be discussed. For simplicity it will be assumed that the objective function consists of a single independent variable.

5.1 Maxima and Minima of Functions

To start with, we assume the function y = f(x) is differentiable (and hence continuous) throughout its range. The function may be represented by one of the following types of graphs:



i) Constant ii) Monotonically increasing



iii) Monotonically decreasing iv) Fluctuating

In the first case, the function is the same for all values of x. We shall consider the values of y as both maximum and minimum value.

In the second case, the value of the function is monotonically increasing. Let us restrict the domain of x to the interval 0 ≤ x ≤ x. Then in this closed interval, y can assume a maximum of B1x1 and a minimum of OB. These maximum and minimum values of the function are called absolute maximum (global) and absolute (global) minimum respectively. Such maximum and minimum values of a function are called extreme values or extrema. Accordingly, in the third case, OC becomes the global maxima and X1C1 the global minima if we restrict in the interval of x to $0 \leq x \leq x$.

Let us consider the fourth case. Here y oscillates with changes in x. The function generates three peaks, D, F and H and two bottoms E and G. The value of the function at D is the highest in comparison to the values at other points in its the immediate neighborhood. Symbolically,

$f(x - \in) < f(x_1)$ and $f(x_1 + \in) < f(x_1)$ as $\in \rightarrow 0$.

Thus, we can say that the value of the function at D is maxima, at least in some small interval $x_1 - \in < x < x + \in$. Similarly, in another small interval $x_3 - \in < x < x + \in$, the value of the function is maximum at $x = x_3$ and in the neighborhood of F,

$f(x_3 - \in) < f(x_3)$ and $f(x_3 + \in) < f(x_3)$

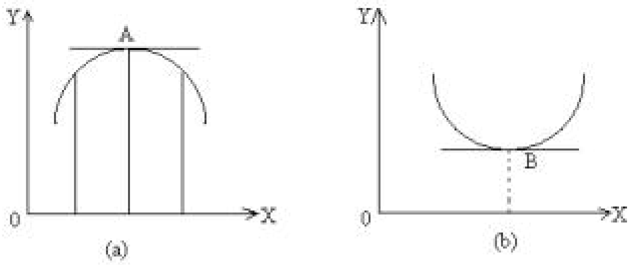
The function attains another maximum in the neighborhood of point F. These maximum values are called local (relative) maxima. A relative maximum gives us the information that the value of the function is maximum in the neighborhood of the maximum point only, and the function can assume higher values elsewhere. For example, in the fourth figure, points D and H describe relative maximum, though the function assumes a higher value at F. The relative maxima are unlike global maxima. The global maximum is the highest value of the function with reference to its entire domain and hence unique.

A function can have several local maxima and unique global maxima. The concept of local minima and global minima can be explained in the same manner. For a continuous function, the global maximum must always be greater than the global minimum.

The local extrema always appear as point interior to an interval, however small. These local extrema are thus strictly interior extrema. In the following discussion, we shall consider only these types of extrema.

5.2 Identification of Maxima and Minima

Consider the following figure:



The extremum described by (a) is a maximum, while that of (b) is a minimum. In Figure (a) the curve $y = f(x)$ takes a turn at point A. To the left of A, the curve is increasing i.e., $f'(x)>0$ while to the its right f(x) is decreasing i.e., $f'(x)<0$. Thus, at the turning point A, $f'(x)$ must be zero i.e., $f'(x0) =0$. Similarly, at point B in Figure (b), $f'(x0)=0$. Therefore, we can conclude that an extremum can take place only at the stationary points where $f'(x)=0$. The extreme values are always stationary values.

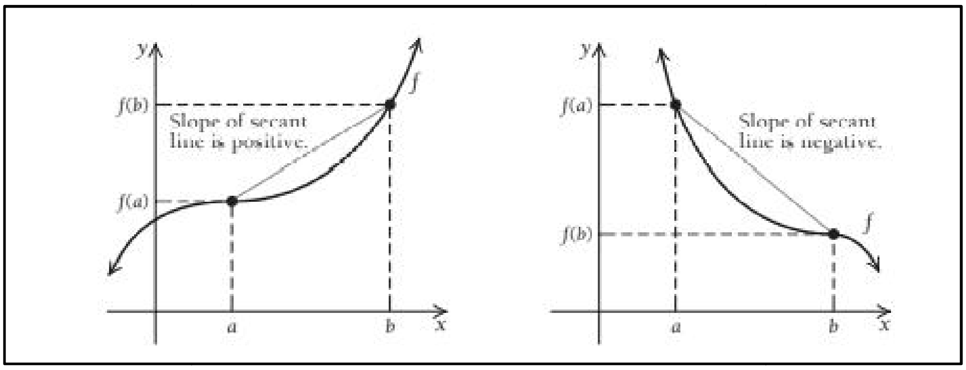
Definition:

Let f be defined on an interval I containing 'c'

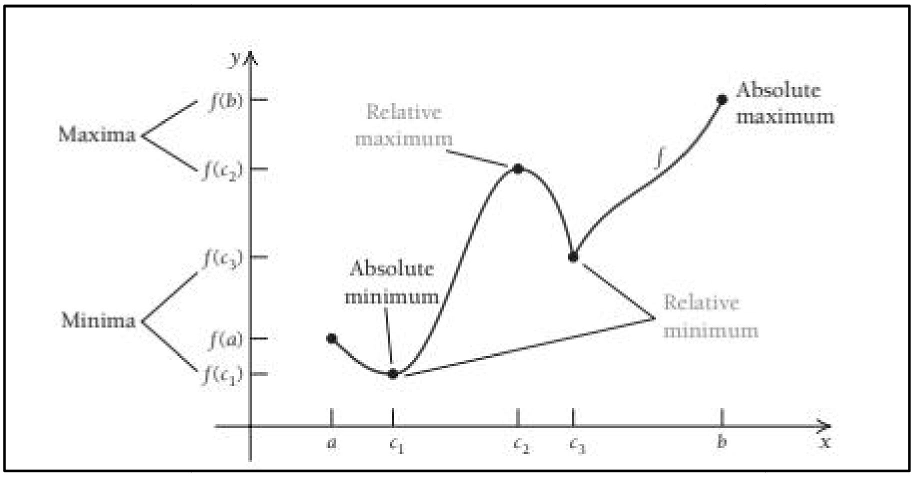
1. f (c) is the (absolute) minimum of f on I if $f(c) \leq f(x)$ for all x in I.

2. f (c) is the (absolute) maximum of f on I if $f(c) \geq f(x)$ for all x in I.

The minimum and maximum of a function on an interval are called the extreme values or extreme, of the function on the interval.



2. If f is a continuous function defined on a closed and bounded interval [a,b], then f has both a minimum and a maximum value on the interval [a,b]. This is called the extreme value theorem and its proof is beyond the scope of our course.



Note that at x = x0, the point A on graph is not an absolute maximum because f(x2) > f(x0). But if we consider the interval (a,b), then f has a maximum value at x = x0 in the interval (a,b). Point A is a point of local maximum off. Similarly, f has a local minimum at point B.

5.3 First Derivative Test and Relative Optima

If the first derivative of a function f (x) at x = x_0i.e., if f ' (x_0) = 0, then the value of the function at x0 i.e., f (x_0) will be

i) A relative maximum if the derivative f ' (x) changes its sign from positive to negative from the immediate left of the point x0to its immediate right.

ii) A relative minimum if the derivative f ' (x) changes its sign from negative to positive from the immediate left of the point x_0 to its immediate right.

iii) Neither a relative maximum nor a relative minimum if f ' (x) has the same sign on both the immediate left and right of the point x_0.

The value of the dependent variable, at which, the first derivative of the function is equal to zero i.e. at x_0 is referred to as the critical value of x. The value of the function at its critical point i.e.f (x_0) is known as the stationary value. The point with the coordinates equal to x_0 and f (x_0) is accordingly called the stationary point.

Point A is a relative maximum because for all values of x in the immediate left of x_2, the function is rising (the first derivative of f (x) is positive) and for all values of x in the immediate right of x_2, the function is falling (the first derivative of f (x) is negative). It is only at x_2, the critical point, the first derivative of the function is zero and f (x_2) is the corresponding stationary value. Note, the slope of the tangent to the function at A i.e. AT is parallel to the x-axis and is equal to zero. Analogously, one can see that point B is a point of relative minimum.

Example 1 $y = 50 + 90x - 5x^2$

Solution:

Step 1: Find the first derivative of the function.

$F'(x) = 90 - 10x$ or $10(9-x)$

Step 2: Equate the first derivative of the function to zero

$10(9-x) = 0$

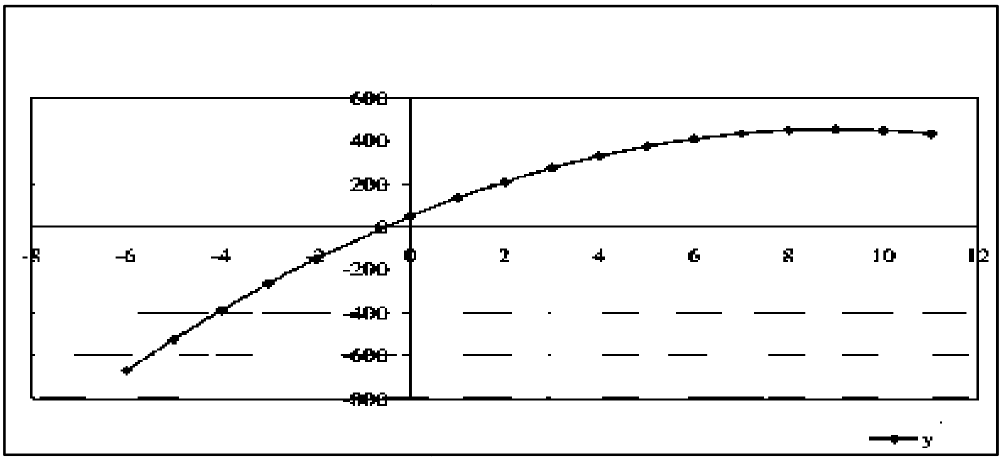
Step 3: Solve for x to obtain the critical value

The critical value of the function is $x = 9$ and the corresponding stationary value is

$y = f(9) = 455$.

x	-6	-3	0	3	6	9	10	11
y	-670	-265	50	275	410	455	450	435

The graph of this function is shown: $y = 50 + 90x - 5x^2$



It is easily verified that for all points in the neighborhood left of $x = 9$, the function is increasing, implying that its first derivative is positive. Similarly, for all points in the immediate neighborhood right of $x = 9$, the function is decreasing, implying its first derivative is negative. This satisfies condition (i) of the first derivative test and establishes, the critical value of $x = 9$ (located in the peak of the hill!) as a relative maximum. The corresponding stationary value of the function is $y = 455$.

Example 2: Consider the following function whose domain is assumed to be the interval $[0, \alpha)$.

$Y = (1/3)x^3 - x^2 + x + 10$

Differentiating with respect to x we get the first derivative as

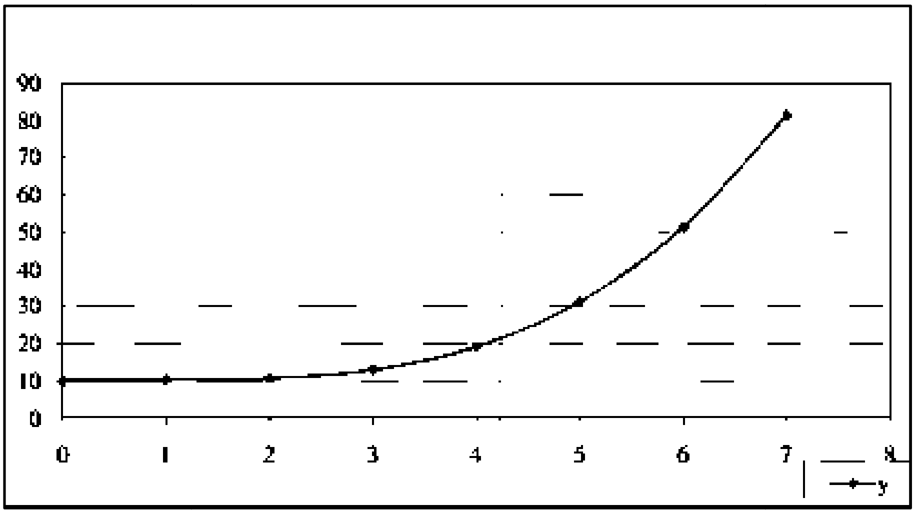
$x^2 - 2x + 1$ or $(x-1)^2$

Setting the first derivative of the function equal to zero yields $x = 1$ as a critical value of x. The corresponding stationary value of y is 10.33. To ascertain whether the stationary value is also a relative extremum we have to perform the first derivative test. The graphic representation of the function in Example 3 is as follows:

x	0	1	2	3	4	5	6
y	10	10.33	10.66	13	19.33	31.66	52

The graph of the function is presented $y = (1/3)x^3 - x^2 + x + 10$

The function attains a zero slope at the point where x = 1. Even though f ′ (1) is zero-which implies f (1) is a stationary value— the derivative does not change its sign from the left -hand side neighborhood of x = 1 to the other. In fact, as confirmed by the graph, the function is more or less flat in the immediate region of x = 1. On the basis of the first derivative test mentioned earlier it can be asserted that the stationary value f (1) = 10.33 is neither a relative maximum nor a relative minimum.



In sum, a relative extremum must be a stationary value (although the reverse is not necessarily true). To find the relative maximum or minimum of a given function, the first step would be to find the critical value of the dependent variable at which the first derivative of the function is equal to zero. This will enable us to find the stationary values of the function. To ascertain whether the stationary value is also a relative maximum or relative minimum one needs to apply the first derivative test.

Note:

First Derivative Test for Local Maxima and Minima Let c be a critical number of f i.e., f´(c) = 0

If f´(x) changes sign from positive to negative at c then f (c) is a local maximum.

If f´(x) changes sign from negative to positive at c then f(c) is a local minimum

5.4 The Second Order Derivative and Second-Order Condition For Optimum

Assuming that the first derivative f'(x) is itself a function of x, the second derivative of the function is obtained by differentiating this function again with respect to x. Symbolically, the second derivative is represented as f''(x). The double prime indicates that the function y = f (x) has been differentiated twice with respect to x. The expression (x) following the double prime indicates that the second derivative is also a function of x. If the second derivative f''(x) exists for all values in the domain, the function f (x) is said to be twice differentiable; if, in addition, f''(x) is continuous, the function f (x) is said to be twice continuously differentiable.

Example 3: Find the second derivative of the following function

$Y= f(x)=4x^3 +5x^2 -3x+10$

Step 1: Differentiate the equation in Example 4 with respect to x to find the first derivative. We obtain the following equation:

$f'(x) = 12x^2 +10x-3$

Step 2: Now differentiate this equation with respect to x to obtain the second derivative of the original function:

$F''(x) = 24x+10$

5.5 Interpretation of The Second Order Derivative

The first derivative of the function i.e., $f'(x)$ measures the slope of the function or the rate of change of the function. If the first derivative is positive, i.e., if $f'(x) > 0$, then the function is increasing; and if the derivative is negative i.e., if $f'(x) < 0$, then the function is decreasing. Analogously the second derivative i.e., $f''(x)$ measures the rate of change of the first derivative $f'(x)$. If the second derivative is positive i.e., if $f''(x) > 0$, then there is an increasing rate of change; and when $f''(x) < 0$, then the rate of change is decreasing. In other words, the second derivative measures the rate of change of the rate of change of the original function f (x). Note that if $f'(x) > 0$and $f''(x) > 0$, then this means that the function has a positive slope which is changing at an increasing rate. In other words, the function is said to be increasing at an increasing rate. Conversely, if $f'(x) < 0$and $f''(x) < 0$, then this means that the function has a negative slope which is changing at a decreasing rate. In other words, the function is said to be decreasing at a decreasing rate.

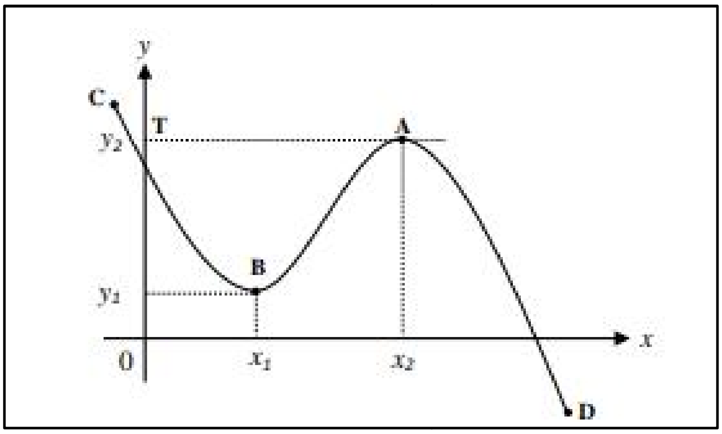
The Second Order Derivative Test

This test uses the second derivative of the function in question, hence the name.

Assume that $f(x_0) = 0$, (so x0 is a critical point!), then

1) If $f''(x_0) > 0$ then $f(x_0)$ is a relative minimum value.

2) If $f''(x_0) < 0$ then $f(x_0)$ is a relative maximum value.



As mentioned earlier, the zero slope condition, in other words $f'(x) = 0$at x = x0, was deemed to be a 'necessary condition' for f (x0) to be a relative extremum. Since this is based on the first derivative of the function, it is also known as the first-order-condition. Once we verify that the first order condition is satisfied, the negative (positive) sign of $f''(x)$ at x = x0 is sufficient to ensure that x0 corresponds to a relative maximum (minimum). Since the sufficiency condition is based on the second derivative of the function, it is also referred to as the second-order-condition.

To get a clearer understanding of how the second derivative enables us to determine whether the stationary value is a 'relative maximum' or a 'relative minimum'. Recall, the extreme values of this function lies in the level stretch, either in the bottom of the hill (point B) or at the peak of the hill (point A). Around the point on the graph corresponding to the relative maximum value (point A), the graph is concave down. This makes sense, since A (x2, y2) is the highest point on the graph in an interval around x2, so the graph must 'bend down' away from the peak, making it concave down. If we pick any two points in this region of the graph, then the straight line joining these two points will lie entirely below the graph except for the two end points on the curve.

This ensures that at x equal to x2, $f''(x2) < 0$ is satisfied. Similarly, around the point B (x1, y1) the graph is convex up. Once again, this makes sense. This point is the lowest point on the graph in an interval around x1, so the graph has to 'bend up' away from the point (x1, y1), making it convex up. If we pick any two points in this region of the graph, then the straight line joining these two points will lie entirely above the graph except for the two end points on the curve. This ensures that at x equal to x1, $f''(x1) > 0$ is satisfied.

For a Relative Maximum		For a Relative Minimum	
$dy/dx = 0$	(First-order Condition)	$dy/dx = 0$	(First-order Condition)
$d^2y/dx^2 < 0$	(Second-order Condition)	$d^2y/dx^2 > 0$	(Second-order Condition)

Example 4: Find the maxima and minima for the following function:

$y = 3x^4 - 10x^3 + 6x^2 + 5$

Solution: First order condition (F.O.C.)

$12x^3 - 30x^2 + 12x = 0$ or,

$3x (4x - 2) (x - 2) = 0.$

Either, $x = 0$ or, $x = 2$ or $x = ½$

Second, order condition (S.O.C.):

At $x = 0$, $f''(x) = 12>0$.

At $x = 2$, $f''(x) = - 9 < 0$

Hence the function attains maximum at $x = ½$ and minimum at $x=0$ and $x = 2$.

Guidelines to find Local Maxima and Local Minima

The function f is assumed to posses the second derivative on the interval I.

Step 1 : Find f ' (x) and set it equal to 0.

Step 2 : Solve f ' (x) = 0 to obtain the critical numbers of f.

Let the solution of this equation be α, ……….

We shall consider only those values of x which lie in I and which are not

end points of I.

Step 3 : Evaluate f ´ (α)

If f (α) < 0, f (x) has a local maximum at x = α and its value if f (α)

If f (α) > 0, f (x) has a local minimum at x = α and its value if f (α)

If f (α) = 0, apply the first derivative test.

Step 4 : If the list of values in Step 2 is not exhausted, repeat step 3, with that

Value

Example 5: Find the points of local maxima and minima, if any, of each of the following functions. Find also the local maximum values and local minimum values.

a. $f (x) = x^3 -6x^2+9x+1$

Solution: $f (x) = x^3 -6x^2+9x+1$

Thus, $f'(x)=3x^2 -12x +9= 3(x^2 -4x+3) =x(x-1)(x-3)$

To obtain critical number of f, we set f ´ (x) = 0 this yields x = 1, 3.

Therefore, the critical number of f are x = 1, 3.

Now f ´ (x) = 6x – 12 = 6(x – 2)

We have f ´ (1) = 6(1 – 2) = – 6 < 0 and f (3) = 6(3 – 2) = 6 > 0

Using the second derivative test, we see that f (x) has a local maximum at

x = 1 and a local minimum at x = 3. The value of local maximum at x = 1 is f (1) = 1– 6+ 9 +1 = 5 and the value of local minimum at x = 3 is

f (3) = 3^3– 6(3^2) + 9(3) + 1 = 27 – 54 + 27 + 1 = 1.

5.6 First-Order Condition for Objective Function with Two Variable

Assume that

Z= f(x,y)

The first-order necessary condition for an extremum (either maximum or minimum) again involves dz = 0. However, now since there are two choice variables, the first-order condition is modified as follows:

dz = 0 for arbitrary non-zero values of dx and dy

The rationale behind this is similar to the explanation of the condition dz = 0 for the one variable case: an extremum point must necessarily be a stationary point; at a stationary point dz = 0 ever for infinitesimal change in the two variables x and y. Totally differentiating equation (3), we get

$D_z = f_x dx + f_y dy$

where f x = df/dx = partial derivative with respect to x and fy = df/dy =partial derivative with respect to y.

Now dx ≠ 0; dy ≠ 0

And at the stationary point dz = 0

(A) and (B) can hold simultaneously only if fy= f x=0

Hence the first-order condition for optimisation (location of extremum points) for an objective function with two variables is:

$$f_x = f_y = 0, \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 0$$

the first-order condition is necessary but not sufficient. To develop the sufficiency condition, we must look into the second-order to the differential, which is related to second-order partial derivatives.

Example 6: Assuming that the second-order condition is satisfied, find out the profit maximising values of quantity (Q) and advertising expenditure (A) for a producer with the following profit function (Π):

$\Pi = 400 - 3Q^2 - 4Q + 2QA - 5A^2 + 48A$

The first-order condition for profit-maximisation requires that the partial derivative

$$\frac{\partial \Pi}{\partial Q} = \frac{\partial \Pi}{\partial A} = 0$$

Partially differentiating equation (4) with respect to A keeping Q as constant gives us:

$$\frac{\partial \Pi}{\partial A} = 2Q - 10A + 48$$

Partially differentiating equation (4) with respect to Q keeping A as constant gives us:

$$\frac{\partial \Pi}{\partial A} = -6Q - 4 + 2A$$

Since $\frac{\partial \Pi}{\partial Q} = \frac{\partial \Pi}{\partial A} = 0$ as the first order condition we get:

$$2Q - 10A + 48 = 0$$

And

$$-6Q - 4 + 2A$$

Equations (5) and (6) can be written as:

Q-5A=-24

-3Q+A=2

From equation (8), we get A=3Q+2 . Substituting for A in equation (7) and we then solve for Q as follows:

Q-5(2+3Q)=-24

Q-10-15Q=-24

-14Q=14

Q=1

Hence A=2+3*1=5

So, we obtain that Q =1 and A = 5 are the quantity and advertising expenditure level which maximizes profit for the firm using the first-order necessary condition.

5.7 Second-Order Condition for Objective Function with Two Variables

Let us now examine the sufficiency condition for optimisation where the objective function has two decision variables. Using the concept of d2 z, we can state the second-order sufficient condition for a maximum of z = f(x,y) as follows:

$d^2 z < 0$ for arbitary non-zero values of dx and dy.

The rationale behind this is similar to that of the d2 z condition explained in the case where the objective function has one variable. Analogously, the second-order sufficiency condition for a minimum of z = f(x,y) is the following:

$d^2 z > 0$ for arbitary non-zero values of dx and dy.

Note that d2 z is a function of the second order partial derivatives fxx, fxy and fyy. Intuitively, it is clear that the second-order sufficiency condition can be translated in terms of these derivatives. However, the actual translation would require knowledge of quadratic form – the discussion of which is given in section 8.5. Hence, we will state the main results here:

For any values of dx and dy, not both zero:

To sum up, the first-order and the second-order condition for optimization in case of an objective function z = f(x,y) is depicted in the following:

Table 1:First-Order and Second-Order Condition for Optimisation

	Maximum	**Minimum**
First-Order Necessary Condition	$f_x = f_y = 0$	$f_x = f_y = 0$
Second-Order Sufficiency Condition	$f_{xx}, f_{yy} < 0$ *and* $f_{xx}f_{yy} > f^2_{xy}$	$f_{xx}, f_{yy} > 0$ *and* $f_{xx}f_{yy} > f^2_{xy}$

The second order condition is applicable only after the first-order condition has been fulfilled.

we had solved for the optimal product (Q) and the advertising expenditure (A), assuming that second-order condition is satisfied at the optimal point. Now let us examine whether it is actually satisfied or not.

Recall that

$\Pi = 400 - 3Q^2 - 4Q + 2QA - 5A^2 + 48A$

$\Pi_A = \frac{\partial \Pi}{\partial A} = 2Q - 10A + 48$

$\Pi_Q = \frac{\partial \Pi}{\partial A} = -6Q + 2A - 4$

Setting them equal to zero and solving for Q and A yields Q* = 1 and A* = 5, where * denotes the optimal level.

For the second-order condition we need to derive the following partial derivatives:

$\frac{\partial \Pi^2}{\partial A^2}, \frac{\partial \Pi^2}{\partial Q^2}$ *and* $\frac{\partial \Pi^2}{\partial A \partial Q}$

Partial differentiation of equation with respect to A gives us:

$\Pi_{AA} = \frac{\partial \Pi^2}{\partial A^2} = -10 < 0$

Partial differentiation of equation with respect to Q gives us:

$\Pi_{QQ} = \frac{\partial \Pi^2}{\partial Q^2} = -6 < 0$

Partial differentiation of equation with respect to A gives us:

$\Pi_{AQ} = \frac{\partial \Pi^2}{\partial A \partial Q} = 2$

Now

$\Pi_{AA} * \Pi_{QQ} = -10 * -6 = 60$

and $(\Pi AQ)^2 = (2)^2 = 4$

Hence $\Pi_{AA} * \Pi_{QQ} > (\Pi_{AQ})^2$

Second-order condition is satisfied for the output level (Q) equal to 1 and the advertising expenditure (A) equal to 5.

5.8 Constrained Optimization Problems

In economic applications, unconstrained optimization is a relatively rare case. More often, we find instances of constrained optimization, that is optimization subject to a constraint. What this means is that there is a side condition on the optimization exercise. The domain of the function that is sought to be optimized is restricted by one or more side relations. Consider the case of utility maximization by a consumer. An individual as consumer has unlimited wants. But she is constrained by her budget set. So, she maximizes her utility subject to her budget constraint. Similarly, a producer would want to minimize her cost subject to a given level of output.

The unit begins with a discussion of how to find stationary values of the objective function. It is shown that in simple cases we can use the method of substitution. But in more complex cases, different methods have to be employed. We will see the importance of the Lagrangian multiplier and how it is used in static constrained optimization. After this the unit discusses second-order conditions for constrained optimization.

The Method of Substitution:

With the help of an example from co-ordinate geometry, we shall learn to maximize/minimize a function subject to a constraint which restricts the domain of the function. Consider a simple example. Find the smallest circle centered on (0,0) which has a point common with the straight-line

$x+y = 10$.

The equation for the circle is

$x^2 + y^2 = r^2$.

The smallest circle will be one with the smallest radius. The restriction is that the circle must have a point in common with a given straight line. Without this restriction, the smallest circle can easily be seen to be a circle with radius zero, i.e. a point.

Thus, our problem is the following:

Minimise

$x^2 + y^2$ subject to $x + y = 10$

Here the unconstrained solution x=0, y=0 will not be available. The constraint prohibits this solution. What we have to do is to consider as the domain only those values of x and y for which x + y = 10. So, we see that the constraint has diminished the domain. How to find the solution? From the constraint we find y x = − 10.

Now if we substitute this into the minim and x2 + y2 we get (X2) + (10-X)2, which incorporates the constraint.

Let us now minimize this expression

$$\frac{d^2}{dx^2}[x^2 + (10-x)^2] = 2x + 2(10-x)(-1) = 4x - 20$$

For a stationary value 4x-20 =0 or x=5

To check whether the stationary value is truly a minimum we differentiate the function once again with respect to x. Thus,

$$\frac{d^2}{dx^2}[x^2 + (10-x)^2]$$

= d/dx=(4x-20)

=4>0

Which proves that we have minimized the expression $x^2 + y^2$

We know that the constraint makes y =10-x= 5, thus giving us a solution x=5,y = 5 for which $x^2 + y^2$ =5

In solving the problem our method involved several steps:

Step 1: Solve the constraint equation to get one variable in terms of the other. Above we found y in terms of x. Thus, y x = − 10

Step 2: Substitute this solution in the objective function to ensure that the domain of the function is a set of pairs of values of x and y which satisfy the constraint. Note that the objective function thus modified is a function of x only.

Step 3: Differentiate this modified function to get the first derivative. Find the value of x for which this derivative equal zero.

Step 4: Put this value of x in the constraint to find the value of y.

Step 5: Calculate the value of the objective function for this pair of values of x and y.

Step 6: Check the second order condition to find whether the stationary point is an extremum

In general, it may not be easy or indeed possible to solve the constraint equation. In such cases it would appear that we are stuck at step (1) of the procedure described above. For instance, we may have a constraint equation like x 3 + 2x2 y + 9y3 - 2y – 117 = 0. To find y in terms of x, or x in terms of y from this complicated equation is extremely difficult. Suppose now that the problem has been referred to a mathematician and while we wait for the solution, we prepare the ground for our computation.

The Lagrange Multiplier Method

Consider once more the problem:

Maximise f(x, y) subject to g(x, y) = 0

If gy(x, y) ≠ 0, then y = h(x), so that the problem is transformed into:

Maximise f(x,h(x))

1st order condition gives us $f_x + f_y h'(x) = 0$ ……..(i)

g(x,h(x)) = 0 is an identity so that we have

$g_x + g_y h'(x) = 0$

Now define $\lambda = \frac{f_y}{g_y}$Multiplying (ii) by $\lambda = \frac{f_y}{g_y}$, we get

$\lambda g_x + f y^{h'(x)} = 0$

From (i) and (iii), fx- λ gx=0

From, $2 = \frac{f_y}{g_y}, f_y - h g_y = 0$

Also the constraint is g(x,y) = 0

(1), (2) and (3) are three equations in three variables, x, y and λ. When we solve these three simultaneous equations, we get the value of x and y which solve our problem; we also get the value of λ which is Lagrangian multiplier, something which does not seem to have any relevance to our problem.

Example 12: Maximise $5x^2 + 6y^2 - xy$, subject to constraint x+2y=24

Solution: Given OF: $5x^2 + 6y^2 - xy$

CF: x+2y = 24

24-x-2y= 0

Let v=OF+ λCF $=5x^2 + 6y^2 - xy + 24\lambda - \lambda x - 2\lambda y$

$v_x = 10x + 0 - y + 0 - 2 - 0 = 0, 10x - y = \lambda$

$v_y = 0 + 12y - x + 1 - 0 - 2\lambda = 0 \; or \; 12y - x = 2\lambda$

From (13) and (14) we get 20x-2y=12y-x, x=2/3y

Since we do not get clear values of x and y,

Therefore, we find $v\lambda$ =2u-x-2y=0

24-2/3y-2y=0

(2/3+2) y=24

Y=9

X=6

Hence constrained maximization takes place when x= 6 and y=9

A consumers utility function is given as u=(y+1) (x+2) . If his budget constraint is 2x+5y=51, how of x and y he should consumer to maximize his satisfaction.

5.9 Economics Applications of Maximum-Minimum

Example 7: A stereo manufacturer determines that in order to sell x units of a new stereo, the price per unit, in dollars, must be

P(x) = 1000 – x

The manufacturer also determines that the total cost of producing x units is given by

C(x) = 3000 + 20x.

a) Find the total revenue.

b) Find the total profit.

c) How many units must the company produce and sell in order to maximize profit?

d) What is the maximum profit?

e) What price per unit must be charged in order to make this maximum profit?

Solution:

a. R(x) = Total revenue

= (Number of units)*(Price per unit)

=x*p

$=x(1000-x)=1000x-x^2$

b. P(x)= Total revenue - Total cost

= R(x)-C(x)

=(1000-x^2)-(3000+20x)

=-x^2 +980x-3000

c. To find the maximum value of , we first find P'(x)

P'(x) = - 2x + 980.

This is defined for all real numbers, so the only critical values will come from solving P'(x)

- 2x + 980=0

X=490

There is only one critical value. We can therefore try to use the second derivative to determine whether we have an absolute maximum.

P''(x) = - 2

Thus,P''(490) is negative, and so profit is maximized when 490 units are produced and sold.

d) The maximum profit is given by

P(490)=-$(490)^2$ +980*490-3000

=237100

Thus, the stereo manufacturer makes a maximum profit of $237,100 by producing and selling 490 stereos.

e) The price per unit needed to make the maximum profit is

4 p = 1000 - 490 = 510.

Example 8: A monopolist sells two products x and y for which the demand functions are

x =25- 0.5Px

y =30 – Py

and the combined cost function is

c = x^2 + 2xy + y^2 + 20

Find (a) the profit-maximizing level of output for each product, (b) the profit-maximizing price for each product, and (c) the maximum profit.

a) Since Π = TRx+ TRy -TC, in this case

$\Pi = P_X x + P_y y - c$

Px =50 -2x

Py =30-2y

Substituting in

Π =(50 -2x)x + (30 - y)y -(x^2 +2xy + y^2 + 20)

=50x– 3x^2 +30y-2y^2 -2xy-20

The first-order condition for maximizing is:

Π x = 50 - 6x -2y = 0

Π y = 30 – 4y -2x= 0

Solving simultaneously, $\bar{x}$ = 7 and $\bar{y}$ =4. Testing the second-order conditon, Π_{xx} =-6, Π_{yy} =-4, and Π_{xy} =-2. With both direct partials negative and $\Pi_{xx}\Pi_{yy} > (\Pi_{xy})^2$, Π is maximized

b) Substituting $\bar{x}$ = 7, $\bar{y}$ = 4

Px=50 -2(7) = 36 Py=30-4= 26

c) Substituting $\bar{x}$ = 7, $\bar{y}$ = 4 in Π =215.

(xy) 2 , is maximized.

Example 8:Let the inverse demand function and the cost function be given by

P = 50 − 2Q and C = 10 + 2q

respectively, where Q is total industry output and q is the firm's output.

First consider first the case of uniform-pricing monopoly, as a benchmark. Then in this case Q = q and the profit function is

$\pi(Q) = (50 - 2Q)Q - 10 - 2Q = 48Q - 2Q^2 - 10$

solving $d\pi/dQ = 0$ we get Q = 12, P = 26, π = 278, CS = 12(50−26) /2 = 144, TS = 278 + 144 = 422

Monopoly:

Q	P	π	CS	TS
12	26	278	144	422

Now let us consider the case of two firms, or duopoly. Let q1 be the output of firm 1 and q2 the output of firm 2. Then Q = q1 + q2 and the profit functions are:

$\pi_1(q_1, q_2) = q_1[50 - 2(q_1 + q_2)] - 10 - 2q_1$

$\pi_2(q_1, q_2) = q_2[50 - 2(q_1 + q_2)] - 10 - 2q_2$

A Nash equilibrium is a pair of output levels (q_1, q_2) such that:

$$\pi 1(q_{1,}^*q_2^*) \geq \pi_1(q_{1,}q_2^*)q_1 \geq 0$$

And

$$\pi_1(q_{1,}^*q_2^*) \geq \pi_1(q_{19}^*q_2) \geq 0$$

This means that, fixing q_2 at the value q_2 and considering π * 1 as a function of q_1 alone, this function is maximized at q_1 = q. But a necessary condition for this to be true is that $\frac{\partial \Pi_1}{\partial q_1}(q_{1,}^*q_2^*) = 0$

Similarly, fixing q_1 at the value q_1 and considering π * 2 as a function of q_2 alone, this function is maximized at q_2 = q. But a necessary condition for this to be true is that $\frac{\partial \Pi_2}{\partial q_2}(q_{1,}^*q_2^*) = 0$

Thus the Nash equilibrium is found by solving the following system of two equations in the two unknowns q_1 and q_2:

$$\frac{\partial \Pi_1}{\partial q_1}(q_{1,}^*q_2^*) = 50 - 4q1 - 2q2 - 2 = 0$$

$$\frac{\partial \Pi_2}{\partial q_2}(q_{1,}^*q_2^*) = 50 - 2q1 - 4q2 - 2 = 0$$

The solution is q1*=q2*= 8 , Q = 16, P = 18, π * * = = 1 = π2 = 118, CS = 16(50−18) 2 = 256, TS = 118 + 118 + 256 = 492.

5.10 Economic Order Quantity

Inventory control is an important consideration in business. In particular, for each shipment of raw materials, a manufacturer must pay an ordering fee to cover handling and transportation. When the raw materials arrive, they must be stored until needed, and storage costs result. If each shipment of raw materials is large, few shipments will be needed, so ordering costs will be low, while storage costs will be high. On the other hand, if each shipment is small, ordering costs will be high because many shipments will be needed, but storage costs will be low. Example 8 shows how the methods of calculus can be used to determine the shipment size that minimizes total cost.

Example 9: A bicycle manufacturer buys 6,000 tires a year from a distributor. The ordering fee is $20 per shipment, the storage cost is 96 cents per tire per year, and each tire costs rupees 21. Suppose that the tires are used at a constant rate throughout the year and that each shipment arrives just as the preceding shipment is being used up. How many tires should the manufacturer order each time to minimize cost?

Solution: The goal is to minimize the total cost, which can be written as

Total cost =storage cost +ordering cost +purchase cost

Let x denote the number of tires in each shipment and C(x) the corresponding total cost in dollars. Then,

Ordering cost =(ordering cost per shipment)(number of shipments)

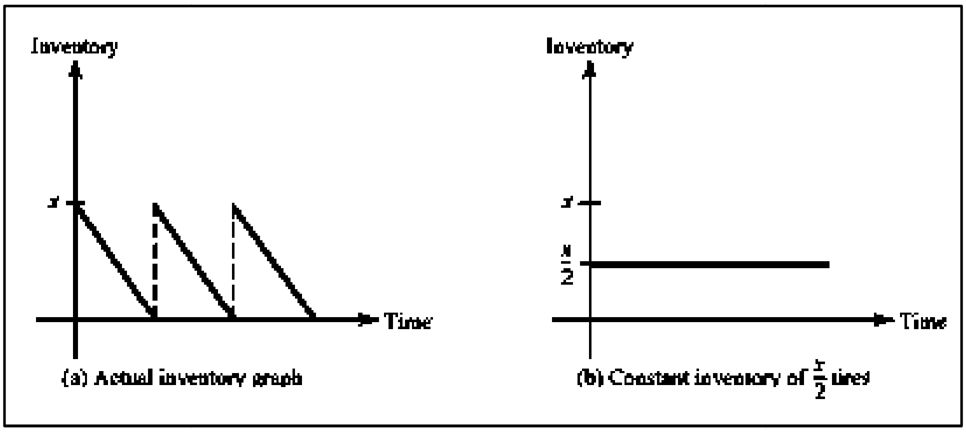
Since 6,000 tires are ordered during the year and each shipment contains x tires, the number of shipments is 6000/x and so

Ordering cost=20*(6000/x)

=120000/x

Purchase cost= (total number of tires ordered)(cost per tire) = 6,000(21)=126,000

The storage cost is slightly more complicated. When a shipment arrives, all x tires are placed in storage and then withdrawn for use at a constant rate. The inventory decreases linearly until there are no tires left, at which time the next shipment arrives. The situation is illustrated in Figure This is sometimes called just-in-time inventory management.



The average number of tires in storage during the year is x/2, and the total yearly storage cost is the same as if x/2 tires were kept in storage for the entire year. This assertion, although reasonable, is not really obvious, and you have every right to be unconvinced. It follows that

Storage cost= (average number of tires stored) (storage cost per tire)

=x /2 (0.96) =0.48

Putting it all together, the total cost is

C(x)= 0.48x+120000/x+126000

and the goal is to find the absolute minimum of C(x) on the interval

$0 \le x \le 6{,}000$

The derivative of C(x) is

C′(x) = 0.48 - 120,000 /x

which is zero when

x^2 =120000/0.48=250000 or x= ±500.

Since x = 500 is the only critical number in the relevant interval $0 \le x \le 6{,}000$, you can apply the second derivative test for absolute extrema. You find that the second derivative of the cost function is

C′′(x)= 240000/x^3

which is positive when x > 0. Hence, the absolute minimum of the total cost C(x) on the interval 0 ≤ x ≤ 6,000 occurs when x = 500; that is, when the manufacturer orders the tires in lots of 500.

5.11 Bilateral Monopoly

Example 10: As can he computed from the previous models, the profit under the joint profit maximizing model (Rs. 21,870) is greater than the sum of the profits under either monopoly 5400 + 12,150 = 17,550) or monopsony (18,225 + 3037.5 = 21,262.5) market. Hence, there is an incentive for the single buyer to move as close as possible to the profit level under the monopsony market (ap = 8,225). Similarly, there is an incentive for the single seller to move as close as possible to the profit level under monopoly market (xp= Rs. 12,150). Goals cannot be achieved simultaneously. In this case some hind of compromises are needed for the bilateral model to have a unique solution of the intermediate product price. Within this framework, we reformulate the bilateral monopoly problem into the following goal programming model.

Minimize $\alpha\Delta\pi_{zz} + \beta\Delta\Pi yy$

Subject to $\pi_z + \Delta\pi_{zz} = \pi_{zz}$

$\Pi y + \Delta\Pi yy = \Pi yy$

$\pi_z + \Pi y = \Pi$

π_{zz} and $\pi_{yy} \geq 0$

Where

$\pi_z = actualprofitlevelofthesinglebuyerinthebilateralmonopolymodel$

$\Pi y = actualprofitlevelofthesinglesellerinthebilateralmonopolymymodel$

$\Delta\pi_{zz} = thedeviationalvariable\ (underachievement)\ thatrepresentsthedifferferencebetweenthe$

$actualprofitlevelofthesinglebuyer\ (ora,)\ andtheprofitlevelunderthememonopsony$

ΔΠyy= the deviational variable (underachievement) that represents the difference between the actual profit level of the single seller (ore,) and the profit level under the monopoly market (ˊ˒»'d

Π *= totalprofitofthejointprofitmaximizationproblem

On one end, we could assume β= 0, i.e., the objective function is reduced to minimizing $\Delta\pi_{zz}$. As a result, the deviational variable for the single buyer would be zero, that is, $\Delta\pi_{zz}$= 0 and $\pi_z = \pi *_{zz}$= Rs. 18,225, Πy - 3645, and $\Delta\Pi yy$ = 8505 (or 12,150 - 3,645) with the same value ofx, y and z. This is equivalent to the revised solution with Py = 148.5. On the other end, we assume that α = 0, i.e., the objective function is reduced to minimizing $\Delta\Pi yy$. As a result, the deviational variable for the single seller would be zero, that is, $\Pi yy = 0$ and $\pi_y = \pi *_{yy}$ = 12,150, π_z = 9720, $\Delta\pi_{zz}$ = 8505 (or 18,225 - 9,720). This is equivalent to the revised solution with Py = 306. The value of the deviational variables (Rs. 8505) represents the amount of revenue either the single seller or single buyer falls short of the goal if the market were monopoly or monopsony, respectively. If the market were a monopoly, the maximum revenue goal would $\pi *_{yy}$= 12,150 for the single seller; and $\pi *_{zz}$ = 18,225 for the single buyer if the market were a monopsony. The sum of both a and a*p exceeds the maximum joint profit under the collusion (Rs.21,870) by Rs.8,505. The solution bounds on the revised model are a special case of the goal programming model or equations through in which either single seller or single buyer takes up all the amount of the underachievement from the goal. However, it is not likely that a single party would shoulder all the burden. An alternative solution to the bilateral monopoly model would be to assume that both parties stiare equal burden, i.e., $\Delta\pi_{zz}$ = $\Delta\Pi yy$.

5.12 Law of Equi-Marginal Utility

Let us assume that a consumer is consuming only two goods X1 and X2 . The utility which she receives from consuming X1 and X2 is given by the utility function U=f (X1 , X2) and satisfies the property of eventual diminishing marginal utility. The consumer has a given money income to be spent on these two goods during the period we are analysing her behaviour. She cannot influence the prices, P1 and P2,, of the goods, through her own action. Prices are given as parameters in decision-making (consumption) as this consumer is one of the numerous consumers demanding X1 and X2 . Thus, she has no market power. Since she is required to spend her entire income on X1 and X2 , the budget equation is given as,

$M = P_1 X_1 + P_2 X_2$

Where M is her nominal income. Since the consumer is a utility maximiser, her consumption problem can be formulated as follows:

Maximise

$U = f (X_1, X_2)$

subject to the budget constraint

$M = P_1 X_1 + P_2 X_2$.

By using the Marshallian Equi-Marginal Principle (which is based on the Lagrange Multiplier Technique), we get the equilibrium condition,

$MU_1 / P_1 = MU_2 / P_2 = \lambda$

This is the first order condition (necessary) for achieving equilibrium. The second order condition (sufficiency) of equilibrium is given by the law of eventual diminishing utility. The second order condition will be automatically fulfilled so long as the marginal utility schedules for each good MU1 and MU2 are both downward sloping. It must be noted that whenever a consumer maximizes utility, equilibrium is said to be attained. What the equilibrium condition says is that to maximise total utility [U = f (X1 , X2)] an individual consumer must equalize the ratio of marginal utility to price for each and every good and which in turn must be equal to constant marginal utility of money. In other words, to obtain maximum total utility a rational consumer must equalize the marginal utility per rupee of expenditure on each and every line of expenditure (that is, on each and every good).

This relationship represents the consumer's equilibrium condition. A consumer attains equilibrium when she maximizes total utility from consuming X_1 and X_2 . The equilibrium condition can also be stated in an alternative form:

$MU_I / MU_2 = P_1 / P_2$... (1)

The ratio of marginal utilities of goods X1 and X2 must equal the price ratio of the same two goods. This in turn, must equal marginal utility of money, which is constant by assumption.

Condition (1) above is the famous Marshallian law of the equi-marginal utility. It can be shown that if the ratio of marginal utility of the two goods is not equal to the price ratio, then without spending more in the aggregate, just by re-allocating the given amount of money income as between the two goods X1 and X2, the consumer can increase her total utility from consumption.

Example 11: The following table gives an individual's marginal utility schedules for goods X1 and X2. If the prices of X1 and X2 are Rs. 2.00 each and that the individual has Rs. 20.00 of Income, which she spends on X1 and X2, what is the individual's equilibrium purchase of X1 and X2?

Q 1 2 3 4 5 6 7 8 9 10 11
MU_1 16 14 11 10 9 8 7 6 5 3 1
MU_2 15 13 12 8 6 5 4 3 2 1 0

$MU_1/P_1 = MU_2 /P_2$

and the budget constraint must be fully satisfied. From the above table we derive the following.

Q 1 2 3 4 5 6 7 8 9 10 11
MU_1 / P_1 8 7 5.5 5 4.5 4 3.5 3 2.5 1.5 0.5
MU_2 / P_2 7.5 6.5 6 4 3 2.5 2 1.5 1 0.5 0

At $X_1 = 6$ MU_1 / P_1 is 4 At $X_2 = 4$ units MU_2 / P_2 is 4 Hence $MU_1/ P_1 = MU_2 / P_2 = 4$.

The amount spent is $P_1 X_1 + P_2 X_2$ which is Rs.20.00 (2*6+2*4 = 12 + 8). Money income is also Rs. 20.00. Hence, the budget constraint is satisfied. The equilibrium purchase is $X_1 = 6$ units and $X_2 = 4$ units.

Since MU_1 falls from 16 to 1 as X_1 increases from 1 to 11 and MU_2 declines from 15 to zero as X_2 increases from 1 to 11, the second order condition is also fulfilled.

When there are more than one combination of two goods (X_1 , X_2) at which the equimarginal principle holds, one has to take recourse to the budget constraint to obtain the equilibrium combination and all other combinations violating the budget constraint have been rejected.

It should be noted that when the consumer consumes n goods, the law of equi- marginal utility would then read as:

$MU_1 / P_1 = MU_2 / P_2 = MU_3 / P_3 = ... = ..MUn / Pn = \lambda$ (the marginal utility of money)

with the second-order conditions (the law of eventual diminishing marginal utility must hold for each of the n goods).

Summary

- Stationary Points: The points, at which first order derivatives are zero, are called stationary points. The values of the function at those points are called stationary values.
- Local (relative) Maxima and Global (absolute) Maxima: A function attains local maximum at any particular point in the domain of definition implies that the value of the function is maximum in the neighborhood of that point only, and the function can assume higher values elsewhere. This can be non-unique. The global maximum is the highest value of the function with reference to its entire domain and hence unique.
- Local (relative) Minima and Global (absolute) Minima: A function attains local minimum at any particular point in the domain of definition implies that the value of the function is minimum in the neighborhood of that point only, and the function can assume lower values elsewhere. This can be non-unique. The global minimum is the lowest value of the function with reference to its entire domain and hence unique.
- Points of Inflextion: The point of inflextion is defined as a point at which a curve changes its curvature. The sufficient condition for a point of inflextion is f x and f x ''( ) 0 '''( ) 0 = ≠ . Thus, if a function has '( ) 0, ''( ) 0 '''( ) 0 f x f x and f x == ≠ at the point x, the point is said to be stationary and inflextional.

Keywords

Minima: minimum point

Maxima: Maximum Point

Absolute minima: exactly minimum of all point sin consideration

Relative minima: comparatively minimum point

Saddle point: stationary point

Self Assessment

1. Select the correct necessary condition, in case of Maxima and Minima in Multi-variable function.

A. $\partial u/ \partial x=\partial u/ \partial y=0$

B. $\partial u/ \partial x=\partial u/ \partial y\neq 0$

C. ∂u/ ∂x=∂u/ ∂y=1

D. ∂u/ ∂x=∂u/ ∂y≠1

2. What is the saddle point?
A. Point where function has maximum value
B. Point where function has minimum value
C. Point where function has zero value
D. Point where function neither have maximum value nor minimum value
3. For function f(x,y) to have minimum value at (a,b) value is?
A. fxx*fyy– (fxy)2>0 and fxx, fyy<0
B. fxx*fyy– (fxy)2<0 and fxx, fyy <0
C. fxx*fyy– (fxy)2>0 and fxx, fyy >0
D. fxx*fyy– (fxy)2=0 and fxx, fyy =0

4. For function f(x,y) to have maximum value at (a,b) value is?
A. fxx*fyy– (fxy)2>0 and fxx, fyy>0
B. fxx*fyy– (fxy)2>0 and fxx, fyy <0
C. fxx*fyy– (fxy)2<0 and fxx, fyy >0
D. fxx*fyy– (fxy)2=0 and fxx, fyy =0

5. For homogeneous function with no saddle points we must have the minimum value as ______________
A. 90
B. 1
C. equal to degree
D. 0

6. The drawback of Lagrange's Method of Maxima and minima is?
A. Maxima or Minima is not fixed
B. Nature of stationary point is cannot be known
C. Accuracy is not good
D. Nature of stationary point is known but cannot give maxima or minima

7. Maximize the function x + y – z = 1 with respect to the constraint xy=36.
A. 0
B. -8
C. 8
D. No Maxima exists

8. What does the first order partial derivative shows in cost maximization?
A. Slope of isoquant curve= slope of Iso-cost lines
B. Rate of change in isoquant curve= slope of Iso-cost lines
C. Slope of isoquant curve= Rate of change in Iso-cost lines
D. Slope of isoquant curve ≠ slope of Iso-cost lines

9. What will be the slope under first order condition for maximization of utility?

A. $dx_2/dx_1=fx_1/fx_2$

B. $dx_2/dx_1=-fx_1/fx_2$

C. $dx_1/dx_2=fx_1/fx_2$

D. $dx_2/dx_1=fx_2/fx_1$

10. What is the second order condition for maximization of utility?

A. $d_2x_1/dx_{22}>0$

B. $d_2x_1/dx_{22}<0$

C. $d_2x_2/dx_{21}>0$

D. $d_2x_1/dx_{22}=0$

11. What does the first order condition show about maximization of output?

A. Slope of isoquant curve be greater than slope of iso-cost line

B. Slope of isoquant curve different from slope of iso-cost line

C. Slope of isoquant curve must be same as slope of iso-cost line

D. None of the above

12. What is the second order condition for maximization of utility?

A. $p_2/p_1=fx_1/fx_2$

B. $p_1/p_2=fx_1/fx_2$

C. $p_2/p_1=fx_2/fx_1$

D. $p_2/p_1=1$

13. What is Nash Equilibrium?

A. A pair of output levels at which the firms are being maximized.

B. Nash equilibrium in game theory is a situation in which a player will continue with their chosen strategy, having no incentive to deviate from it.

C. Only a

D. Both a and b

14. What is duopoly?

A. Where a number of firms are more than 2

B. Where number of firms are equal to 2

C. Where only single firms exist

D. None of the above

15. What are two type of costs on which Economic order quantity depends?

A. Inventory carrying costs

B. Procurement or set up costs.

C. Both a and b

D. None of the above

Answers for Self Assessment

1.	A	2.	D	3.	C	4.	B	5.	D

6.	A	7.	D	8.	A	9.	B	10.	A
11.	C	12.	B	13.	D	14.	B	15.	C

Review Questions

1. Find the maximum profit and the number of units that must be produced and sold in order to yield the maximum profit. Assume that revenue, R(x), and cost, C(x):
   a. $R(x)= 2x$, $C(x) = 0.01x^2 + 0.6x + 30$
   b. $R(x) = 50x - 0.5x^2$, $C(x)= 10x + 3$
2. Raggs, Ltd., a clothing firm, determines that in order to sell x suits, the price per suit must be
   $P=150-0.5x$
   It also determines that the total cost of producing x suits is given by
   $C(x)= 4000+0.25x^2$
   a. Find the total revenue, .
   b. Find the total profit, .
   c. How many suits must the company produce and sell in order to maximize profit? d) What is the maximum profit?
   d. What price per suit must be charged in order to maximize profit
3. A firm faces the production function Q = 20K 0.4L 0.6. It can buy inputs K and L for Rs. 400 a unit and Rs.200 a unit respectively. What combination of L and K should be used to maximize output if its input budget is constrained to Rs. 6,000?
4. A consumer spends all her income of Rs. 120 on the two goods A and B. Good A costs Rs.10 a unit and good B costs Rs.15. What combination of A and B will she purchase if her utility function is $U = 4A^{0.5} B^{0.5}$?
5. A store expects to sell 800 bottles of perfume this year. The perfume costs Rs.20 per bottle, the ordering fee is Rs.10 per shipment, and the cost of storing the perfume is 40 cents per bottle per year. The perfume is consumed at a constant rate throughout the year, and each shipment arrives just as the preceding shipment is being used up. a. How many bottles should the store order in each shipment to minimize total cost? b. How often should the store order the perfume?
6. Use this information to classify each critical number of f (x) as a relative maximum, a relative minimum, or neither:
   a. $f'(x) = x^3 (2x- 3)^2 (x +1)^5 (x+7)$
   b. $f^1(x) = \frac{x(x-2)^2}{x^4+1}$
   c. $f'(x) = \sqrt[3]{x}(3-x)(x+1)^2$
7. The first derivative of a certain function is $f'(x) = x(x - 1)^2$.
   a. On what intervals is f increasing? Decreasing?
   b. On what intervals is the graph of f concave up? Concave down?
   c. Find the x coordinates of the relative extrema and inflection points of f.
   d. Sketch a possible graph of f(x).
8. Suppose that q = 500 - 2p units of a certain commodity are demanded when p dollars per unit are charged, for $0\leq p \leq 250$.
   a. Determine where the demand is elastic, inelastic, and of unit elasticity with respect to price.

b. Use the results of part (a) to determine the intervals of increase and decrease of the revenue function and the price at which revenue is maximized.

c. Find the total revenue function explicitly and use its first derivative to determine its intervals of increase and decrease and the price at which revenue is maximized.

d. Graph the demand and revenue functions

Further Readings

- Sydsaeter, Knut and Hammond, Peter J. (1995) Mathematics for Economic Analysis, Pearson Education, Inc, NOIDA, India.
- Hoy, Michael, Livernois, John, McCenna, Chris, Rees, Ray, and Stengos, Thanasis (2001) Mathematics for Economics the MIT Press, Cambridge, Massachusetts, USA. Indian Reprint (Prentice-Hall of India Pvt Ltd, New Delhi India
- Chiang, Alpha C. and Wainwright, Kevin (2005) Fundamental Methods of Mathematical Economics 4th edition, McGraw-Hill International edition, Boston, USA.
- Renshaw, Geoff (2009) Maths for Economics, Oxford University Press, New York, USA. Indian Reprint available

Web Links

- https://www.cl.cam.ac.uk/teaching/2003/Probability/prob07.pdf
- https://www.eco.uc3m.es/~rimartin/Teaching/AMATH/NOTES2EN.pdf
- https://people.math.aau.dk/~matarne/11-imat/notes2011a.pdf
- https://nitkkr.ac.in/docs/14-Difference%20Equations%20and%20their%20Applications.pdf

Unit 05: Linear Programming I

CONTENTS

Objectives

After studying this unit, you will be able to,

- enable you to grasp the basic idea of linear programming principles;
- formulate a management problem as a linear programming problem in suitable cases,
- identify the characteristics of a linear programming problem,
- make a graphical analysis of the problem,
- formulate the dual linear programming problem and analyse the dual variables.

Introduction

Linear programming (LP) is a technique used for deriving optimum use of limited resources. Specifically, it deals with maximizing a linear function of variables subject to linear constraints. Applications range from economic planning and environmental management to the diet problem. In principle, any linear programming problem (often called an LP problem) can be solved numerically, provided that a solution exists. This is because the simplex method introduced by G. B. Dantzig in 1947 provides a very efficient numerical algorithm that finds the solution in a finite number of steps.

5.1 Linear Programming: Basic Concept

Linear programming deals with the optimization of the total effectiveness expressed as a linear function of decision variables, known as the objective function, subject to a set of linear equalities/inequalities known as constraints. Decision variables are the variables in terms of which the problem is defined.

Optimization problems consist of two basic features, viz.,

1. An objective function that you want to minimize or maximize.

2. The objective function describes the behavior of the measure of effectiveness and captures the relationship between that measure and those variables that cause it to change. System variables can be categorized as decision variables and parameters. A decision variable is directly controlled by the decision-maker. Thus, constraints are the relations between decision variables and the parameters.

Thus, any LP problem will have an objective function and a set of constraints. In most cases, constraints come from the nature of problem you work with. For example, if the desirable objective (maximisation or minimisation) in view is of keeping some constraints (i.e., the difficulties, restrictions).

Check the following conditions while formulating a LP problem:

- The objective function must be linear. That is, check if all variables have power of 1 and they are added or subtracted (not divided or multiplied).
- The objective must be either maximisation or minimisation of a linear function. The objective must represent the goal of the decision-maker.
- The constraints must also be linear. Moreover, the constraint must be of the following forms ($\leq, \geq$ or =, that is, the LP-constraints are always closed) giving the LP in standard format:
- Max $c^T x$ subject to (s.t.) $Ax = b$, for $x \geq 0$

As the above quotation indicates, the simplex method has made linear programming a mathematical technique of immense practical importance. It is reported that when Mobil Oil Company's multimillion-dollar computer system was installed in 1958, it paid off this huge investment in two weeks by doing linear programming. That said, the simplex method will not be discussed in this book. After all, faced with a nontrivial LP problem, it is natural to use one of the great numbers of available LP computer programs to find the solution. In any case, it is probably more important for economists to understand the basic theory of LP than the details of the simplex method.

Indeed, the importance of LP extends even beyond its practical applications. In particular, the duality theory of linear programming is a basis for understanding key theoretical properties of more complicated optimization problems with an even larger range of interesting economic applications.

5.2 Feasible and Optimal Solutions

A solution value for decision variables, where all of the constraints are satisfied, is called a feasible solution. Most solution algorithms proceed by first finding a feasible solution, then seeking to improve upon it, and finally changing the decision variables to move from one feasible solution to another feasible solution. This process is repeated until the objective function has reached its maximum or minimum. This result is called an optimal solution. The basic goal of the optimization process is to find values of the variables that minimize or maximize the objective function while satisfying the constraints. This result is called an optimal solution.

5.3 Formulation: Structure & Variables of Linear Programming

The mathematical formulation of linear programming problem (LPP) is described in the following steps:

1. Identify the decision variables of the problem.

2. Express the objective function, which is to be optimised, i.e., maximised or minimised, as a linear function of the decision variables.

3. Identify the limited available resources, i.e., the constraints and express them as linear inequalities or equalities in terms of decision variables.

4. Since negative values of the decision variables do not have any valid physical interpretation, introduce non-negative restrictions. Let us take an example to illustrate these steps.

Example 1: A small scale industry manufactures two products P and Q which are processed in a machine shop and assembly shop. Product P requires 2 hours of work in a machine shop and 4 hours of work in the assembly shop to manufacture while product Q requires 3 hours of work in machine shop and 2 hours of work in assembly shop. In one day, the industry cannot use more than 16 hours of machine shop and 22 hours of assembly shop. It earns a profit of `3 per unit of product P and `4 per unit of product Q. Give the mathematical formulation of the problem so as to maximize profit.

Solution: Let x and y be the number of units of product P and Q, which are to be produced. Here, x and y are the decision variables. Suppose Z is the profit function.

Since one unit of product P and one unit of product Q gives the profit of 3 and 4, respectively, the objective function is

Maximise: $Z = 3x + 4y$

The requirement and availability in hours of each of the shops for manufacturing the products are tabulated as follows:

	Machine Shop	Assembly	Shop Profit
Product P	2 hours	4 hours	3 per unit
Product Q	3 hours	2 hours	4 per unit
Available hours per day	16 hours	22 hours	

Total hours of machine shop required for both types of product = $2x + 3y$ Total hours of assembly shop required for both types of product = $4x + 2y$ Hence, the constraints as per the limited available resources are:

$$2x + 3y \le 16 \text{ and}$$

$$4x + 2y \le 22$$

Since the number of units produced for both P and Q cannot be negative, the non-negative restrictions are:

$$x \ge 0, y \ge 0$$

Thus, the mathematical formulation of the given problem is

Maximise $Z = 3x + 4y$

subject to the constraints

$2x + 3y \le 16$

$4x + 2y \le 22$ and

non-negative restrictions

$x \ge 0, v \ge 0$

Example 2: A manufacturer has two types of machines. Her production requires that she must have at least three A type of machines and one B type of machine. The cost of production in type A machine is Rs.1000 while that in type B is Rs.1200 for. The floor areas taken up by each of these machines are 4m2 and 5m2 respectively. The total cost of production must not exceed Rs. 15000/- and the available floor space is $40m^2$.

Solution: Let x, y be the number of type A and type B machines. Then the equalities will be l00Ox + 1200y ≤15000 for the constraint as regards money.

Such a formulation implies $5x + 6y \le 75$ .... (1)

For space constraint we have $4x + 5y \leq 40$ .... (2)

Again for A type of machine $x \geq 3$ . . . . (3)

And for B type of machine $x \geq 1$ . . . . (4)

Now, suppose the weekly profit from the output as Rs.120 for each type A and Rs.100 for each type B machine. Our problem is to find the combination of machine use giving maximum profit. If the total profit $p = 120x + 100y$, then the problems is to find x and y which will maximize the objective function. p = 120x + l00y subject to constraints (I), (2), (3) and (4) above. Here x and y are decision variables.

5.4 Graphic Solution

The graphical method is used to solve linear programming problems having two decision variables. For solving LPPs involving more than two decision variables, we use another method called the Simplex method.

The graphical method of solving a linear programming problem comprises the following steps:

Step 1 : Plot the constraints on the graph paper and find the feasible region.

Step 2: Find the coordinates of that verifies (comer points) of the feasible region.

Step 3: Find the values of the objective function of these comer points.

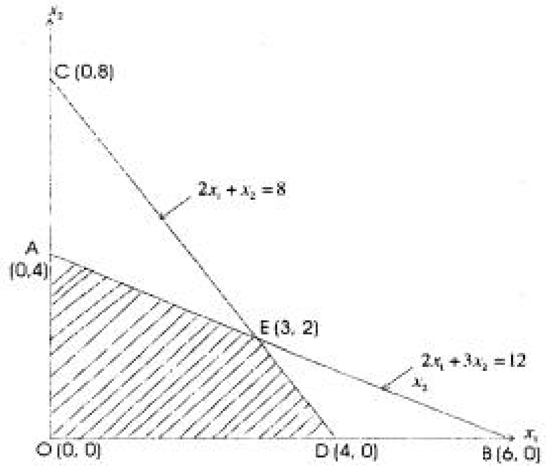
Step 4: Select the comer point at which the value of the objective function is maximum (or minimum as in minimisation problem). This comer point will be the required solution of the given L.P.P.

Example 3: Maximise $z = 6x_1 + 7x_2$

Subject to $2x_1 + 3x_2 \leq 12$

$2x_1 + x_2 \leq 8$

$(x_1, x_2) \geq 0$



The shaded region OAED is the set of points that are feasible under all the constraints it is called the feasible region. The task now is to find which combination of x_1 and x_2 in OAED for which the value of z is maximised.

The task is added by a theorem: If there exists an optimum solution' to the problem, it will be found among the combinations of xl and x2 values represented by the vertices (or the extreme comers) of the feasible solution polygon.

Comer Point	Coordination (x_1, x_2)	Value of z =6x1 + 7x2
O	(0,0)	6(0)+7(0)=0

D	(4,0)	6(4)+7(0)=24
E	(3,2)	6.(3)+7(2)=32
A	(0,4)	6.(0)+7(4)=28

A firm is producing two goods, A and B. It has two factories that jointly produce the two goods in the following quantities (per hour):

The value of z is maximum at E whose coordinates are (3, 2). Hence, the required solution is xl = 3, x2 = 2 and the corresponding maximum value of z =32.

The same process is followed for a minimization process but only then we have to consider the minimum value of z.

Example 3: A firm is producing two goods, A and B. It has two factories that jointly produce the two goods in the following quantities (per hour):

	Factory 1	Factory 2
Good A	10	20
Good B	25	25

The firm receives an order for 300 units of A and 500 units of B. The costs of operating the two factories are 10 000 and 8 000 per hour. Formulate the linear programming problem of minimizing the total cost of meeting this order.

Solution:

Let u_1 and u_2 be the number of hours that the two factories operate to produce the order. Then

$10u_1 + 20u_2$ units of good A are produced, and $25u_1 + 25u_2$ units of good B. Because 300 units of A and 500 units of B are required, u_1 and u_2 must satisfy

$10u_1 + 20u_2 \geq 300$

$25u_1 + 25u_2 \geq 500$

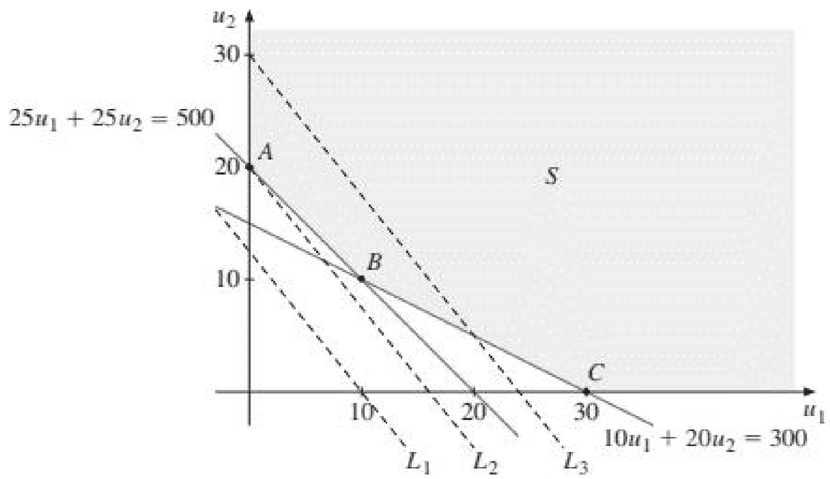
In addition, of course, $u_1 \geq 0$ and $u_2 \geq 0$. The total costs of operating the two factories for u_1 and u_2 hours, respectively, are $10\,000\ u_1 + 8\,000\ u_2$. The problem is, therefore,

min $10\,000\ u_1 + 8\,000\ u_2$ subject to

$10u_1 + 20u_2 \geq 300$

$25u_1 + 25u_2 \geq 500$ $\qquad u_1 \geq 0,\ u_2 \geq 0$

The feasible set S is shown in Fig. Because the inequalities in (i) are of the ≥ type and all the coefficients of u1 and u2 are positive, the feasible set lies to the north-east. Figure includes three of the level curves 10 000u1 + 8 000u2 = c, marked L1, L2, and L3. These three correspond to the values 100 000, 160 000, and 240 000 of the cost level c. As c increases, the level curve moves farther and farther to the north-east.



The solution to the minimization problem is clearly the level curve that touches the feasible set S at point A with coordinates (0, 20). Hence, the optimal solution is to operate factory 2 for 20 hours and not to use factory 1 at all, with minimum cost 160000.

The graphical method of solving linear programming problems works well when there are only two decision variables. One can extend the method to the case with three decision variables. Then the feasible set is a convex polyhedron in 3-space, and the level surfaces of the objective function are planes in 3-space. However, it is not easy to visualize the solution in such cases. For more than three decision variables, no graphical method is available.

5.5 Simplex Method

The simplex method is based on the property that the optimum solution of a L.P. problem if exists can always be found in one of the basic feasible solutions.

Extra Variables

The solution to an LP problem is a set of optimal values for each of the variables. However, the output that comes with the solution to a LP problem usually contains much more information than just this. In addition to the optimal values of the variables, the output will typically include reduced cost values, slack or surplus values, and dual prices (also known as shadow prices).

Reduced Cost

Associated with each variable is a reduced cost value. However, the reduced cost value is only non-zero when the optimal value of a variable is zero. A somewhat intuitive way to think about the reduced cost variable is to think of it as indicating how much the cost of the activity represented by the variable must be reduced before any of that activity will be done. More precisely, ... the reduced cost value indicates how much the objective function coefficient on the corresponding variable must be improved before the value of the variable will be positive in the optimal solution. In the case of a minimization problem, "improved" means "reduced."

So, in the case of a cost-minimization problem, where the objective function coefficients represent the per-unit cost of the activities represented by the variables, the "reduced cost" coefficients indicate how much each cost coefficient would have to be reduced before the activity represented by the corresponding variable would be cost-effective. In the case of a maximization problem, "improved" means "increased." In this case, where, for example, the objective function coefficient might represent the net profit per unit of the activity, the reduced cost value indicates how much the profitability of the activity would have to increase in order for the activity to occur in the optimal solution. The units of the reduced cost values are the same as the units of the corresponding objective function coefficients.

If the optimal value of a variable is positive (not zero), then the reduced cost is always zero. If the optimal value of a variable is zero and the reduced cost corresponding to the variable is also zero, then there is at least one other corner that is also in the optimal solution. The value of this variable will be positive at one of the other optimal corners.

Slack or Surplus

A slack or surplus value is reported for each of the constraints. The term "slack" applies to less than or equal constraints, and the term "surplus" applies to greater than or equal constraints. If a constraint is binding, then the corresponding slack or surplus value will equal zero. When a less-than-or-equal constraint is not binding, then there is some unutilized, or slack, resource. The slack value is the amount of a resource, as represented by a less-than-or equal constraint, that is not being used. When a greater-than-or-equal constraint is not binding, then the surplus is the extra amount over the constraint that is being produced or utilized. The units of the slack or surplus values are the same as the units of the corresponding constraints.

Dual Prices (a.k.a. Shadow Prices)

The dual prices are some of the most interesting values in the solution to a linear program. A dual price is reported for each constraint.

The dual price is only positive when a constraint is binding. The dual price gives the improvement in the objective function if the constraint is relaxed by one unit.

In the case of a less-than-or-equal constraint, such as a resource constraint, the dual price gives the value of having one more unit of the resource represented by that constraint. In the case of a greater-than-or-equal constraint, such as a minimum production level constraint, the dual price gives the cost of meeting the last unit of the minimum production target.

The units of the dual prices are the units of the objective function divided by the units of the constraint. Knowing the units of the dual prices can be useful when you are trying to interpret what the dual prices mean.

Standard form of LPP

Standard form of LPP must have following three characteristics:

1. Objective function should be of maximization type
2. All the constraints should of equality type
3. All the decision variables should be nonnegative

The procedure to transform a general form of a LPP to its standard form is discussed below. Let us consider the following example.

Example 4: Maximise $z = x + y$,

subject to $x + y \leq 5$

$x+3y \leq 12,$

$x\geq 0, y\geq 0$

Step 1: Convert the inequalities to equalities by addition of non-negative slack variables. Let s1 and s2 be the slack variables, which convert the inequalities into equations. Then

$x + y + s_1 = 5$ and $3x + 2y + s_2 = 10$

where $s_1 \geq 0$ and $s_2 \geq 0$.

The problem can then be written as follows:

maximise $z = 5x + 6y + 0s_1 + 0s_2$

subject to $x + y + s_1 + 0s_2 = 5$ and

$3x + 2y + s_2 + 0s_1 = 12$

$x\geq 0,\ y\geq 0,\ s_1\geq 0,\ s_2 \geq 0$

Step 2: Put the problem in a simplex tableau

			5	6	0	0	
C	Basic variable	Values of the basic variable	x	y	s_1	s_2	Ratio
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	s_1	5	1	1	1	0	5/1 =5
0	s_2	12	2	3	0	1	12/3=4
	z_j	0	0	0	0	0	
	$C_j- z_j$	-	5	6	0	0	

The 1" column denoted by Cj, is known as the objective column, which represents the coefficients of the objective function of the basic variables listed in Column (2). The 2"d column shows the basic in the solution. The 3rd column represents the values of basic variables listed in Column (2).

Columns (4), (5), (6) and (7) represent the 4 variables x, y, s_1, and s_2 respectively. The row shows the coefficients of the respective variables in the objective function. This row is known as the objective row. The Column (8) known as the ratio column. The elements in zj, row are obtained by multiplying the elements of that column by the compounding elements of the objective column (Cj column) and then added up. Similarly, other elements of the zj row all zero since all the elements in the Cj column are zero. The row Cj- zj is known as the net evaluation row or index row.

Step 3: (a) Calculate the net evaluation of (Cj- zj). To get an element in the net Values row under any column multiply the entries in that column by the corresponding entities in the objective column (Cj) and all them up. Next, subtract this sum from the element in the objective row listed at the top of the table

(b) After examining that the net evaluation row of all the elements are zero or negative, the optimum solution is reached. But if any positive element is present it indicates that a better program can be formulated.

(c) Revise the program.

1) Find the pivot column: The column under which falls the largest positive element of the net evolution row is pivot column.

2) Find the pivot row and the pivot number: Divide the elements of the Application constant column by corresponding non-negative elements of the pivot column to form replacement ratio. The row in which the replacement ratio is the smallest is the pivot row. The number lying at the intersection of the pivot row and the pivot column is the pivot number.

3) Transform the pivot row: Divide all the elements of the pivot row (starting from the constant column) by the pivot number. The resulting numbers will form the corresponding row of the next table.

4) The non-pivot rows are transformed by using the rule

New Number= Old Number - $-\frac{Corresponding\ number\ in\ the\ pivot\ row*corresponding\ number\ in\ the\ pivot\ column}{Pivot\ number}$

5) With the results of (3) and (4) form a new table representing a new basic solution. In the new table the variable of the pivot row of previous table will be replaced by the variable of the column of the previous table.

Then steps 3 and 4 are repeated until an optimal column is reached.

Example 5: Maximise $z = 3x_1 + 7x_2 + 6x_3$

Subject to $2x_1 + 2x_2 + 2x_3 \leq 8$

$x_1 + x_2 \leq 3$

$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0$

Next step is to convert it into equalities by two slack variables s_l and s_2

Maximise $z = 3x_1 + 7x_2 + 6x_3 + 0s_1 + 0s_2$

Subject to $2x_1 + 2x_2 + 2x_3 + s_1 + 0s_2 = 8$

$x_1 + x_2 + 0s_1 + s_2 = 3$

$(x_1, x_2, x_3, s_1, s_2) \geq 0$

				3	7	6	0	0	
	C_j	Basic variable	Values of the basic variable	x_1	x_2	x_3	s_1	S_2	Ratio
Table 1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	0	s_1	8	2	2	2	1	0	8/2 =4
	0	s_2	3	1	1	0	0	1	3/1=3
		z_j	0	0	0	0	0	0	
		C_j-z_j	-	3	7	6	0	0	
Table 2	0	s_1	2	0	0	2	1	-2	2/2=1
	7	x_2	3	1	1	0	0	1	
		z_j.	21	7	7	0	0	7	
		C_j-z_j	-	-4	0	0	0	-7	
Table 3	6	x_3	1	0	0	1	½	-1	
	7	x_2	3	1	1	0	0	1	
		zj.	27	7	7	6	3	1	
		C_j-z_j	-	-4	0	0	-3	-1	

In the lst program we include only the slack variables sl = 8, s2 =3, xl =0, x2 = 0 and x3 = 0. In this solution, we find that the value of the objective function is zero. Further, from the Cj-zj row it is found that there are positive elements. Hence, a better program can be formulated. The highest positive element in the net evaluation row is 7, which lies in the x2 column. Then the x2 column is the pivot column and in the next program x2 has to be included as one of the basic variables. Now dividing the elements of the constant column by the nonnegative elements of the pivot column, we get the replacement ratios of column (9). The lowest ratio appears to be in the second column (s2) column, which is now the pivot column. Hence in the next program s2 will be replaced by x2. The pivot number is 1, which lies at the intersection of the pivot row and pivot column. The 2nd program is formulated in Table 2, which is given below Table 1. In Table 2 the basic variables are sl and s2. The 2nd row (x2 row) of Table 2 has been obtained by dividing the elements of the 2nd row of Table 1 by the pivot number, 1. The elements of the first now (sl row) of Table 2 have been obtained by following the rule for the transformation of the non-pivot row. The calculations are shown below:

$$8 - \frac{3 \times 2}{1} = 2 \qquad 2 - \frac{1 \times 2}{1} = 0$$

$$2 - \frac{1 \times 2}{1} = 0 \qquad 2 - \frac{0 \times 2}{1} = 0$$

$$1 - \frac{0 \times 2}{1} = 1 \qquad 0 - \frac{1 \times 2}{1} = -2$$

Thus the 2nd table represents a basic solution where $x_1 = 0$, $x_2 = 3$, $x_3 = 0$, $s_l = 2$, $s_2 = 0$ and where the value of the objective function is 21. Inspecting the net evaluation row (Cj-zj) it is found that one element remains positive. The highest positive number in the net evaluation row lies in the x_3 column, which is now the pivot column. The variable x_3 will have to be introduced as a basic variable in the next programme. The smallest replacement ratio in Table 2 falls in the s_l row. Therefore, this row is the pivot row and in the new program s_l will be replace by x_3. The s_l row in Table 2 is the pivot row and 2 is the pivot number.

The 3rd program is shown in Table 3, which is placed immediately below Table 2. By diving all the elements of the sl row of Table 2 by 2, the pivot number we get the corresponding elements of the first row (x_3 row) in Table 3. The elements of the second row (x_2 row) of Table 3 have been obtained from the corresponding elements of the 2" row (x_2 row) of Table 2 by using the rule for transformation of the non-pivot row(s). These calculations are shown below:

$$3-\frac{2\times 0}{2}=3 \qquad 1-\frac{0\times 0}{2}=1$$

$$1-\frac{0\times 0}{2}=1 \qquad 0-\frac{2\times 0}{2}=0$$

$$0-\frac{1\times 0}{2}=0 \qquad 1-\frac{(-2)\times 0}{2}=1$$

In the third the basic solution is given by xl = 0, x2 = 3, x3 =1, s1 = 0, s2 = 0. In the net evaluation row of the table it is found that all the elements are either zero or negative. This means that the optimum program has been attained and there is no scope for further improvement. Hence, the required optimum solution is x1= 0, x2 = 3 and x3 = 1. The corresponding value of z = 27.

Summary

- A problem wherein the objective is to allot the limited available resources to the jobs in such a way as to optimise the overall effectiveness, i.e., minimise the total cost or maximise the total profit, is called mathematical programming. Mathematical programming wherein constraints are expressed as linear equalities/inequalities is called linear programming. Linear programming deals with the optimisation of the total effectiveness expressed as a linear function of decision variables, known as the objective function, subject to a set of linear equalities/inequalities known as constraints.
- Steps involved in the mathematical formulation of a linear programming problem (LPP) are: i) Identification of the decision variables of the problem; ii) expressing the objective function as a linear function of the decision variables; iii) identifying the limited available resources to write the constraints as linear inequalities or equalities in terms of decision variables; iv) introducing the nonnegative restrictions.
- Graphical method is used to solve linear programming problems having two decision variables. The graphical method of solving linear programming programs comprises the following steps: i) The graphs are plotted for the equations corresponding to the given inequalities for constraints as well as restrictions; ii) The region corresponding to each inequality is shaded; iii) After shading the regions for each inequality, the most common shaded portion, i.e., the region obtained on superimposing all the shaded regions is determined. This is the region where all the given inequalities, including non-negative restrictions are satisfied. This common region is known as the feasible region or the solution set or the polygonal convex set; iv) Each of the corner points (vertices) of the polygon is then determined; v) The objective function at each corner point is evaluated.
- How to formulate the dual of an LPP, weak and strong duality theorems which relate the primal and dual basic feasible solutions. The Dual simplex method to solve LPP for which a basic optimal, but infeasible, solution is known.

Keywords

- Basic Feasible Solutions: These solutions are basic as well as feasible.

- Basic Solution: Any set of values of the variables in which the number of non-zero valued variables is equal to the number of constraints is called a Basic Solution.
- Constraints: The linear inequalities or the side condition.
- Dual Problem: Associated with every linear programming there is a linear programming problem. Which is called its dual problem.
- Feasible Solution: A set of values of decision variables, which satisfies the set of constraints and the non-negativity restrictions.
- Linear Programming: Linear programming is the analysis of problems in which linear function of a number of variables is to be maximized (or minimized) when those variables are subject to a number of restraints in the form of linear inequalities.
- Maximinior Minimax Principle: If a player lists the worst possible outcomes of all his potential strategies then he will choose that strategy to be the most suitable for him which corresponds to the best of these worst outcomes.
- Mixed Strategy: Decision making rule in which a player decides in advance, to choose his courses of action with some definite probability distribution.
- Primal: The original L.P.P. is called the primal problem
- Objective Function: The function to be maximized or minimized.

Self Assessment

1. A constraint in an LP model restricts

A. value of the objective function
B. value of the decision variable
C. use of the available resources
D. all of the above

2. In graphical method of linear programming problem if the iso-cost line coincide with a side of region of basic feasible solutions we get

A. Unique optimum solution
B. unbounded optimum solution
C. no feasible solution
D. Infinite number of optimum solutions

3. The linear function of the variables which is to be maximize or minimize is called

A. Constraints
B. Objective function
C. Decision variable
D. None of the above

4. The first step in formulating a linear programming problem is

A. Identify any upper or lower bound on the decision variables
B. State the constraints as linear combinations of the decision variables
C. Understand the problem
D. Identify the decision variables

5. A basic solution is called non-degenerate, if

A. All the basic variables are zero
B. None of the basic variables is zero
C. At least one of the basic variables is zero
D. None of these

6. The graph of $x \leq 2$ and $y \geq 2$ will be situated in the
A. First and second quadrant
B. Second and third quadrant
C. First and third quadrant
D. Third and fourth quadrant

7. Which of the terms is not used in a linear programming problem
A. Slack variables
B. Objective function
C. Concave region
D. Feasible solution

8. In. L.P.P----
A. objective function is linear
B. constraints are linear
C. Both objective function and constraints are linear
D. None of the above

9. Constraints means----
A. limitations are expressed in mathematical equalities (or inequalities)
B. Assumption
C. goal is to be achieved
D. None of the above.

10. Dual of the dual is the
A. Primal
B. Dual
C. Either primal or dual
D. None of these

11. If the 'i' constraint of a primal (maximization) is equality, then the dual (minimization) variable 'yi' is:
A. $\leq$
B. $\geq$
C. Unrestricted in sign
D. None of the above

12. In linear programming, dual price represents
A. Mean and maximum price
B. Unit worth of a resource
C. Minimum and mean price

D. Minimum and maximum price

13. In primal-dual solutions, the dual problem solution can be obtained by solving other problems classified as
A. Unrestricted problem
B. Original problem
C. Double problem
D. Restricted problem

14. When using a graphical solution procedure, the region bounded by the set of constraints is called the
A. solution.
B. feasible region.
C. infeasible region.
D. maximum profit region.

15. The graphic method of LP uses
A. objective equations.
B. constraint equations.
C. linear equations.
D. all of the above.

Answers for Self Assessment

1.	A	2.	D	3.	B	4.	D	5.	B
6.	B	7.	C	8.	C	9.	A	10.	A
11.	C	12.	B	13.	B	14.	B	15.	D

Review Questions

1. What is the difference between formulating and solving a linear programming problem?
2. What are the basic steps in formulating a linear program?
3. What is the feasible region of an LP?
4. Solve the following LPP by two phase simplex method.
   Minimise: $Z= 15/2x_1-3x_2$
   subject to $3x_1 - X_2 - X_3 \geq 3$
   $x_l - x_2 + x_3 \geq 2$
   $(x_1, x_2, x_3) \geq (0,0,0)$
5. Maximise $-2x_1 + 3x_2 – x_4 + x_5$subject to $x_l +x_2 -x_3+x_4 \leq 5$
   $3x_1-5x_2+X_4 \leq 7$
   $4x_1 +2x_2 -x_3 -6x_5 \leq 10$
   $4x_1-2x_2 +x_3 +6x_5 \leq -10$
   $(x_l,x_2,x_3, x_4, x_5)\geq(0,0,0,0,0)$
6. Minimize $Z = 10x_1 + 4x_2$subject to the constraints

$4x_1+x_2 \geq 80$

$2x_1+ x_2 \geq 60$

$x_1 \geq 0,\ x_2 \geq 0$

Further Readings

- S. Hillier / G. J. Liebennan (2001), Introduction to Operations Research, McGraw-Hill ,7th Edition.
- M. Glicksman (2001). An Introduction to Linear Programming and the Theory of Games, Dover Publ., Mineola, NY.
- Essential Mathematics for Economists- Nutt Sedester, peter Hawmond, Prentice Hall Publication

Web Links

- https://www.brainkart.com/article/Solution-of-LPP-by-graphical-method_37041/
- http://www.math.wsu.edu/students/odykhovychnyi/M201-04/Ch06_1-2_Simplex_Method.pdf

Unit 06: Linear Programming II

CONTENTS

Objectives

- Understand Shadow Price
- describe Identification of unique and multiple optimal solutions
- describe dual simplex method

Introduction

In linear programming, shadow prices refer to the change in the optimal value of the objective function that results from a one-unit change in the right-hand side of a constraint. In other words, the shadow price of a constraint represents the marginal value of an additional unit of the constrained resource.

For example, consider a linear programming problem where the objective is to maximize profit subject to a set of constraints, including a constraint that limits the availability of a particular resource (e.g., labor or raw materials). The shadow price associated with this constraint represents the increase in profit that would result from an additional unit of the resource becoming available.

6.1 Shadow Prices

Shadow prices can be useful for decision-making, as they provide information on the value of additional resources and can help in determining which constraints are most critical to achieving the objective. In addition, shadow prices can be used to perform sensitivity analysis to evaluate the impact of changes in the problem data on the optimal solution.

Shadow prices in linear programming have both advantages and disadvantages, as described below:

Advantages:

1. Provide valuable information: Shadow prices provide valuable information on the marginal value of additional resources and help in making informed decisions regarding resource allocation.
2. Allow sensitivity analysis: Shadow prices can be used to perform sensitivity analysis and assess the impact of changes in the problem data on the optimal solution. This helps in identifying critical constraints and making adjustments to the problem formulation as needed.

3. Support decision-making: Shadow prices enable decision-makers to compare the costs and benefits of alternative courses of action and identify the most cost-effective solutions.

Disadvantages:

1. Limited applicability: Shadow prices are only applicable to linear programming problems and cannot be used in other types of optimization problems.
2. Assumptions must hold: Shadow prices assume that the problem parameters, such as the objective function coefficients and constraint RHS values, remain constant and do not change as additional resources are allocated. If these assumptions do not hold, shadow prices may not accurately reflect the true value of additional resources.
3. Lack of precision: Shadow prices are estimates and may not provide an accurate representation of the true marginal value of additional resources. The accuracy of shadow prices depends on the quality of the underlying data and the assumptions made in the problem formulation.

In summary, shadow prices can be a useful tool in linear programming to inform decision-making and support sensitivity analysis. However, their use should be accompanied by an awareness of their limitations and the need for careful interpretation and validation of the results.

To illustrate the concept of shadow prices in linear programming with a numerical example, let's consider the following linear programming problem:

Example 1: Maximize Z = 5x1 + 6x2

Subject to:

2x1 + 3x2 <= 12

4x1 + 2x2 <= 16

where x1 and x2 are decision variables.

Solution: We can solve this problem using a linear programming solver, such as Excel Solver or any other software. The optimal solution is x1=2, x2=2 and Z=22.

Now, let's consider the first constraint: 2x1 + 3x2 <= 12. The shadow price associated with this constraint is the change in the optimal value of Z resulting from a one-unit increase in the right-hand side (RHS) of the constraint. To determine the shadow price, we can solve the following modified linear programming problem:

Maximize Z = 5x1 + 6x2

Subject to:

2x1 + 3x2 <= 13

where the RHS of the first constraint is increased by one unit. Solving this problem, we obtain a new optimal solution: x1=3, x2=1.3333, and Z=21.3333.

The shadow price associated with the first constraint is the difference in Z values between the original problem and the modified problem, divided by the change in the RHS of the constraint. In this case, the shadow price is:

(21.3333 - 22) / (13 - 12) = -0.6667

This means that for each additional unit of resource available beyond the current level of 12, the value of the objective function Z would decrease by $0.6667.

Similarly, we can compute the shadow price associated with the second constraint (4x1 + 2x2 <= 16) by solving a modified problem with the RHS of the second constraint increased by one unit. The shadow price associated with the second constraint is:

(21.5 - 22) / (17 - 16) = -0.5

This means that for each additional unit of the second resource available beyond the current level of 16, the value of the objective function Z would decrease by $0.5.

In summary, shadow prices in linear programming are a useful tool for assessing the marginal value of additional resources and evaluating the sensitivity of the optimal solution to changes in the problem data.

Example 2: Suppose a company produces two types of products, A and B. The company has limited resources of labor and machine time, which it uses to manufacture the products. The objective is to maximize the profit from the production and sale of the products subject to the constraints on labor and machine time. The problem can be formulated as a linear programming problem as follows:

Maximize Z = 5A + 8B

Subject to:

2A + 3B <= 240 (Labor constraint)

A + 2B <= 200 (Machine time constraint)

Solution: Suppose that the optimal solution to this problem is A=60, B=80, and the maximum profit is $940. The shadow prices of the labor and machine time constraints can be determined by finding the dual of the problem. The dual problem is:

Minimize W = 240X + 200Y

Subject to:

2X + Y >= 5

3X + 2Y >= 8

where X and Y are the dual variables associated with the labor and machine time constraints, respectively. The dual problem represents the minimum cost of the resources necessary to produce one unit of the two products.

The optimal solution to the dual problem is X=2.8, Y=1.2, and the minimum cost is $12.8. The shadow prices of the labor and machine time constraints are the coefficients of the dual variables in the optimal solution, which are 2.8 and 1.2, respectively.

Interpreting the shadow prices, we can say that the marginal value of one additional unit of labor is $2.8 and the marginal value of one additional unit of machine time is $1.2. For example, if the company were able to obtain additional labor at a cost of less than $2.8 per unit, it would be profitable to do so, as the additional labor would generate more profit than the cost. Similarly, if the company could obtain additional machine time at a cost of less than $1.2 per unit, it would be profitable to do so.

In summary, shadow prices in linear programming provide valuable information on the marginal value of additional resources and can be used to inform decision-making and sensitivity analysis. The shadow prices are found by determining the dual of the problem and interpreting the coefficients of the dual variables.

Identification of Unique and Multiple Optimal Solutions

Shadow prices are a useful tool in linear programming for assessing the sensitivity of the optimal solution to changes in the problem's constraints. The shadow price of a resource represents the amount by which the objective function value would increase if one additional unit of the resource were made available. In other words, it measures the marginal value of the resource to the objective function.

In linear programming, the objective is to maximize or minimize a linear function subject to linear constraints. A solution is optimal if it satisfies all the constraints and maximizes or minimizes the objective function. A linear program can have one, no, or many optimal solutions.

a. If a linear program has a unique optimal solution, it means there is only one feasible solution that maximizes or minimizes the objective function. This is often the case when the constraints are binding, meaning that the solution satisfies all constraints with equality.
b. If a linear program has multiple optimal solutions, it means there are several feasible solutions that maximize or minimize the objective function. This occurs when there are multiple feasible solutions that satisfy all the constraints and provide the same optimal objective function value. This can happen when some of the constraints are non-binding, meaning that the solution satisfies the constraint with inequality.

To identify whether a linear program has a unique or multiple optimal solutions, one can use sensitivity analysis to examine how the optimal objective function value changes as the problem's constraints are adjusted. This can be done by examining the shadow prices of the resources, as mentioned earlier, or by performing a range analysis to determine the range of values for the coefficients in the constraints that would not change the optimal solution.

When there is a unique optimal solution, the shadow prices can be obtained by finding the dual of the linear programming problem and examining the coefficients of the dual variables in the optimal solution. The shadow price associated with a particular constraint represents the marginal value of the resources associated with that constraint.

In contrast, when there are multiple optimal solutions, the shadow prices can vary depending on which optimal solution is considered. In such cases, the range of the shadow prices can be determined by performing sensitivity analysis.

For example, consider a linear programming problem with two optimal solutions:

Example 3: Maximize Z = 5x1 + 3x2

Subject to:

x1 + x2 <= 4

2x1 + 3x2 <= 9

x1, x2 >= 0

Solution: The two optimal solutions are (x1 = 2, x2 = 2) and (x1 = 3, x2 = 1). The shadow prices associated with the constraints can be determined by finding the dual of the problem as follows:

Minimize W = 4y1 + 9y2

Subjectto:

y1 + 2y2 >= 5

y1 + 3y2 >= 3

The optimal solution to the dual problem is (y1 = 1.5, y2 = 0.5), and the corresponding shadow prices are 1.5 and 0.5 for the first and second constraints, respectively.

However, the shadow prices associated with the constraints depend on which optimal solution is considered. For the first optimal solution (x1 = 2, x2 = 2), the shadow prices are 1.5 and 0.5 for the first and second constraints, respectively. For the second optimal solution (x1 = 3, x2 = 1), the shadow prices are 3 and 0, respectively.

In summary, when there are multiple optimal solutions, the shadow prices associated with the constraints can vary depending on which optimal solution is considered. Sensitivity analysis can be used to determine the range of the shadow prices and help in making informed decisions regarding resource allocation.

6.2 Duality of Linear Programming

In linear programming, duality refers to a relationship between two optimization problems, known as the primal and the dual problems, that allows us to obtain valuable insights and solutions to both problems. The primal problem seeks to maximize a linear objective function subject to linear constraints, while the dual problem seeks to minimize a linear objective function subject to linear constraints. The key idea of duality is that the dual problem provides a lower bound on the optimal value of the primal problem, and the primal problem provides an upper bound on the optimal value of the dual problem. This is known as the weak duality theorem.

The strong duality theorem states that if the primal problem has an optimal solution and some technical conditions are satisfied, then the optimal value of the primal problem is equal to the optimal value of the dual problem. This means that solving either problem can give us the optimal solution to the other problem as well. Moreover, the dual problem allows us to derive additional insights and information about the primal problem, such as shadow prices, which indicate the marginal value of each constraint in the primal problem, and sensitivity analysis, which helps us understand how changes in the constraints affect the optimal solution.

In summary, duality is a powerful tool in linear programming that provides a deeper understanding of optimization problems and allows us to obtain optimal solutions and useful information about both the primal and the dual problems.

Every L.P. problem is intimately related another called its "dual". For purposes of identification, the original problem is called the primal problem. The relationship between the primal problem and its dual can be summarized as follows:

1) The dual has as many variables as there are constraints in the original problem.

2) The dual has as many constrains as there are variables in the original problem.

3) The dual of a maximization problem is a minimization problem and vice versa.

4) The coefficients of the objective function of the original problem appear as the constant terms of the constraints of the dual and the constant terms of the original constraints are the coefficients of the objective function of the dual.

5) The coefficients of a single variable in the original constraints become the coefficients of a single constraint in the dual. Stated visually, each column of coefficients in the constraints of the original problem becomes a row of coefficients in the dual.

6) The sense of the inequalities, the dual is the reverse of the inequalities in the original problem, except that the inequalities restricting the variables to be non-negative have the same in the primal and the dual.

Example 6: Suppose the primal problem is

Maximise $z=c_1x_1+c_2x_2$

subject to $a_{11}\, x_1. + a_{12}\, x_2 \leq b_1$

$a_{21}\, x_1. + a_{22}\, x_2 \leq b_2$

$a_{31}\, x_1. + a_{32}\, x_2 \leq b_3$

$x_1 \geq 0,\ x_2 \geq 0$

We have to form the dual of this problem by applying the 6 rules described above. Since the original problem has 3 constraints the dual problem will have 3 variables. Let yl, y2 and y3 be the dual variables. Again, since the original problem has two variables, the dual problem will have 2 constraints. Since the original problem is a maximization problem, the constants bl, b2 and b3 will appear as the coefficients in the objective function of the dual and the constants cl and c2 will appear as the constant terms in the right hand side of the constraints of the dual.

Further, the first column of coefficients $\begin{bmatrix} a11 \\ a21 \\ a31 \end{bmatrix}$ in the constraints will be the first row of coefficients in the constraints of the dual. Similarly, the second column of coefficients $\begin{bmatrix} a12 \\ a22 \\ a32 \end{bmatrix}$ will be the second rowof coefficients in the constraints of the dual. Again, since the constraints in the original problem all the "less than equal to" type, the constraints in the dual problem will be "greater than equal to" type.

The dual problem will then be as follows:

maximize $w = b_1y_1 + b_2y_2 + b_3y_3$

subject to $a_{11}y_1 + a_{21}y_2 \geq c_1$

$a_{21}y_1 + a_{22}y_2 \geq c_2$

$y_1 \geq 0,\ y_2 \geq 0,\ y_3 \geq 0$

Example3: Original problem:

Maximise $z=4x+6y$

subject to $1/2\, x + y \leq 4$

$2x+y \leq 8$

$4X-2y \leq 2$

x≥0,y ≥ 0.

Dual problem:

Minimize w = 4u + 8v + 2w.

Subject to 1/2 u + 2v + 4w ≥4

u+v-2w≥ 6

u≥0, v≥ 0,w≥ 0.

Last Table for the solution of the original problem

			4	6	0	0	0
C_j	Basic variable	Values of the basic variable	x	y	s_1	s_2	s_3
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
6	y	3	0	1	4/5	0	-1/10
0	s_2	1	0	0	-8/5	1	-3/10
4	x	2	1	0	2/5	0	1/5
	z_j	26	4	6	32/5	0	1/5
	C_j-z_j	-	0	0	-32/5	0	-1/5

Last Table for the solution of the dual problem

			4	8	2	0	0	M	M
C_j	Basic variable	Values of the basic variable	u	v	w	s_1	s_2	A_1	A_2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2	w	1/5	0	3/10	1	-1/5	1/10	1/5	- 1/10
4	u	32/5	1	16/10	0	-2/5	-4/5	2/5	4/5
	z_j	26	4	7	2	-2	-3	2	3
	C_j-z_j	-	0	1	0	2	3	M-2	M-3

Here also it is seen that the value of the objective function is the same in both the problems. In the original problem the elements in the net evaluation row under columns sl, s2 and s3 are -32/5, 0 and -1/5. Hence, the solution of the three dual variables will be u = 32/5, v = 0, and w =-115.

The knowledge of the dual is important for two main reasons.

1) The dual variables have economic interpretations. The values of the dual variables may be useful in taking managerial decisions.

2) The solution of a L.P. problem may be easier through the dual than through the primal problem.

Example 5 : Maximize: 3x + 4y

Subjectto:x + y ≤ 4

$x + 2y \leq 7$

$x, y \geq 0$

Solution: To find the dual problem, we first introduce dual variables u and v, one for each constraint, and formulate the Lagrangian:

$L(x, y, u, v) = 3x + 4y - u(x + y - 4) - v(x + 2y - 7)$

We then minimize the Lagrangian with respect to x and y, and obtain the dual function g(u, v):

g(u, v) = min L(x, y, u, v) = -4u - 7v + 4u^2 + 5uv + 2v^2

Finally, the dual problem is to maximize g(u, v) subject to $u, v \geq 0$.

Maximise: -4u - 7v + 4u^2 + 5uv + 2v^2

Subject to:

$u, v \geq 0$

We can now solve the dual problem using various methods. For example, we can use the simplex method to solve it by converting it to standard form:

Maximise: -4u - 7v + 4u^2 + 5uv + 2v^2

Subject to:

4u + v + s1 = 0

u + 2v + s2 = 0

$u, v, s1, s2 \geq 0$

We start with the initial feasible solution u = v = s1 = 0, s2 = 0, and find the most negative coefficient in the objective function, which is -7v. We then choose v as the entering variable and find the leaving variable, which is s2. We pivot on (2, 1) to obtain the new tableau:

u	v	s1	s2	RHS
4	1	0	0	0
1	2	0	1	0
4	0	0	-7	0

The optimal solution is u = 4, v = 0, and the optimal value of the dual problem is -28.

We can now use strong duality to conclude that the optimal value of the primal problem is also 28, and the optimal solution is x = 2, y = 2, with a dual solution of u = 2, v = 0.

Note that the dual problem provides us with additional information, such as the shadow prices or dual variables u and v, which represent the marginal value of each constraint in the primal problem. In this case, u = 2 indicates that the first constraint has a marginal value of 2, while v = 0 indicates that the second constraint is binding and has a marginal value of 0. This means that a small increase in the right-hand side of the first constraint would increase the optimal objective value by 2, while a small increase in the right-hand side of the second constraint would not affect the optimal objective value.

Summary

The simplex method is an appropriate method for solving a linear programming problem involving more than two decision variables. For "less-than-or-equal-to" type of constraints, slack variables are introduced to convert the inequalities into equations. A particular type of solution known as a basic feasible solution is important for simplex computation. Every basic feasible solution is an extreme point of the convex set of feasible solutions and vice- versa. We can always find a basic feasible solution with the help of the slack variables. The objective function is maximized or minimized at one of the basic feasible solutions. Starting with the initial basic feasible solution obtained from the slack variables, the simplex method improves the value of the objective function step by step by bringing in a new basic variable and making one of the present basic variables as non-basic. The selection of the new basic variable and the omission of a current basic variable are performed by following certain rules so that the revised basic feasible solution improves the value of the objective

function. The interactive procedure stops when it is no longer possible to get a better value of the objective function than the present one. The existing basic feasible solution is the optimum solution which maximizes or minimizes the objective function as the case may be When one or more of the constraints are "greater-than-or-equal-to" type, then the surplus variables are introduced to convert inequalities into equations. Surplus variables, however cannot be used to obtain an initial basic feasible solution. If some of the constraints are "greater-than-or equal-to" type or some are equations, then the artificial variables are used to initiate the simplex computation. WO methods, namely Big-M method and Wo Phase Method are available to solve linear programming problems in these cases. The simplex method can also identify unbounded solutions and infeasible problems.

Every linear programming problem has an accompanying linear programming problem, known as dual problem. The variables of the dual problem are known as dual variables. The dual variables have an economic interpretation which can be used by management for planning its resources. The solution of the dual problem can be obtained from the simplex computation of the original problem. The solution has a number of important properties which can also be helpful for computational purposes.

Keywords

1. Slack Variable- A slack variable corresponding to a "less-than-or equal-to" type of constraint is a non-negative variable introduced to convert the constraint into an equation.
2. Basic Solution: A basic solution of a system of m equations and n variables ($m < n$) is a solution where at least ($n - m$) variables are zero.
3. Surplus Variable: A surplus variable corresponding to a "greater-than-or-equal to" type of constraint is a non-negative variable introduced to convert the constraint into an equation.
4. Dual Problem: The dual problem corresponding to a linear programming problem is another linear programming problem formulated from the parameters of the original problem.
5. Primal Problem: The primal problem is the original linear programming problem.
6. Dual Variables: The dual variables are the variables of the dual linear programming problem.
7. Shadow Price: The shadow price of a resource is the change in the optimum value of the objective function per unit increase of the resource.

Self Assessment

1. The dual of the dual is ____________.

A. dual-primal
B. primal-dual
C. dual
D. primal

2. If the given Linear Programming Problem is in its standard form then primal-dual pair is _______

A. symmetric
B. unsymmetric
C. slack
D. square

3. If the primal problem has n constraints and m variables then the number of constraints in the dual problem is __________.

A. mn
B. m+n

C. m-n

D. m/n

4. In an LPP functions to be maximized or minimized are called ___________.

A. constraints

B. objective function

C. basic solution

D. feasible solution

5. A feasible solution of an LPP that optimizes then the objective function is called _________

A. basic feasible solution

B. optimum solution

C. feasible solution

D. solution

6. The region of feasible solution in LPP graphical method is called ____.

A. Infeasible region

B. Unbounded region

C. Infinite region

D. Feasible region

7. When it is not possible to find solution in LPP, it is called as case of__________.

A. Unknown solution

B. Unbounded solution

C. Infeasible solution

D. Improper solution

8. When the feasible region is such that the value of objective function can extend to infinity, it is called a case of __________.

A. Infeasible solution

B. Alternate optimal

C. Unbounded solution

D. Unique solution

9. When the constraints are a mix of 'less than' and 'greater than' it is a problem having __________.

A. Multiple constraints

B. Infiniteconstraints

C. Infeasible constraints

D. Mixed constraints

10. In case of an '________' constraint, the feasible region is a straight line.

A. less than or equal to

B. greater than or equal to

C. mixed

D. equal to

11. In linear programming, unbounded solution means ______.

A. Infeasible solution

B. Degenerate solution

C. Infinite solutions

D. Unique solution

12. The value of one extra unit of resource is called __________ in simplex.

A. unit price

B. extra price

C. retail price

D. shadow price

13. In simplex, a maximization problem is optimal when all Delta J, i.e. Cj – Zj values are __________.

A. Either zero or positive

B. Either zero or negative

C. Only positive

D. Only negative

14. The variable added to the LHS of a less than or equal to constraint to convert it into equality is called __________.

A. surplus variable

B. artificial variable

C. slack variable

D. additional variable

15. A resource which is completely utilized is called __________ in simplex.

A. null resource

B. scarce resource

C. zero resource

D. abundant resource

Answers forSelfAssessment

1.	D	2.	B	3.	A	4.	B	5.	B
6.	D	7.	C	8.	C	9.	D	10.	D
11.	C	12.	D	13.	B	14.	C	15.	B

Review Questions:

1. Obtain the dual LP problem of the following primal LP problem: Minimize Z = x1 + 2x2 subject to the constraints (i) 2x1 + 4x2 ≤ 160, (ii) x1 – x2 = 30, (iii) x1 ≥ 10 and x1, x2 ≥ 0
2. Obtain the dual LP problem of the following primal LP problem: MinimizeZx = x1 – 3x2 – 2x3 subject to the constraints (i) 3x1 – x2 + 2x3 ≤ 7, (ii) 2x1 – 4x2 ≥ 12, (iii) – 4x1 + 3x2 + 8x3 = 10 and x1, x2≥ 0; x3 unrestricted in sign.

3. XYZ Company has three departments – Assembly, Painting and Packing. The company can make three types of almirahs. An almirah of type I requires one hour of assembly, 40 minutes of painting and 20 minutes of packing time, respectively. Similarly, an almirah of type II needs 80 minutes, 20 minutes and one hour, respectively. The almirah of type III requires 40 minutes each of assembly, painting and packing time. The total time available at assembly, painting and packing departments is 600 hours, 400 hours and 800 hours, respectively. Determine the number of each type of almirahs that should be produced in order to maximize the profit. The unit profit for types I, II and III is Rs 40, 80 and 60, respectively. Suppose that the manager of this XYZ company is thinking of renting the production capacities of the three departments to another almirah manufacturer – ABC company. ABC company is interested in minimizing the rental charges. On the other hand, the XYZ company would like to know the worth of production hours to them, in each of the departments, in order to determine the rental rates. (a) Formulate this problem as an LP problem and solve it to determine the number of each type of almirahs that should be produced by the XYZ company in order to maximize its profit. (b) Formulate the dual of the primal LP problem and interpret your results.
4. Define the dual of a linear programming problem. State the functional properties of duality.
5. Explain the primal-dual relationship.
6. How can the concept of duality be useful in managerial decision-making?
7. State and prove the relationship between the feasible solutions of an LP problem and its dual.
8. Prove that the necessary and sufficient condition for any LP problem and its dual, in order to have optimal solutions is that both have feasible solutions.
9. Three food products are available at costs of Rs. 10, Rs. 36 and Rs. 24 per unit, respectively. They contain 1,000, 4,000 and 2,000 calories per unit, respectively and 200, 900 and 500 protein units per unit, respectively. It is required to find the minimum-cost diet containing at least 20,000 calories and 3,000 units of protein. Formulate and solve the given problem as an LP problem. Write the dual and use it to check the optimal solution of the given problem.
10. A company produces three products: P, Q and R from three raw materials A, B and C. One unit of product P requires 2 units of A and 3 units of B. A unit of product Q requires 2 units of B and 5 units of C and one unit of product R requires 3 units of A, 2 units of B and 4 units of C. The company has 8 units of material A, 10 units of material B and 15 units of material C available to it. Profits per unit of products P, Q and R are Rs. 3, Rs. 5 and Rs. 4, respectively. (a) Formulate this problem as an LP problem. (b) How many units of each product should be produced to maximize profit? (c) Write the dual of this problem.

Further Readings

- "Microeconomic Theory" by Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green - This is a comprehensive textbook on microeconomic theory that covers the concept of shadow prices in detail.
- "Resource Allocation in Project Management" by Michael E. Taylor - This book provides a practical guide to resource allocation in project management, including the use of shadow prices.
- "Applied Cost-Benefit Analysis" by Robert J. Brent - This book covers the application of cost-benefit analysis to public policy and includes a discussion of shadow prices.

Unit 07: Statistical Data

CONTENTS

Objectives

- describe the factors affecting choice of data
- explain the problem5 of collecting primary data
- narrate the different methods of collecting primary data
- state the sources of secondary data
- explain the precautions to be taken while using secondary data

Introduction

The views commonly held about statistics are numerous, but often incomplete. It has different meanings to different people depending largely on its use. For example, (i) for a cricket fan, statistics refers to numerical information or data relating to the runs scored by a cricketer; (ii) for an environmentalist, statistics refers to information on the quantity of pollution released into the atmosphere by all types of vehicles in different cities; (iii) for the census department, statistics consists of information about the birth rate per thousand and the sex ratio in different states; (iv) for a share broker, statistics is the information on changes in share prices over a period of time; and so on.

The average person perceives statistics as a column of figures, various types of graphs, tables and charts showing the increase and/or decrease in per capita income, wholesale price index, industrial production, exports, imports, crime rate and so on. The sources of such statistics for a common man are newspapers, magazines/journals, reports/bulletins, radio, and television. In all such cases the relevant data are collected, numbers manipulated and information presented with the help of figures, charts, diagrams, and pictograms; probabilities are quoted, conclusions reached, and discussions held. Efforts to understand and find a solution (with certain degree of precision) to problems pertaining to social, political, economic, and cultural activities, seem to be unending. All such efforts are guided by the use of methods drawn from the field of statistics.

Now-a-days the term 'statistics' has become a household word, although different people comprehend it in different senses. Today an educated person has to be a person of statistics, broadly understanding its meaning and applying it to his/her life in different ways. In daily life it means general calculation of items, in railway statistics means the number of trains operating, number of passenger's freight etc. and so on.

Thus, statistics is used by people to take decision about the problems on the basis of different type of quantitative and qualitative information available to them. However, in behavioral sciences, the word 'statistics' means something different from the common concern of it. Prime function of statistic is to draw statistical inference about population on the basis of available quantitative information. Overall, statistical methods deal with reduction of data to convenient descriptive terms and drawing some inferences from them. In order to improve our understanding of the world around us, it is necessary to

i) measure what is being said,

ii) express it numerically, and

7.1 Utilize Quantitative Information or Expression to Draw Conclusions and Suggest Policy Measures

Meaning of Statistics

The term 'statistics' finds it origin in an Italian term 'Statista' that is a person who deals with State related affairs and activities. During early period, these words were used for political state of the region. The word 'Statista' was used to keep the records of census or data related to wealth of a state. Gradually, its meaning and usage extended and there onwards its nature also changed.

It was initially called 'state arithmetic' in which the information about the nation, for instance, tax related information and war plans, were tabulated (Aron, Aron and Coups, 2009). Thus, statistics was earlier known for its application to government related activities and data, like census. However, today it is increasingly used in various fields like economics, psychology, education, management and so on. Further, Statistics can be described as a branch or sub-field of mathematics that mainly deals with the organization as well as analysis and interpretation of a group of numbers (Aron, Aron and Coups, 2009). In simple terms, statistics can be described as "the science of classifying, organizing and analyzing data".

The word statistics is used to convey different meanings in singular and plural sense. Therefore, it can be defined in two different ways.

Statistics in Singular Sense

In singular sense, 'Statistics' refers to what is called statistical methods. It deals with the collection of data, their classification, analysis and interpretations of statistical data. Therefore, it is described as a branch of science which deals with classification, tabulation and analysis of numerical facts and make decision as well. Every statistical inquiry should pass through these stages.

Statistics in Plural Sense

Statistics used in plural sense means that quantitative information is available called 'data'. For example, information on population or demographic features, enrolment of students in Psychology programmes of IGNOU, and the like. According to Websters "Statistics are the classified facts representing the conditions of the people in a State specifically those facts which can be stated in number or in tables of number or classified arrangement".

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." Thus Secrist's definition highlights following features of statistics:

i. ***Statistics are aggregate of facts***: Single or unrelated items are not considered as statistics.

ii. ***Statistics are affected by multiplicity of causes***: In statistics the collected information are greatly influenced by a number of factors and forces working together.

iii. ***Statistics are numerical facts***: Only numerical data constitute statistics.

iv. ***Statistics are enumerated or estimated with a reasonable standard of accuracy***: While enumerating or estimating data, a reasonable degree of accuracy must be achieved.

v. ***Statistics are collected in a systematic manner***: Data should be collected by proper planning by utilising tool/s developed by trained personnel.

vi. ***Statistics are collected for a predetermined purpose***: It is necessary to define the objective of enquiry, before collecting the statistics. The objective of enquiry must be specific and well defined.

vii. ***Statistics should be comparable***: Introduction to Statistics Only comparable data will have some meaning. For statistical analysis, the data should be comparable with respect to time, place group, etc.

7.2 Definition of Statistics

"Statistics may be called the science of counting". A. L. Bowley

"Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry".

Selligman

"Statistics as the collection, presentation, analysis, and interpretation of numerical data"

Croxton and Cowden

7.3 Nature of Statistics

If Science is knowledge, Art is action or the actual application of science. While Science teaches us to know, Art teaches us to do. Further, Art has the following characteristics:

- It is a group of actions which solve a problem
- It does not describe the facts but examines the merits and demerits and suggests ways to achieve the objective

Based on these characteristics, we can define statistics as an art of applying the science of scientific methods. As an art, statistics offer a better understanding and solution to problems in real life as it offers quantitative information.

While there are several statistical methods, the successful application of the methods is dependent on the statistician's degree of skill and experience.

According to Tippet,

"Statistic is both a science and an art. It is a science in that its methods are basically systematic and have general application and art in that their successful application depends, to a considerable degree, on the skill and special experience of the statistician, and on his knowledge of the field of application."

According to Mohanty and Misra (2016),

- Statistics can be termed as a science in which the facts related to social events are observed, recorded and computed.
- Organization of data, its classification and analysis are the processes involved in statistics.

- Various events and phenomenon can be described, explained and compared with the help of statistics.

A scientific enquiry can be systematically interpreted and predicted with the help of statistics. And in this regard statistics can also help in decision making.

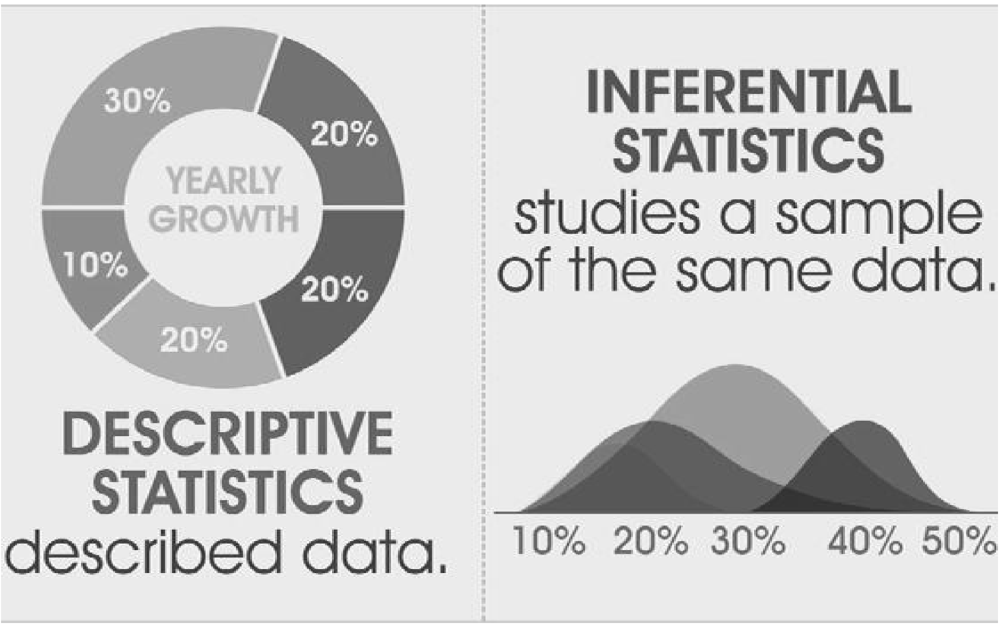
7.4 Types of Statistics

Though various bases have been adopted to classify statistics, following are the two major ways of classifying statistics:

i. on the basis of function and

ii. on the basis of distribution.

1. **On the Basis of Functions**



Descriptive Statistics: As it obviously implies, descriptive statistics deals with collecting, summarizing and simplifying data, which are otherwise quite unwieldy and voluminous. It seeks to achieve the intended in a manner that meaningful conclusions can be readily drawn from the data. Descriptive statistics may thus be seen as comprising methods that highlight the latent characteristics present in any sent of data. In the process, it facilitates an easy understanding of the data and makes them amenable to further analysis and interpretation.

Once the data have been collected, these are assembled, organized and presented in the form of appropriate tables to make them conveniently readable. Wherever needed, charts and graphs are drawn for better grasp of the data. No manner of data presentation is useful unless the data are properly classified in keeping with the objectives of the investigation.

These includes measures of central tendency, dispersion, skewness and kurtosis, which constitute the essence of descriptive statistics.

Inferential statistics:

Organizing and summarizing data is only one step in the process of analyzing the data. In any scientific investigation either the entire population or a sample is considered for the study.

In most of the scientific investigations a sample, a small portion of the population under investigation, is used for the study. On the basis of the information contained in the sample we try to draw conclusions about the population. This process is known as statistical inference.

Statistical inference is widely applicable in behavioral sciences, especially in psychology. For example, before the Lok Sabha or Vidhan Sabha election process starts or just before the declaration of election results print media and electronic media conduct exit poll to predict the election result.

In this process all voters are not included in the survey, only a portion of voters i.e. sample is included to infer about the population. This is called inferential statistics.

Inferential statistics deals with drawing of conclusions about large group of individuals (population) on the basis of observation of a few participants from among them or about the events which are yet to occur on the basis of past events. It provides tools to compute the probabilities of future behavior of the subjects.

Inferential statistics is the mathematics and logic of how this generalization from sample to population can be made. There are two types of inferential procedures:

- Estimation,
- Hypothesis testing.

Estimation:

In estimation, inference is made about the population characteristics on the basis of what is discovered about the sample. There may be sampling variations because of chance fluctuations, variations in sampling techniques, and other sampling errors. Estimation about population characteristics may be influenced by such factors. Therefore, in estimation the important point is that to what extent our estimate is close to the true value.

Characteristics of Good Estimator:

A good statistical estimator should have the following characteristics,

i. Unbiased: An unbiased estimator is one in which, if we were to obtain an infinite number of random samples of a certain size, the mean of the statistic would be equal to the parameter. The sample mean, (x) is an unbiased estimate of population mean (μ) because if we look at possible random samples of size N from a population, then mean of the sample would be equal to μ.

ii. Consistent: A consistent estimator is one that as the sample size increased, the probability that estimate has a value close to the parameter also increased. Because it is a consistent estimator, a sample mean based on 20 scores has a greater probability of being closer to (μ) than does a sample mean based upon only 5 scores.

iii. Accuracy: The sample mean is an unbiased and consistent estimator of population mean (μ). But we should not overlook the fact that an estimate is just a rough or approximate calculation. It is unlikely in any estimate that (x) will be exactly equal to population mean (μ). Whether or not x is a good estimate of (μ) depends upon the representative ness of sample, the sample size, and the variability of scores in the population.

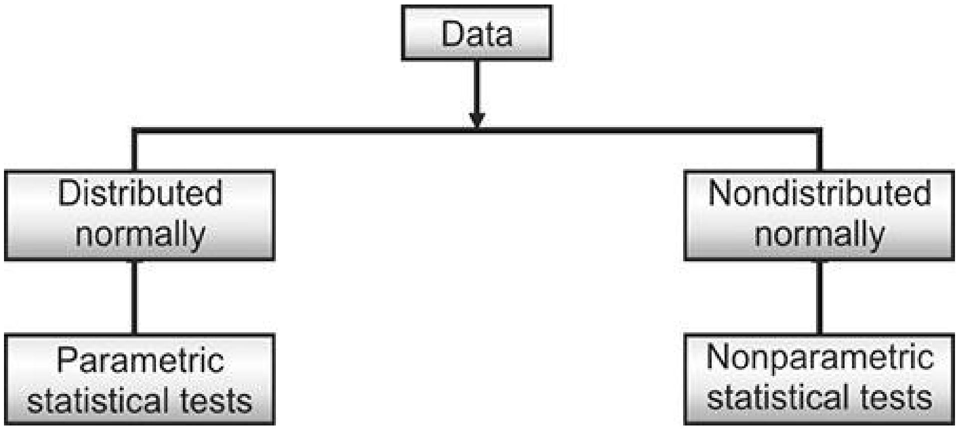
Hypothesis Testing:

Inferential statistics is closely tied to the logic of hypothesis testing. In hypothesis testing we have a particular value in mind. We hypothesize that this value characterizes the population of observations. The question is whether that hypothesis is reasonable in the light of the evidence from the sample. In estimation no particular population value need to be stated. Rather, the question is: What is the population value? For example, Hypothesis testing is one of the important areas of statistical analyses. Sometimes hypothesis testing is referred to as statistical decision-making process. In day-to-day situations we are required to take decisions about the population on the basis of sample information. For example, on the basis of sample data, we may have to decide whether a new method of teaching is better than the existing one, whether new medicine is more effective in curing the disease than the previously available medicine, and so forth.

Null Hypothesis:The probability of chance occurrence of the observed results is examined by the null hypothesis (H0). Null hypothesis is a statement of no differences.

Alternative hypothesis: It's a hypothesis that a random cause may influence the observed data or sample. It is represented by Ha or H1.

2. **On the Basis of Distribution of Data**



Parametric and nonparametric statistics are the two classifications on the basis of distribution of data. Both are also concerned to population or sample. By population we mean the total number of items in a sphere. In general, it has infinite number therein but in statistics there is a finite number of a population, like the number of students in a college. According to Kerlinger (1968) "the term population and universe mean all the members of any well-defined class of people, events or objects." In a broad sense, statistical population may have three kinds of properties – (a) containing finite number of items and knowable, (b) having finite number of articles but unknowable, and (c) keeping infinite number of articles.

Parametric Statistics:

It is defined to have an assumption of normal distribution for its population under study. "Parametric statistics refers to those statistical techniques that have been developed on the assumption that the data are of a certain type. In particular, the measure should be an interval scale and the scores should be drawn from a normal distribution".

There are certain basic assumptions of parametric statistics. The very first characteristic of parametric statistics is that it moves after confirming its population's property of normal distribution. The normal distribution of a population shows its symmetrical spread over the continuum of –3 SD to +3 SD and keeping unimodal shape as its mean, median, and mode coincide. If the samples are from various populations, then it is assumed to have same variance ratio among them. The samples are independent in their selection. The chances of occurrence of any event or item out of the total population are equal and any item can be selected in the sample. This reflects the randomized nature of sample which also happens to be a good tool to avoid any experimenter bias.

In view of the above assumptions, parametric statistics seem to be more reliable and authentic as compared to the nonparametric statistics. These statistics are more powerful to establish the statistical significance of effects and differences among variables. It is more appropriate and reliable to use parametric statistics in case of large samples as it consists of more accuracy of results. The data to be analyzed under parametric statistics are usually from interval scale.

However, along with many advantages, some disadvantages have also been noted for the parametric statistics. It is bound to follow the rigid assumption of normal distribution and further it narrows the scope of its usage. In case of small sample, normal distribution cannot be attained and thus parametric statistics cannot be used. Further, computation in parametric statistics is lengthy and complex because of large samples and numerical calculations. T-test, F-test, r-test, are some of the major parametric statistics used for data analysis.

Non-Parametric Tests

Non-parametric statistics covers techniques that do not rely on data belonging to any particular distribution. These include (i) distribution free methods (ii) nonstructural models. Distribution free means its interpretation does not depend on any parametrized distributions. It deals with statistics based on the ranks of observations and not necessarily on scores obtained by interval or ratio scales.

Non-parametric statistics is defined to be a function on a sample that has no dependency on a parameter. The interpretation does not depend on the population fitting any parametrized distributions. Statistics based on the ranks of observations are one example of such statistics and these play a central role in many non-parametric approaches.

Non-parametric techniques do not assume that the structure of a model is fixed. Typically, the model grows in size to accommodate the complexity of the data. In these techniques, individual variables are typically assumed to belong to parametric distributions, and assumptions about the types of connections among variables are also made.

Non-parametric methods are widely used for studying populations that are based on rank order (such as movie reviews receiving one to four stars). The use of non-parametric methods may be necessary when data have a ranking but no clear numerical interpretation. For instance, when we try to assess preferences of the individuals, (e.g., I prefer Red more than White color etc.), we use non parametric methods. Also, when our data is based on measurement by ordinal scale, we use non-parametric statistics.

As non-parametric methods make fewer assumptions, their applicability is much wider than those of parametric methods. Another justification for the use of non-parametric methods is its simplicity. In certain cases, even when the use of parametric methods is justified, non-parametric methods may be easier to use. Due to the simplicity and greater robustness, non-parametric methods are seen by some statisticians as leaving less room for improper use and misunderstanding.

Assumptions of Non-parametric Statistics:

- An independence of observations, except with paired data
- Continuity of variable under study

Characteristics of non-parametric techniques:

- Fewer assumptions regarding the population distribution
- Sample sizes are often less stringent
- Measurement level may be nominal or ordinal
- Independence of randomly selected observations, except when paired
- Primary focus is on the rank ordering or frequencies of data
- Hypotheses are posed regarding ranks, medians, or frequencies of data.

7.5 Scope of Statistics

The scope of applications of statistics has assumed unprecedented dimensions these days. Statistical methods are applicable in all diversified fields such as economics, trade, industry, commerce, agriculture, bio-sciences, physical sciences, educations, astronomy, insurance, accountancy and auditing, sociology, psychology, meteorology, and so on.

According to the statistician, Bowley, 'A knowledge of statistics is like a knowledge of foreign language or of algebra, it may prove of use at any time under any circumstances'.

1. Statistics and the State

A state in the modern setup collects the largest amount of statistics for various purposes. It collects data relating to prices, production, consumption, income and expenditure, investments, and profits. Popular statistical methods such as time-series analysis, index numbers, forecasting, and demand analysis are extensively practiced in formulating economic policies. Governments also collect data on population dynamics in order to initiate and implement various welfare policies and programmes. In addition to statistical bureaus in all ministries and government departments in the Central and state governments, other important agencies in the field are the Central Statistical Organization (CSO), National Sample Survey Organization (NSSO), and the Registrar General of India (RGI).

2. Statistics in Economics:

Statistical methods are extensively used in all branches of economics. For example:

(i) Time-series analysis is used for studying the behaviour of prices, production and consumption of commodities, money in circulation, and bank deposits and clearings.

(ii) Index numbers are useful in economic planning as they indicate the changes over a specified period of time in (a) prices of commodities, (b) imports and exports, (c) industrial/agricultural production, (d) cost of living, and the like.

(iii) Demand analysis is used to study the relationship between the price of a commodity and its output (supply).

(iv)Forecasting techniques are used for curve fitting by the principle of least squares and exponential smoothing to predict inflation rate, unemployment rate, or manufacturing capacity utilization.

3. Statistics in Business Management

According to Wallis and Roberts, 'Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainty.' Ya-Lin-Chou gave a modified definition over this, saying that 'Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risks.' These definitions reflect the applications of statistics in the development of general principles for dealing with uncertainty. Statistical reports provide a summary of business activities which improves capability of making more effective decisions regarding future activities. Discussed below are certain activities of a typical organization where statistics plays an important role in their efficient execution.

1. Marketing: Before a product is launched, the market research team of an organization, through a pilot survey, makes use of various techniques of statistics to analyze data on population, purchasing power, habits of the consumers, competitors, pricing, and a hoard of other aspects. Such studies reveal the possible market potential for the product. Analysis of sales volume in relation to the purchasing power and concentration of population is helpful in establishing sales territories, routing of salesman, and advertising strategies to improve sales.
2. Production: Statistical methods are used to carry out R&D programs for improvement in the quality of the existing products and setting quality control standards for new ones. Decisions about the quantity and time of either self-manufacturing or buying from outside are based on statistically analyzed data.
3. Finance: A statistical study through correlation analysis of profit and dividend helps to predict and decide probable dividends for future years. Statistics applied to analysis of data on assets and liabilities and income and expenditure, help to ascertain the financial results of various operations. Financial forecasts, break-even analysis, investment decisions under uncertainty— all involve the application of relevant statistical methods for analysis.
4. Personnel: In the process of manpower planning, a personnel department makes statistical studies of wage rates, incentive plans, cost of living, labor turnover rates, employment trends, accident rates, performance appraisal, and training and development programmers'. Employer-employee relationships are studied by statistically analyzing various factors— wages, grievances handling, welfare, delegation of authority, education and housing facilities, and training and development

1. Statistics in Physical Science:

Currently there is an increasing use of statistical methods in physical sciences such as astronomy, engineering, geology, meteorology, and certain branches of physics. Statistical methods such as sampling, estimation, and design of experiments are very effective in the analysis of quantitative expressions in all fields of most physical sciences.

2. Statistics in Social Sciences:

In social sciences where both types (quantitative and qualitative) of information are used, statistics helps the researchers to alter the information in a comprehensive way to explain and predict the patterns of behavior/ trend. Where the characteristics of the population being studied are normally distributed, the best and statistically important decision about variables being investigated is possible by using parametric statistics or nonparametric statistics to explain the pattern of activities.

3. Statistics in Medical Sciences:

The knowledge of statistical techniques in all natural sciences—zoology, botany, meteorology, and medicine—is of great importance. For example, for proper diagnosis of a disease, the doctor needs and relies heavily on factual data relating to pulse rate, body temperature, blood pressure, heart beats, and body weight.

An important application of statistics lies in using the test of significance for testing the efficacy of a particular drug or injection meant to cure a specific disease. Comparative studies for effectiveness of a particular drug/injection manufactured by different companies can also be made by using statistical techniques such as the t-test and F-test.

To study plant life, a botanist has to rely on data about the effect of temperature, type of environment, and rainfall, and so on.

4. Theoretical researches:

Theories evolve on the basis of facts obtained from the field. Statistical analyses establish the significance of those facts for a particular paradigm or phenomena. Researchers are engaged in using the statistical measures to decide on the facts and data whether a particular theory can be maintained or challenged. The significance between the facts and factors helps them to explore the connectivity among them.

5. Industries:

Statistics is a basic tool to handle daily matters not only in big organizations but also in small industries. It is required, at each level, to keep data with care and look at them in different perspectives to mitigate the expenditure and enable each employee to have his/ her share in the benefit. Psychologists/ personnel officers dealing with selection and training in industries also use statistical tools to differentiate among employees.

7.6 Limitations of Statistics

1. Statistics Does Not Study Qualitative Phenomena:

Since statistics deals with numerical data, it cannot be applied in studying those problems which can be stated and expressed quantitatively. For example, a statement like 'Export volume of India has increased considerably during the last few years' cannot be analyzed statistically. Also, qualitative characteristics such as honesty, poverty, welfare, beauty, or health, cannot directly be measured quantitatively. However, these subjective concepts can be related in an indirect manner to numerical data after assigning particular scores or quantitative standards. For example, attributes of intelligence in a class of students can be studied on the basis of their Intelligence Quotients (IQ) which is considered as a quantitative measure of the intelligence.

2. Statistics Does Not Study Individual

According to Horace Secrist 'By statistics we mean aggregate of facts effected to a marked extent by multiplicity of factors . . . and placed in relation to each other.' This statement implies that a single or isolated figure cannot be considered as statistics, unless it is part of the aggregate of facts relating to any particular field of enquiry. For example, price of a single commodity or increase or decrease in the share price of a particular company does not constitute statistics. However, the aggregate of figures representing prices, production, sales volume, and profits over a period of time or for different places do constitute statistics.

3. Statistics Can be Misused

Statistics are liable to be misused. For proper use of statistics, one should have enough skill and experience to draw accurate and sensible conclusions. Further, valid results cannot be drawn from the use of statistics unless one has a proper understanding of the subject to which it is applied.

The greatest danger of statistics lies in its use by those who do not possess sufficient experience and ability to analyze and interpret statistical data and draw sensible conclusions. Bowley was right when he said that 'statistics only furnishes a tool though imperfect which is dangerous in the hands of those who do not know its use and deficiencies.' For example, the conclusion that smoking causes lung cancer, since 90 per cent of people who smoke die before the age of 70 years, is statistically invalid because here nothing has been mentioned about the percentage of people who do not smoke and die before reaching the age of 70 years. According to W. I. King, 'statistics are like clay of which you can make a God or a Devil as you please.' He also remarked, 'science of statistics is the useful servant but only of great value to those who understand its proper use.

7.7 Distrust and Misuse of Statistics

Despite its importance and usefulness, the science of statistics is looked upon with suspicion. Quite often is discredited, by people who donot know its real purpose and limitations. We often hear statements such as:

"There are three types of lies: Lies, damned lies and statistics. "Statistics can prove anything". By distrust of statistics, we mean lack of confidence in statistical data, statistical methods and the conclusions drawn. You may ask, why distrust in statistics? Some of the important reasons for distrust in statistics are as follows:

1. Arguments based upon data are convincing. But data can be manipulated according to wishes of an individual.

2. Even if correct figures are used, they may be incomplete and presented in such a manner that the reader is misled. Suppose, it has been found that the number of traffic accidents is lower in foggy weather than on clear weather days. It may be concluded that it is safer to drive in fog. the conclusion drawn is wrong. To arrive at a valid conclusion, we must take into account the difference between the rush of traffic under the two weather conditions.

3. Statistical data does not bear on their face the label of their quality. Sometimes even unintentionally inaccurate or incomplete data is used leading to faulty conclusions.

4. The statistical tools have their own limitations. The investigator must use them with precaution. But sometimes these tools or methods are handled by those who have little or no knowledge about them. As a result, by applying wrong methods to even correct and complete data, faulty conclusions may be obtained. This is not the fault of statistical methods but of the persons who use them.

7.8 Need for Data

Data is required to make a decision in any business situation. The researcher is faced with one of the most difficult problems of obtaining suitable, accurate and adequate data. Utmost care must be exercised while collecting data because the quality of the research results depends upon the reliability of the data. Suppose, you are the Director of your company. Your Board of Directors has asked you to find out why the profit of the company has decreased since the last two years. Your Board wants you to present facts and figures. What are you going to do?

The first and foremost task is to collect the relevant information to make an analysis for the above-mentioned problem. It is, therefore, the information collected from various sources, which can be expressed in quantitative form, for a specific purpose, which is called data. The rational decision maker seeks to evaluate information in order to select the course of action that maximizes objectives. For decision making, the input data must be appropriate. This depends on the appropriateness of the method chosen for data collection. The application of a statistical technique is possible when the questions are answerable in quantitative nature, for instance; the cost of production, and profit of the company measured in rupees, age of the workers in the company measured in years. Therefore, the first step in statistical activities is to gather data.

7.9 Basics Concepts

The information collected from various sources through the use of different tools and techniques generally comprise numerical figures, ratings, descriptive narrations, responses to open-ended"question, quotations, field notes, etc. This information is called data. In nutritional research, usually two types of data are used universally. They are, quantitative data and qualitative data.

Quantitative Data:

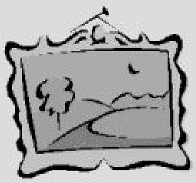
Quantitative data are obtained by applying various scales of measurement. The traits1 experiences of people are fit into standard responses to which numerical values are attached. To illustrate, respiratory rate has a numeric outcome as 68 per minute, birth weight 2 2.5 kg or hemoglobin level 2 11.0 mg/dl etc. These data are close-ended and hardly provide any depth or details. Quantitative data are either parametric or non-parametric. Parametric data undergo metric scale measurement. For example, in measuring respiratory rate, we make use of metric scale measurement. The score on a nutrition educational test or inventory is another illustration of numeric scale measurement. Non-parametric data are obtained by applying nominal or ordinal scales of measurement. These data are either coded, counted or ranked.

Qualitative Data

Qualitative data are verbal or symbolic. The detailed descriptions of observed behaviors. people, situations and events, are some examples of qualitative data. For example, the responses to open ended questions of a questionnaire or a schedule, first-hand information from people about their experiences, ideas, beliefs, and selected content or excerpts from documents, case studies personal diaries and letter are other examples of qualitative data.

Example 1:

Oil Painting

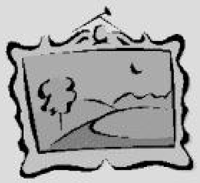


Qualitative data:

- blue/green color, gold frame
- smells old and musty
- texture shows brush strokes of oil paint
- peaceful scene of the country
- masterful brush strokes

Example 1:

Oil Painting



Quantitative data:

- picture is 10" by 14"
- with frame 14" by 18"
- weighs 8.5 pounds
- surface area of painting is 140 sq. in.
- cost $300

Cross-sectional Data

These are data from units observed at the same time or in the same time period. The data may be single observations from a sample survey or from all units in a population. Examples of Norwegian cross-section data are the Household Budget Survey for the year 1999, The Manufacturing Statistics for the year 2000, the Population Census for the year 2001.

Cross-Section data show spatial variation: Variation across units (individuals, households, firms, ...)

Time series Data

These are data from a unit (or a group of units) observed in several successive periods. Examples of Norwegian timeseries data are National Accounts data (production, private and public consumption, investment, export, import etc.), the Index of Manufacturing Production, the Consumer Price Index and Financial statistics (money stock, exchange rates, interest rates, bank deposits, etc.)

Time-Series data show temporal variation: Variation over periods (years, months, weeks, seconds, ...)

Most often cross-section data are data for micro units – individuals, households, firms, companies, etc. But macro-like cross-section data may well occur; examples are cross-section data for municipalities, other local units, counties or or even countries. In cross-section data, all data variation goes across units; we have variation across space (spatial variation).

Most often time-series data are macro data or macro-type data, for example timeseries for macro-economic variables from the National Accounts. But micro-data may also occur as time-series, for example time-series for a particular household or time-series for a particular firm. In time-series data the data variation goes over time periods; we have variation over time (time serial variation).

Population VS Sampling

Populationis the entire aggregation of cases that meet the designated set of criteria. It includes a complete set of persons or objects that possess same characteristics that is of interest to the researcher. The population usually is described as Target population which is also called the universe, is composed of entire group of people or objects to which the researcher wishes to generalize the findings of a study, target population consists of people or things that meet the designated set of criteria of interest to researcher. Population is not restricted to human subjects. A population might consist of all the hospital records on file or all blood samples. The researcher usually samples from an available group, called the accessible population or study population. The researcher needs to identify the accessible population from which generalization of the study finding can be drawn. The conclusions of research study are based on data obtained from accessible population and the statistical inferences should 60 be made only to the group from which sample was randomly selected.

The samplerepresents the population of those critical characteristics you plan to study. In other words, if the sample is representative of the population you can say that what you have found out about the sample is true of the population. The term representative means that sample subjects are not selected haphazardly, but deliberately so that every element in the population has an equal chance of being selected for the study.

If you take all the rural mothers of village "A" as your population, you will perhaps study one whole population. Whereas if you define your population as all rural mothers of India-you cannot possibly study the whole group. Even if you can, it is a waste of money and time. The process of selecting a fraction of the sampling unit of your target population for inclusion in your study is called sampling.

The population "element" consists of the characteristics or attributes of the subjects that the researcher wants to study. A true representative sample consists of similar elements; these are called the sampling elements. Each sampling unit can be the sampling element itself or it may contain a number of sampling elements. For example, the researcher wants to study the health status of rural mothers and finds that there are Hindu, Muslim and Christian mothers from which he/she can select the sample; hence each group becomes a sampling unit. In each unit, a number of sampling elements, mothers in this case are present. Sampling technique will then involve getting representatives from each unit.

Sampling refers to the process of selecting portion of the population that represents the entire population. A sample then consists of the subsets of the Population, Sample and Sampling population. A single member of a population is called elements

	Population	**Sample**
Definition	Complete enumeration of items is considered	Part of the population chosen for study
Characteristics	Parameters	Statistics
Symbols	Population Size = N Population Mean =μ Population S.d = σ	Sample Size = n Sample mean = x Bar Sample S.d = s

Elements

States In India	**Population**
Uttar Pradesh	199,812,341
Bihar	103,804,637
Maharashtra	112,372,972

Each of the states listed in the table is an element or member of the sample. For example, Uttar Pradesh is a member or element of the sample. An element could be an item, a state, a person, and so forth.

Observations:

The value assigned to only one element is called an observation. For example, 199,812,641 is an observation. An observation is also called measurement.

Variable:

Anything that can take on different values is called a variable. Therefore, Population is a variable since it has different values.

Other examples of variables could be number of students who graduate from college, income of senior citizens, types of health insurance plans people enrolled in.

7.10 Sources of Data

The increasingly complex nature of business and government has focused attention on the uses of research methodology in solving managerial problems. The credibility of the results derived from the application of such methodology is dependent upon the upto date information about the various pertinent characters included in the analysis. To illustrate, the demand of disc records has dropped dramatically after cassettes have entered into the market commercially. This information must be taken into consideration for formulating marketing strategy by a dealer selling musical products. Information expressed in appropriate quantitative form are known as data. The necessity and usefulness of information gathering or data collection cannot be overemphasized in government policies. The government must be aware of the actual scenario of the acceptance of family planning before it can formulate any policy in this matter. The components of this scenario are provided by appropriate data to be collected from various families. In industrial disputes regarding wages, cost of living index, a data-based indicator of inflation is often accepted as a guideline for arbitration.

In short, neither a business decision nor a governmental decision can be made in a casual manner in the highly involved environment prevailing in this age. It is through appropriate data and their analysis that the decision maker becomes equipped with proper tools of decision making.

Primary and Secondary Data

The primary data are those which are collected afresh and for the first time, and thus happen to be original in character. Such data are published by authorities who themselves are responsible for their collection.

The secondary data, on the other hand, are those which have already been collected by some other agency and which have already been processed. Generally speaking, secondary data are information which have previously been collected by some organization to satisfy its own need but it is being used by the department under reference for an entirely different reason.

However, the difference between primary and secondary data is largely of degree, and there is hardly any watertight difference between them. The data collected through primary sources by one investigator may be secondary in the hands of others.

For example, the census figures are published every tenth year by the Registrar General of India. But the census figures are also used by demographers and other social scientists for planning and research. Thus, the officials of the department of Registrar General will visualize the census figures as primary data. But a demographer using the same census figures to prepare a mortality table will consider them as secondary data.

Even when an organization is interested to collect primary data it is necessary to take help of various types of secondary data to design a proper sampling scheme.

Methods of Collecting Primary Data

The collection of primary data for business research is of paramount importance to assist management in making decisions. Generally, information regarding a large number of characteristics is necessary to analyze any problem pertaining to management. For instance, a study relating to employment in rural areas requires data on income, wages, types of crops and land holdings. The collection of primary data thus requires a great deal of deliberation and expertise.

Depending upon the nature of information necessary the following methods of collecting primary data are available.

A. Observation

Observation is a systematic viewing, coupled with consideration of the seen phenomena. The Concise Oxford Dictionary defines observation as accurate watching and noting of phenomena as they occur in nature with regard to cause and effect or mutual relations. The required information is obtained directly through observation rather than through the reports of others. In the case of behavior, one finds out what the individual does, rather than what the individual says he does. If the informants are unable to provide the information or can' give only very in exact answer, questioning is not useful and observation is the only way to proceed. For instance, when you are studying the behavior of small children who cannot speak, you can collect the information by observing the children under different circumstances. You should remember that all phenomena are not open to observation. Even if a phenomenon is open to observation, it may not find a ready observer at hand. Observation may be participant observation or non-participant observation. In Participant Observation Method the observer joins in the daily life of the group or organization he is studying. He watches what happens to the members of the community and how they behave. He also engages in conversation with them to find out their reactions to, and interpretations of, the events that have occurred In the Non-Participant Observation Method in order to. collect information the observer will not join the group or organization he is studying but will watch it from outside.

B. Personal Interviewing

Under this method data are collected by the investigator himself through interviews. Therefore, the enquiry is intensive rather than extensive. Under this method the investigator meets the informants personally, asks them questions pertaining to enquiry and collects the desired information. Thus, if a person wants to collect data on the wages of workers of the National Ball Bearing Company, he would go to the factory site of this company, contact the workers and collect the relevant information. Thus, this method is generally used in small size surveys confined to a small locality.

Interviews can be formal or informal. In Formal Interviewing set questions are asked and the answers are recorded in a standardized form. This is the practice in large scale interviews where a number of investigators are assigned to the job of interviewing. In a formal interview, the interviewer's bias is minimized. This type of interview is most suitable when you know very clearly what type of information you require for your survey. In the case of Informal Interviewing, the investigator may not have a set of questions but have only a number of key points around which to build the interview. The interviewer is at liberty to vary the sequence of questions, to explain their meaning, to add additional ones and even to change the wording. Informal interviews are preferred in the case of an explorative survey where you are not sure about the type of data you collect.

C. 'Through Local Reports and Correspondents

Under this method the investigator appoints local agents or correspondents in different places of the field of enquiry and the relevant information is obtained through them. These correspondent's and transmit the information to the office of the investigator. Newspaper agencies generally adopt this method. This method is also used by various departments of the Government in cases where regular information is to be collected from a relatively wide area. In case of making crop estimates or for obtaining regular information regarding prices of different commodities for the preparation of price index this method is generally used.

D. Questionnaire

A popular and common method of collection of primary data is by personally interviewing individuals, recording their answers in a structured questionnaire. The complete enumeration of Indian decennial census is performed by this method. The enumerators visit the dwellings of individuals and put questions to them which elicit the relevant information about the subject of enquiry. This information is recorded in the questionnaire. Occasionally a part of the questionnaire is unstructured so that the interviewee can feel free to share information about intimate matters with the interviewer. As the data are collected by the field staff personally it is also known as personal interview method. Much of the accuracy of the collected data, however, depends on the ability and tactfulness of investigators, who should be subjected to special training as to how they should elicit the correct information through friendly discussions.

The questionnaire can be broadly categorized into two types:

1. structured questionnaire

2. unstructured questionnaire.

1. ***Structured questionnaires***are prepared in advance. They contain definite and concrete questions. The structured questionnaire may contain close ended questions and open-ended responses. In the close ended questionnaire, the question setter gives alternative options for which the respondent has to give definite response. The best example of the close ended questionnaire format is the one that leads respondents to the "Yes" or "No"/ "True" or "False" answers.
2. ***Unstructured questionnaires***are those that are not structured in advance, and the investigators may adjust questions according to their needs during an interview

E. Schedule

This method of data collection is similar to that of the questionnaire. The schedule is also a proforma containing a set of questions. The 'difference between the questionnaire and the schedule is that the schedule is being filled in by the enumerators who are specially appointed for the purpose. These enumerators go to respondents with the schedules and ask them the questions from the schedule in the order they are listed. The enumerator records the replies in the space meant for the same in the schedule itself. In certain situations, schedules are handed over to respondents and the enumerators help the respondents in recording the answers. Enumerators explain, the objectives of the investigation and also remove the difficulties which the respondent may feel in understanding the implications of a particular question(s) or the definition or concept of difficult terms. Thus, the essential difference between the questionnaire and schedule is that the former (i.e., questionnaire) is sent to the informants by post and in the latter case the enumerators carry the schedule personally to informants and fill them in their own handwriting. This method is usually adopted in investigations conducted by governmental agencies or by some big organizations. For instance, population census all over the world is conducted through this method.

Data collection through schedules requires enumerators for filling up schedules and as such they should be very carefully selected. They should be trained to perform their job well. They should be intelligent and must possess the capacity of cross-examination in order to find out the fact. Above all, they should be honest, sincere, hardworking and should have the patience and perseverance. In drafting the schedules, all points stated for a good questionnaire, must as well be observed.

Choice of Method:

You have noticed that there are various methods and techniques for the collection of primary data. You should be careful while selecting the method which should be appropriate and effective. The selection of the methods depends upon various factors like scope and objectives of the inquiry, time, availability of funds, subject matter of the research, the kind of information required, degree of accuracy etc. As apprised, every method has its own merits and demerits. For example, the observation method is suitable for field surveys when the incident is really happening, the interview method is suitable where direct observation is not possible. Local reporter/correspondent method is suitable when information is required at regular intervals. The questionnaire method is appropriate in extensive enquiries where sample is large and scattered over large geographical areas and the respondents are able to express their responses in writing. The Schedule method is suitable in case respondents are illiterate.

Methods of Collecting Secondary Data

a. In case you decide to collect secondary data, you have to look into various sources from where you can obtain it. The source from which you actually collect the data depends upon the nature of the problem. The sources of secondary data can broadly be classified into two categories:

i) published, and ii) unpublished sources.

Let us discuss these two sources in detail.

A. Published Source

1. **Various publications of Central, State and local governments**: The important official publications are Statistical Abstract, India-Annual; Monthly Abstract of Statistics (both published by Central Statistical Organization); Indian Agricultural Statistics (Annual)

(Published by Ministry of Food and Agriculture); Index Number of Wholesale Prices in India (Weekly) (Published by Ministry of Commerce and Industry); Reserve Bank of India Bulletin (Monthly) (Published by Reserve Bank of India).

2. **Various publications of foreign governments or of international bodies**: The important publications are publications of international bodies like UNO, FAO, WHO, UNESCO, ILO, Statistical Year Book (Published by the Statistical Office of the United Nations), Yearbook of Labour Statistics (Published by ILO, Geneva). The secondary data provided by such publications are authentic, but along with other things, one must be especially careful about the units in respect of currency, weight etc. which greatly vary from one country to another.
3. **Journals of trade, commerce, economics, engineering** etc. published by responsible trade associations, Chambers of Commerce provide secondary data in respect of some important items. Some examples of this kind of publications are "Annual Report of the Chief Inspector of Mines in India" (issued annually by the office of the Chief Inspector of Mines, Dhanbad) and "Indian Textile Bulletin (issued monthly by the Textile Commissioner, Bombay).
4. **The other sources of secondary data**are books, magazines and newspapers, reports prepared by various universities, historical documents, diaries, letters, unpublished biographies and autobiographies.

B. Unpublished Source

All statistical material is not necessarily available in published form. There are various sources of unpublished data which can also be used wherever necessary. Unpublished data may generally be found in diaries, letters, unpublished biographies and autobiographies. Unpublished data may also be available with scholars and research workers, trade associations, labor bureau and other public/private organizations and individuals.

Precautions In Using Secondary Data

Primary data are to be scrutinized after the questionnaires are completed by the interviewers. Likewise, the secondary data are to be scrutinized before they are compiled from the source. The scrutiny should be made to assess the suitability, reliability, adequacy and accuracy of the data to be compiled and to be used for the proposed study.

1. Suitability: The compiler should satisfy himself that the data contained in the publication will be suitable for his study. In particular, the conformity of the definitions, units of measurement and time frame should be checked. For example, one US gallon is different from one British gallon.
2. Reliability: The reliability of the secondary data can be ascertained from the collecting agency, mode of collection and the time period of collection. For instance, secondary data collected by a voluntary agency with unskilled investigators are unlikely to be reliable.
3. Adequacy: The source of data may be suitable and reliable but the data may not be, adequate for the proposed enquiry. The original data may cover a bigger or narrower geographical region or the data may not cover suitable periods. For instance, per capita income of Pakistan prior to 1971 is inadequate for reference during the subsequent periods as it became separated into two different countries with-considerable variation in standard of living.
4. Accuracy: The user must be satisfied about the accuracy of the secondary data. The process of collecting raw data, the reproduction of processed data in the publication, the degree of accuracy desired and achieved should also be satisfactory and acceptable to the researcher.

The pattern of business and industry in the present-day environment has become quite complex and involved due to a variety of reasons. Any meaningful decision to be made in this context must be objective and fact based in nature. This is achieved by collecting and analyzing appropriate data. Data may broadly be divided into two categories, namely primary data and secondary data. The primary data are those which are collected for the first time by the organization which is using them. The secondary data, on the other hand, are those which, have already been collected by some other agency but also can be used by the organization under consideration. Primary data maybe collected by observation, oral investigation, questionnaire method or by telephone interviews. Questionnaires may be used for data collection by interviewers. They may also be mailed to prospective respondents. The drafting of a good questionnaire requires utmost skill. The process of interviewing also requires a great deal of tact, patience and, competence to establish rapport with the respondent. Secondary data are available in various published and unpublished documents. The suitability, reliability, adequacy and accuracy of the secondary data should, however, be ensured before they are used for research problems.

Summary

In this unit we elaborated on the meaning of data, methods of data collection, merits and limitations of data collection, precautions which are needed for the collection of data. The information collected from various processes for a specific purpose is called data.

Statistical data may be either primary data or secondary data. Data which is collected originally for a specific purpose is called primary data. The data which is already collected and processed by someone else and is being used now in the present study, is called secondary data. Secondary data can be obtained either from published sources or unpublished sources. It should be used if it is reliable, suitable and adequate, otherwise it may result in misleading conclusions. It has its own merits and demerits. There are several problems in the collection of primary data. These are: tools and techniques of data collection, degree of accuracy, designing the questionnaire, selection and training of enumerators, problem of tackling non-responses and other administrative aspects. Several methods are used for collection of primary data. These are: observation, interview, questionnaire and schedule methods. Every method has its own merits and demerits. Hence, no method is suitable in all situations. The suitable method can be selected as per the needs of the investigator which depends on objective nature and scope of the enquiry, availability of funds and time. Further, arrangement of data is discussed which is very necessary for taking decisions in the market.

Keywords

- **Data**: Quantitative or/ and qualitative information, collected for study and analysis.
- **Interview:** A method of collecting primary data by meeting the informants and asking the questions.
- **Observation**: The process of observing individuals in controlled situations.
- **Questionnaire**: is a device for collection of primary data containing a list ofquestions pertaining to enquiry, sent to the informants, and the informant himselfwrites the answers.
- **Primary Data:** Data that is collected originally for the first time.
- **Secondary Data**: Data which were collected and processed by someone elsebut are being used in the present study.
- **Published Sources**: Sources which consist of published statistical information.
- **Schedule**: is a device for collection of primary data containing a list ofquestions to be filled in by the enumerators who are specially appointed for thatpurpose.
- **Elements:**Each of the states listed in the table is an element or member of the sample.
- **Observations:**The value assigned to only one element is called an observation.
- **Variables:**Anything that can take on different values is called a variable.
- **Sample:** Subset of a population selected to participate in research study.

SelfAssessment

1. A variable is any characteristic which can assume ____________________ values.

A. Different
B. Similar
C. Fixed
D. Assumed

2. In ___________________ type of classification, the data is grouped together according to some distinguished characteristic or attribute, such as religion, sex, age, national origin, and so on.

A. Quantitative
B. Chronological
C. Qualitative
D. All of the above

3. A summary measure that describes any given characteristic of the population is known as a __________________.

A. Parameter
B. Information
C. Inference
D. Statistics

4. Which of the following is a qualitative random variable?

A. height of the cable car station on the top of Table Mountain in Cape Town
B. distance in meters that you have walked already today
C. nicotine content of a cigarette
D. highest educational qualification of respondents in an interview

5. The degree of randomness of selection would depend upon the process of selecting the items from the ________________________.

A. Population
B. Region
C. Sample
D. Data

6. When an investigator uses the data which has already been collected by others, such data is called _______________ _________.

A. Primary data
B. Collected data
C. Processed data
D. Secondary data

7. Questionnaire survey method is used to collect

A. Secondary data
B. Qualitative variable
C. Primary data
D. None of these

8. _____________________ is one which is collected by the investigator himself for the purpose of a specific inquiry or study.

A. Secondary data
B. Primary data
C. Statistical data
D. Published data

9. Census reports used as a source of data is

A. Primary source
B. Secondary source
C. Organized data
D. None

10. Data collected by NADRA to issue computerized identity cards (CICs) are

A. Unofficial data
B. Qualitative data
C. Secondary data
D. Primary data

11. Which of the following is a branch of statistics?
A. Descriptive statistics
B. Inferential statistics
C. Industry statistics
D. Both A and B

12. Review of performance appraisal, labour turnover rates, planning of incentives, and training programs are the examples of which of the following?
A. Statistics in production
B. Statistics in marketing
C. Statistics in finance
D. Statistics in personnel management

13. Statistics branches include
A. Applied Statistics
B. Mathematical Statistics
C. Industry Statistics
D. Both A and B

14. Inferential statistics help in generalizing the results of a sample to the entire population.
A. True
B. False

15. The specific statistical methods that can be used to summarize or to describe a collection of data is called:
A. Descriptive statistics
B. Inferential statistics
C. Analytical statistics

D. All of the above

16. The need for inferential statistical methods derives from the need for ______________.

A. Population
B. Association
C. Sampling
D. Probability

Answers for SelfAssessment

l.	A	2.	B	3.	A	4.	D	5.	C
6.	D	7.	C	8.	B	9.	B	10.	D
11.	D	12.	D	13.	D	14.	A	15.	A
16.	C								

Review Questions

1. What do you mean by data? Why it is needed for research?
2. State whether the following data are primary or secondary?

a. The Secretary, Merchant Chamber of Commerce is using the figures published in "Reserve Bank of India Bulletin" (Published monthly by Reserve Bank of India) as the basis of forecasting money supply during the next month.

b. The Secretary, department of mines is writing a report on various types of mining accidents using the data available in the "Annual Report of the Chief Inspector of Mines in India" issued by the Office of the Chief Inspector of Mines.

c. The Textile Commissioner is preparing a report on the prospect of textile export based on the data available in "Indian Textile Bulletin" published by his own office.

d. National Thermal Power Corporation is examining the supply of coal to their thermal power stations using the data available in "Monthly Coal Bulletin" published monthly by the office of the Chief Inspector of Mines.

3. State with reason the type of data collection method to be followed in the following cases.

a. A railway accident occurred at a distance of 20 km from New Delhi railway station. You are asked to ascertain the cause of the accident and the extent of disaster.

b. The Directorate of Cottage and Small Scale Industries desires to find out the reasons for non-payment of loan taken by a number of artisans and small traders. You are asked to find out the reasons of non-repayment.

c. The new General Manager claims that the telephone service of the city has become vastly improved during the last one year. You are assigned the responsibility of providing a databased report on this claim expeditiously.

4. Describe the various methods of collecting primary data and comment on their relative advantages and disadvantages.
5. Define secondary data. State their chief sources and point out the dangers involved in their use and the precautions necessary to use them. Illustrate with examples.
6. What are the advantages of using a frequency distribution to describe a body of raw data? What are the disadvantages?

7. A portfolio contains 51 stocks whose prices are given below:

   67 34 36 48 49 31 61 34 43 45 38 32 27 61 29 47 36 50 46 30 40 32 30 33 45 49 48 41 53 36 37 47 47 30 50 28 35 35 38 36 46 43 34 62 69 50 28 44 43 60 39

   Summarize these stock prices in the form of a frequency distribution.

8. Form a frequency distribution of the following data. Use an equal class interval of 4 where the lower limit of the first class is 10.

   10 17 15 22 11 16 19 24 29 18 25 26 32 14 17 20 23 27 30 12 15 18 24 36 18 15 21 28 33 38 34 13 10 16 20 22 29 29 23 31

9. The distribution of ages of 500 readers of a nationally distributed magazine is given below:

Age (in Years)	*Number of Readers*
Below 14	20
15–19	125
20–24	25
25–29	35
30–34	80
35–39	140
40–44	30
45 and above	45

   Find the cumulative frequency distributions for this distribution.

10. A. Illustrate two methods of classifying data in class-intervals.

    B. Why is it necessary to summarize data? Explain the approaches available to summarize data distributions.

Further Readings

- Kothari, C.R. 2004. Research Methodology Methods and Techniques, New Age International (P) Limited : New Delhi.
- Rao K.V. 1993. Research Methodology in Commerce and Management, Sterling Publishers Private Limited : New Delhi.
- Wilkinson, T.S. and P.L. Bliandarkar. 1979. Methodology and Techniques of Social Research, Himalaya Publishing House: Bombay.

Web Links

- Joselyn, R. W. (1977) Designing the marketing research, Petrocellis/Charter, New York, p.15
- Patton MQ (1990, 2nd ed.) Qualitative Evaluation and Research Methods. Newbury Park, USA: Sage Publications.

Unit 08: Descriptive Statistics

CONTENTS

Objectives

- Describe what is central tendency.
- Define and compute the arithmetic mean, geometric mean harmonic means, median etc.
- Explain the merits and limitations of different measures of central tendency or averages.

Introduction

As we know that after the classification and tabulation of data one often finds too much detail for many uses that may be made of information available. We, therefore, need further analysis of the tabulated data to draw inference. In this unit we are going to discuss about measures of central tendencies. For the purpose of analysis, very important and powerful tool is a single average value that represents the entire mass of data.

The term average in Statistics refers to a one figure summary of a distribution. It gives a value around which the distribution is concentrated. For this reason, that average is also called the measure of central tendency. For example, suppose Mr. X drives his car at an average speed of 60 km/hr. We get an idea that he drives fast (on Indian roads of course!). To compare the performance of two classes, we can compare the average scores in the same test given to these two classes. Thus, calculation of average condenses a distribution into a single value that is supposed to represent the distribution. This helps both individual assessments of a distribution as well as in comparison with another distribution.

8.1 Concepts of Central Tendency

Measures of central tendency i.e., condensing the mass of data in one single value, enable us to get an idea of the entire data. For example, it is impossible to remember the individual incomes of millions of earning people of India. But if the average income is obtained, we get one single value that represents the entire population. Measures of central tendency also enable us to compare two or more sets of data to facilitate comparison. For example, the average sales figures of April may be compared with the sales figures of previous months.

According to **Professor Bowley**, averages are "statistical constants which enable us to comprehend in a single effort the significance of the whole". They throw light as to how the values are concentrated in the central part of the distribution.

According to **King and Minium (2013)** described measures of central tendency as a summary figure that helps in describing a central location for a certain group of scores.

According to **Tate (1955)** defined measures of central tendency as "a sort of average or typical value of the items in the series and its function is to summarize the series in terms of this average value".

For a proper appreciation of various statistical measures used in analyzing a frequency distribution, it is necessary to note &at most of the statistical distributions have some common features. If we move from lowest value to the highest value of a variable, the number of items at each successive stage increases till we reach a maximum value, and then as we proceed further, they decrease. The statistical data which follow this general patternmay differ from one variable to another in the following three ways:

1) They may differ in the values of the variables around which most of the item's cluster (i.e., Average) 2) They may differ in the extent to which items are dispersed (i.e., Dispersion).

3) They may differ in the extent of departure from some standard distributions called normal distribution (i.e., Skewness and Kurtosis).

Accordingly, there are three sets of statistical measures to study these three kinds of characteristics. At present, however, we are confined to the first set of measures which are called Averages or Measures of Central Tendency or Measures of Location.

In the general pattern of distribution, in the data we may identify a value around which many other items of the data congregate. This is a value which is somewhere in the central part of the range of all values. When this typical item of the data is towards the central part of the data, it is known as Central Tendency. As it indicates the location of the clustering of items, it is also called a measure of location. Just as the title of an essay gives the central theme of the essay, the central tendency of the numerical data gives the central idea of the entire data. Look at Figure 10.1 carefully. It shows the central locations of three different curves A, B and C. You must have noticed that the central locations of curve A and curve C are equal. The central location of curve B lies to the right of those of curves A and C.

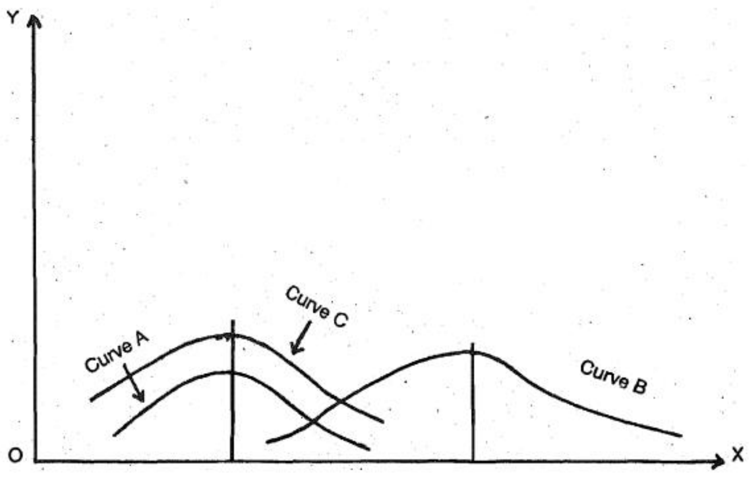


Fig 1: Central Location of Different Curves

8.2 Properties of A Good Measure of Central Tendency

A good measure of central tendency should possess, as far as possible, the following properties,

i. It should be easy to understand.
ii. It should he simple to compute.
iii. It should be based on all observations.
iv. It should be uniquely defined.
v. It should be capable of further algebraic treatment.

vi. It should not be unduly affected by extreme values.

Following are some of the important measures of central tendency which are commonly used in business and industry.

- Arithmetic Mean
- Weighted Arithmetic
- Mean Median
- Quantiles
- Mode
- Geometric
- Mean
- Harmonic
- Mean

8.3 Arithmetic Mean

The arithmetic mean is commonly known as mean. It is a measure of central tendency because other figures of the data congregate around it. Arithmetic mean is obtained by dividing the sum of the values of all observations in the given data set by the number of observations in that set. It is the most commonly used statistical average in the disciplines such as commerce, management, economics, finance, production, etc. The arithmetic mean is also called as simple Arithmetic Mean.

Mean for sample is denoted by symbol 'M or $\bar{x}$ ('x-bar')' and mean for population is denoted by 'μ' (mu). It is one of the most commonly used measures of central tendency and is often referred to as average. It can also be termed as one of the most sensitive measures of central tendency as all the scores in a data are taken in to consideration when it is computed (Bordens and Abbott, 2011). Further statistical techniques can be computed based on mean, thus, making it even more useful.

Mean is a total of all the scores in data divided by the total number of scores. For example, if there are 100 students in a class and we want to find mean or average marks obtained by them in a psychology test, we will add all their marks and divide by 100, (that is the number of students) to obtain mean.

For Ungrouped Data

Mean (M) is the most familiar and useful measure used to describe the central tendency average of a distribution of scores for any group of individuals, objects or events. It is computed by dividing the sum of the scores by the total number of scores.

Any data that has not been categorized in any way is termed as an ungrouped data. For example, we have an individual who is 25 years old, another who is 30 years old and yet another individual who is 50 years old. These are independent figures and not organized in any way, thus they are ungrouped data.

$M = \Sigma X / N$

Where, M = Mean

ΣX= Summation of scores in the distribution

N = Total number of scores.

Example 1:

The scores obtained by 10 students on psychology test are as follows:

58 34 32 47 74 67 35 34 30 39

Solution:

Step 1: In order to obtain mean for the above data we will first add the marks to obtain

ΣX: 58+ 34+ 32+ 47+ 74+ 67+ 35+ 34+ 30+ 39 = 450

Step 2: Now using the formula, we will compute mean

$M = \Sigma X/ N$ $\Sigma X= 450$, $N= 10$ (Total number of students)

Thus, $M= 450/ 10 = 45$

Thus, the mean obtained for the above data is 45

Computation of Mean for Grouped Data

The formula for computing mean for grouped data is

$M = \Sigma fX/ N$

Where, M= Mean

$\sum$= Summation

X= Midpoint of the distribution

f = the respective frequency

N = Total number of scores.

Example 2:

A class of 30 students was given a psychology test and the marks obtained by them were categorized in to six categories. The lowest marks obtained were 10 and highest marks obtained were 35. A class interval of 5 was employed. The data is given as follows:

Marks	Frequencies (f)	Midpoint (X)	fX
35- 39	5	37	185
30-34	7	32	224
25-29	5	27	135
20-24	6	22	132
15-19	4	17	68
10-14	3	12	12
	N=30		ΣfX = 780

The steps followed for computation of mean with grouped data are as follows:

Step 1: The data is arranged in a tabular form with marks grouped in categorieswith class interval of 5.

Step 2: Once the categories are created, the marks are entered under frequencycolumn based on which category they fall under.

Step 3: The midpoints of the categories are computed and entered under X.

Step 4: fX is obtained by multiplying the frequencies and midpoints for eachcategory.

Step 5: fX for all the categories are added to obtain ΣfX, in case of ourexample it is obtained as 780

Step 6: The formula $M = \Sigma fX/ N$ is used, N is equal to 30.

$M = \Sigma fX/ N$ $M = 780/ 30= 26$

Thus, the mean obtained is 26.

Computation of Mean by Shortcut Method (with Assumed mean)

In certain cases data is very large and it is not possible to compute each fX. In such situations, a short cut method with the help of assumed mean can be computed. A real mean can thus be computed with application of correction.

The formula is

M =AM+(Σfx′/N×i)

Where,

AM= Assumed mean,

Σ = Summation

i = Class interval

x′ = {(X − AM)/ i}, X the midpoint of the scores in the interval

f = the respective frequency of the midpoint

N = the total number of frequencies or students.

Marks	Frequencies (f)	Midpoint (X)	x′={(X − AM)/ i}	f x′
35- 39	5	37	3	15
30-34	7	32	2	14
25-29	5	27	1	5
20-24	6	22	0	0
15-19	4	17	-1	-4
10-14	3	12	-2	-6
	N=30			Σ fx′= 24

Step 1: We will assume mean (AM) as 22.

Step 2: Difference is obtained between each of the midpoints and the assumed mean and then the same is divided by 'i' that is the class interval (5 in this case), these are then entered under column with heading x′= {(X − AM)/ i}. The x′ for 22 will be 0 .

Step 3: Frequency (f) is then multiplied with x′ to obtain f x′.

Step 4: All f x′are added to obtain Σ fx′, in the present example it is 24.

Step 5: The formula for mean is now applied

M =AM+(Σfx′/N×i)

M= 22 + (24/30 x 5)

= 22 + 4 = 26

Thus, mean is obtained as 26.

And if you refer to the mean obtained by the direct method and mean obtained with the shortcut method, the mean is the same, that is 26.

Computation of Mean by Shortcut Method (with Assumed mean)

Class-Interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	10	20	14	16	18	22

Solution:

Here, the intervals are of equal size. So we can apply the step-deviation method, in which

A = a + l · ∑di′fi/∑fi

Where a = assumed mean,

l = common size of class intervals

fi = frequency of the ith class interval

di′ = m1−al, mi being the class mark of the ith class interval.

Class-Intervals	Class Marks	Frequency	di = mi - a = mi - 28	di′= di/l =d_i/8	di′fi
0-8	4	10	-24	-3	-30
8-16	12	20	-16	-2	-40
16-24	20	14	-8	-1	-14
24-32	28	16	0	0	0
32-40	36	18	8	1	18
4-48	44	22	16	2	44
		∑fi= 100			∑di′fi = -22

Therefore, A = a + l · ∑di′fi/∑fi

= 28 + 8 · −22100

= 28 – 1.76

= 26.24.

Merits and Demerits of Arithmetic Mean

Merits of Arithmetic Mean

1. It utilizes all the observations;

2. It is rigidly defined;

3. It is easy to understand and compute; and

4. It can be used for further mathematical treatments.

Demerits of Arithmetic Mean

1. It is badly affected by extremely small or extremely large values;

2. It cannot be calculated for open end class intervals; and

3. It is generally not preferred for highly skewed distributions.

Weighted Mean

Weight here refers to the importance of a value in a distribution. A simple logic is that a number is as important in the distribution as the number of times it appears. So, the frequency of a number can also be its weight. But there may be other situations where we have to determine the weight based on some other reasons. For example, the number of innings in which runs were made may be considered as weight because runs (50 or 100 or 200) show their importance. Calculating the weighted mean of scores of several innings of a player, we may take the strength of the opponent (as judged by the proportion of matches lost by a team against the opponent) as the corresponding weight. Higher the proportion stronger would be the opponent and hence more would be the weight. If xi has a weight wi, then weighted mean is defined as:

$$\frac{\sum_{i=1}^{k} xiwi}{\sum_{i=1}^{k} wi}$$

Example 3: Prices of three commodities viz., A, B & C raised by 40 %, 60 % and 90 % respectively. Commodity A is six times more important than C, and B is three times more important than C. What is the mean rise in price of these three commodities?

Solution 3: As the mean rise in price is to be determined, the figures of rise in price will be denoted as x. The relative importanceof A: B:C is 6:3:1. So these figures will be taken as weights 'w'.

Commodity	Percentage rise in prices (x)	Weights (w)	wx
A	40	6	240
B	60	3	180
C	90	1	90
Total		∑w= 10	∑wx=510

Weighted Arithmetic Mean =∑wx/∑w

= 510/10

=51%

Mean rise in the prices is 51 %

It may be noted that for computation purpose, weights of items are treated in the same way as the frequencies of the items. In fact, weights are not frequencies. Frequency means number of times an item is repeated in the data, whereas weights only give the relative importance of various items. The items actually occur only once in the data.

Weighted arithmetic mean is also called Weighted Average. The word 'Average' in statistics, as pointed out earlier, is also used for other measures of central tendency viz., geometric mean, harmonic mean, etc. So, in broader sense, weighted average also includes weighted geometric mean and weighted harmonic mean.

Uses of Weighted Arithmetic Mean

Weighted arithmetic mean. is mainly useful under the following situations:

1) When the given items are of unequal importance

2) When averaging percentages which have been computed by taking different number of items in the denominator a

3) When statistical measures such as mean of several groups are to be combined

To be more specific, weighted arithmetic mean is used in the following cases:

1) Construction of Index Numbers.

2) Computation of standardized birth and death rates.

3) Finding out an average output per machine, where machines are of varying capacities.

4) Determining the average wages of skilled, semi-skilled and unskilled workers of a 'factory.

Harmonic Mean

If x1, x2, x3 ... xn be a set of n observations then the harmonic mean is defined as the reciprocal of the (arithmetic) mean of the reciprocals of the quantities.

Thus,

HM= $\frac{1}{\frac{1}{n}(\frac{f1}{x1}+\frac{f2}{x2}+\cdots\ldots\ldots\frac{fn}{xn})}$

In a frequency distribution harmonic mean is given by

HM = $\frac{1}{\frac{1}{n}(\frac{f1}{x1}+\frac{f2}{x2}+\cdots\ldots\ldots\frac{fn}{xn})}$

where N = ∑fi

Example 4: Find the harmonic mean from the following data.

X 2574, 475, 75, 5, 0.8, 0.08, 0.005, 0.0009

Solution 4: To find harmonic mean, calculate1/x .

x	1/x
2574	0.0004
475	0.0021
75	0.0133
5	0.2000
0.8	1.2500
0.08	12.500
0.005	200.00
0.0009	1111.1111

Here n=8

HM= N/(1/x)= 8/1325.0769

=0.006

Example 5: From the following data, compare the harmonic mean.

Marks	Students
10	20
20	30
25	50
40	15
50	5

Solution: To find HM we form the table as follows.

Marks	Students	f/x
10	20	2.00
20	30	1.500

25	50	2.00
40	15	0.375
50	5	0.100

Here N = 120 and ∑f/x = 5.975

Therefore, HM= N/∑f/x= 120/5.975= 20.08

Merits and Demerits of HarmonicMean

Merits of Harmonic mean

1. It is rigidly defined;

2. It utilizes all the observations;

3. It is amenable to algebraic treatment; and

4. It gives greater importance to small items.

Demerits of Harmonic Mean

1. Difficult to understand and compute.

Geometric Mean

For n numbers (X1 ,X2 ,...Xn)the geometric mean (GM) is defined as the nth root of the product of these n numbers, i.e.,

GM= $(X_1, X_2, \ldots X_n)^{1/n}$ =

$=\{\sum_{i=1}^{n} Xi\}^{1/n}$

Clearly, GM is not defined unless all the n numbers are positive. If any number is negative or zero, we cannot calculate GM. By taking logarithm of GM, we have

Log GM= $(1/n)(\log X_1 + \log X_2 + \log X_3 \ldots\ldots\ldots\ldots \log X_n)$

$= 1/n * \sum_{i=1}^{n} \log X\, i$

Which shows that now GM can be computed by using a log-table? Anti-logarithm of the arithmetic mean of log X values is GM.For the second data set, gross salary increased at the rate of 11% every year. In practice, however,increase/decrease will not be at a fixed rate over the years; and it is meaningful to talk about average rate because fixed rate situation is rare.

In general, GM is more appropriate average for percentage (or proportionate) rates of change than arithmetic mean as in the case of rise in various price indices, cost of living indices, etc.

Example 5: Daily income of ten families of a particular place is given below. Find geometric mean. 85, 70, 15, 75, 500, 8, 45, 250, 40, 60,

Solution: To find GM first, we calculate log x as follows

X	1/x
85	1.9294
70	1.8457
15	1.1761
75	1.8757
500	2.6990
2	0.9030

45	1.6532
250	2.3979
40	1.6021
36	1.5563

∑log x = 17.6373

GM = Antilog (∑log x/ n)

=Antilog (17.6373/10)=58.03

Merits and Demerits of Geometric Mean

Merits of Geometric Mean

1. It is rigidly defined;

2. It utilizes all the observations;

3. It is amenable to algebraic treatment (the reader should verify that if GM1 and GM2 are Geometric Means of two series-Series 1 of size n and Series 2 of size m respectively, then Geometric Mean of the combined series is given by

Log GM = (n GM1 + m GM2) / (n + m);

4. It gives more weight to small items; and

5. It is not affected greatly by sampling fluctuations.

Demerits of Geometric Mean

Difficult to understand and calculate; and

It becomes imaginary for an odd number of negative observations and becomes zero or undefined if a single observation is zero.

8.4 Median

Median is that value of the variable which divides the whole distribution into two equal parts. Here, it may be noted that the data should be arranged in ascending or descending order of magnitude. When the number of observations is odd then the median is the middle value of the data. For even number of observations, there will be two middle values. So we take the arithmetic mean of these two middle values. Number of the observations below and above the median, are same. Median is not affected by extremely large or extremely small values (as it corresponds to the middle value) and it is also not affected by open end class intervals. In such situations, it is preferable in comparison to mean.

Median for Ungrouped Data

Mathematically, if x1, x2,…, xn are the n observations then for obtaining the median first of all we have to arrange these n values either in ascending order or in descending order. When the observations are arranged in ascending or descending order, the middle value gives the median if n is odd. For even number of observations there will be two middle values. So we take the arithmetic mean of these two values.

$M_d = (n+1/2)^{th}$observation ; when n is odd

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2} ; Whenniseven$$

Example 6: Find median of following observations:

Solution: 6,4,3,7,8

Solution: First we arrange the given data in ascending order as

3, 4,6,7,8

Since, the number of observations i.e. 5, is odd, so median would be the middle value that is 6.

Example 7:Calculate median for the following data:

Solution: 7,8,9,3,4,10

Solution: First we arrange given data in ascending order as

3, 4, 7, 8, 9, 10

Here, Number of observations (n) = 6 (even). So we get the median by

$M_d = \frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2}$

$= \frac{\left(\frac{6}{2}\right)^{th} observation + \left(\frac{6}{2}+1\right)^{th} observation}{2}$

= (3^{rd} observation +4^{th} observation)/2

= (7+8)/2= 15/2= 7.5

For Ungrouped Data (when frequencies are given)

If xi are the different value of variable with frequencies fi then we calculate cumulative frequencies from fi then median is defined by

Md = Value of variable corresponding to $\left(\frac{\Sigma f}{2}\right)^{th} = \left(\frac{N}{2}\right)^{th}$ *cumulative frequency*

Note: If N/2 is not the exact cumulative frequency then value of the variable corresponding to next cumulative frequencies is the median.

Example 8:Find Median from the given frequency distribution.

X	F
20	7
40	5
60	4
80	3

Solution: First we find cumulative frequency

X	F	c.f
20	7	7
40	5	12
60	4	16
80	3	19
	$\sum_{i=1}^{n} fi = 19$	

Md = Value of the variable corresponding to the

(19/2)th cumulative frequency

=Value of the variable corresponding to 9.5 since 9.5 is not among c.f

So, the next cumulative frequency is 12 and the value of variable against 12 cumulative frequency is 40. So median is 40.

Median for Grouped Data

For class interval, first we find cumulative frequencies from the given frequencies and use the following formula for calculating the median:

Median = L+ (N/2-C)/f *h

Where, L = lower class limit of the median class, N = total frequency, C = cumulative frequency of the pre-median class, f = frequency of the median class, and h = width of the median class.

Median class is the class in which the (N/2) th observation falls. If N/2 is not among any cumulative frequency then next class to the N/2 will be considered as median class.

Example 9:

X	F	c.f
0.-10	3	3
10-20	5	8
20-30	7	15
30-40	9	24
40-50	4	28

$\sum_{i=1}^{n} fi$ = 28 = N/2= 28/2=14

Since 14 is not among the cumulative frequency so the class with next cumulative frequency i.e. 15, which is 20-30, is the median class.

We have L = lower class limit of the median class = 20 N = total frequency = 28 C = cumulative frequency of the pre median class =8 f = frequency of the median class = 7 h = width of median class = 10

Now substituting all these values in the formula of Median

Median = L+ (N/2-C)/f *h

= 20+ (14-8)/7 *10= 28.57

Therefore, median is 28.57.

Merits and Demerits of Median

Merits of Median

1. It is rigidly defined;

2. It is easy to understand and compute;

3. It is not affected by extremely small or extremely large values; and

4. It can be calculated even for open end classes (like “less than 10” or “50

and above”).

Demerits of Median

1. In case of even number of observations we get only an estimate of themedian by taking the mean of the two middle values. We don’t get itsexact value;

2. It does not utilize all the observations. The median of 1, 2, 3 is 2. If theobservation 3 is replaced by any number higher than or equal to 2 and ifthe number 1 is replaced by any number lower than or equal to 2, themedian value will be unaffected. This means 1 and 3 are not being utilized;

3. It is not amenable to algebraic treatment; and

4. It is affected by sampling fluctuations.

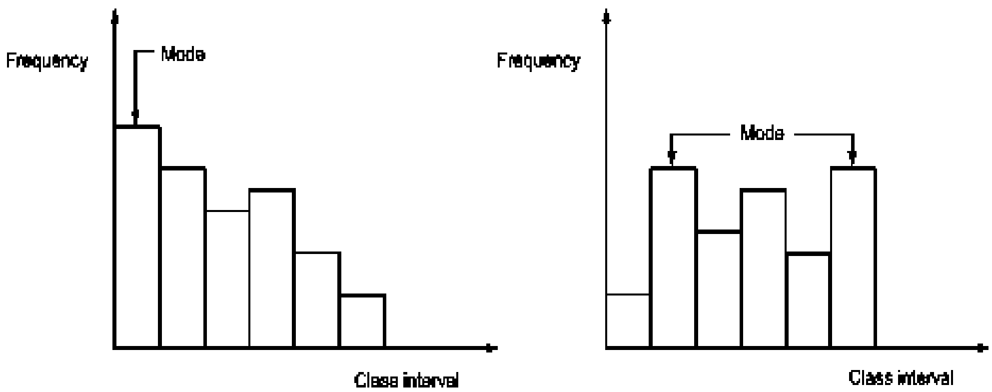
8.5 Mode

The mode is that value of an observation which occurs most frequently in the data set, that is, the point (or class mark) with the highest frequency.

The concept of mode is of great use to large scale manufacturers of consumable items such as ready-made garments, shoe-makers, and so on. In all such cases it is important to know the size that fits most persons rather than 'mean' size.

There are many practical situations in which arithmetic mean does not always provide an accurate characteristic (reflection) of the data due to the presence of extreme values. For example, in all such statements like 'average man prefers . . . brand of cigarettes', 'average production of an item in a month', or 'average service time at the service counter'. The term 'average' means majority (i.e., mode value) and not the arithmetic mean. Similarly, the median may not represent the characteristics of the data set completely owing to an uneven distribution of the values of observations. For example, suppose in a distribution the values in the lower half vary from 10 to 100 (say), while the same number of observations in the upper half very from 100 to 7000 (say) with most of them close to the higher limit. In such a distribution, the median value of 100 will not provide an indication of the true nature of the data. Such shortcomings stated above for mean and median are removed by the use of mode, the third measure of central tendency.

The mode is a poor measure of central tendency when most frequently occurring values of an observation do not appear close to the centre of the data. The mode need not even be unique value. Consider the frequency distributions shown in Fig. 3.3(a) and (b). The distribution in Fig. 3.3(a) has its mode at the lowest class and certainly cannot be considered representative of central location. The distribution shown Fig. 3.3(b) has two modes. Obviously neither of these values appear to be representative of the central location of the data. For these reasons the mode has limited use as a measure of central tendency for decision-making. However, for descriptive analysis, mode is a useful measure of central tendency.



(a) Unimodal Distribution (b) Bimodal Distribution

For Ungrouped Data

Mathematically, if x1, x2,…, xn are the n observations and if some of the observation are repeated in the data, say i x is repeated highest times then we can say the i x would be the mode value.

Example 10: Find mode value for the given data

2, 2, 3, 4, 7, 7, 7, 7, 9, 10, 12, 12

Solution:First we prepare frequency table as

X	F

2	2
3	1
4	1
7	4
9	1
10	1
12	2

This table shows that 7 have the maximum frequency. Thus, mode is 7.

For Grouped Data:

Data where several classes are given, following formula of the mode is used

$M_0 = L + \frac{[f1-f0]}{[f1-f0]+[f1-f2]} * h$

where, L = lower class limit of the modal class, 1 f = frequency of the modal class, 0 f = frequency of the pre-modal class, 2 f = frequency of the post-modal class, and h = width of the modal class

Example 11:

X	F
0.-10	3
10-20	5
30-40	7
40-50	9
50-60	4

Corresponding to highest frequency 9 model class is 40-50 and we have

L = 40, f1 = 9, f 7 o = , f 4 2 = , h = 10

Applying the formula,

Mode= 40+ (9-7)/(2*9-7-4)*10

=42.86

Merits and Demerits of Mode

Merits of Mode

1. Mode is the easiest average to understand and also easy to calculate;

2. It is not affected by extreme values;

3. It can be calculated for open end classes;

4. As far as the modal class is confirmed the pre-modal class and the post modal class are of equal width; and

5. Mode can be calculated even if the other classes are of unequal width.

Demerits of Mode

1. It is not rigidly defined. A distribution can have more than one mode;

2. It is not utilizing all the observations;

3. It is not amenable to algebraic treatment; and

4. It is greatly affected by sampling fluctuations.

Summary

In the present unit, we discussed the concept of central tendency. The measures of central tendency were explained as summary figures that help in describing a central location for a certain group of scores. It was further explained as providing information about the characteristics of the data by identifying the value at or near the central location of the data. The functions of measures of tendency besides the characteristics of good measures of central tendency were also discussed. Further, the unit focused on the three measures of central tendency, namely, mean, median and mode. Mean is a total of all the scores in data divided by the total number of scores. It is one of the most frequently used measure of central tendency and is often referred to as an average. It can also be termed as one of the most sensitive measures of central tendency as all the scores in a data are taken in to consideration when it is computed. Median is the middle score in an ordered distribution. Median is a point in anydistribution below and above which lie half of the scores. Mode is the score in a distribution that occurs most frequently. Certain distributions are bimodal, where there are two modes.When there are three modes, the term used is trimodal and when there are four or more modes, we use the term multimodal. Though, if the scores in a distribution greatly vary, then it is possible that there is no mode. The properties, advantages and limitations of mean, median and mode were also discussed in detail. Further, the computation of each of these measures of central tendency was also discussed for both ungrouped and grouped data with stepwise explanation.

Keywords

Measures of Central Tendency: Measures of central tendency can be explained as a summary figure that helps in describing a central location for a certain group of scores.

Mean: Mean is a total of all the scores in data divided by the total number of scores.

Median: Median is a point in any distribution below and above which lie half of the scores.

Mode: Mode is the score in a distribution that occurs most frequently.

Central tendency:A single value that has a tendency to be somewhere at the center and within range of all values.

Geometric Mean: For n numbers (X_1 ,X_2 ,...$X_{n)}$the geometric mean (GM) is defined as the nth root of the product of these n numbers.

Harmonic Mean: If x1, x2, x3 ... xn be a set of n observations then the harmonic mean is defined as the reciprocal of the (arithmetic) mean of the reciprocals of the quantities.

SelfAssessment

1. Which measure of central tendency takes into account the magnitude of scores?

A. Median

B. Range

C. Mode

D. Mean

2. Which of the following is not a common measure of central tendency?

A. Mean

B. Mode

C. Median

D. Range

3. Which of the following is a characteristic of a mean?
A. The sum of deviations from the mean is zero
B. It minimizes the sum of squared deviations
C. It is affected by extreme scores
D. All of the above

4. The total of all the observations divided by the number of observations is called:
A. Arithmetic mean
B. Geometric mean
C. Median
D. Harmonic mean

5. The values of extreme items do not influence the average for_________.
A. Mean
B. Mode
C. Median
D. None of the above

6. To calculate the median, all the items of a series have to be arranged in a/an ________.
A. Descending order
B. Ascending order
C. Ascending or descending order
D. None of the above

7. The values of extreme items do not influence the average for_________.
A. Mean
B. Mode
C. Median
D. None of the above

8. The number of observations smaller than ________ is the same as the number of observations larger than it.
A. Median
B. Mode
C. Mean
D. None of the above

9. The most frequent observation in a data set is called
A. Mode
B. Median
C. Range
D. Mean

10. A measurement that corresponds to largest frequency in a set of data is called:
A. Mean
B. Median

C. Mode
D. Percentile

11. Mode of the series 0, 0, 0, 2, 2, 3, 3, 8, 10 is:
A. 0
B. 2
C. 3
D. No mode

12. A distribution with two modes is called:
A. Unimodal
B. Bimodal
C. Multimodal
D. Normal

13. When the values in a series are not of equal importance, we calculate the:
A. Arithmetic mean
B. Geometric mean
C. Weighted mean
D. Mode

14. Taking the relevant root of the product of all non-zero and positive values are called:
A. Arithmetic mean
B. Geometric mean
C. Harmonic mean
D. Combined mean

15. The ratio among the number of items and the sum of reciprocals of items is called:
A. Arithmetic mean
B. Geometric mean
C. Harmonic mean
D. Mode

Answers for SelfAssessment

1.	D	2.	D	3.	D	4.	A	5.	C
6.	C	7.	C	8.	A	9.	A	10.	C
11.	A	12.	B	13.	C	14.	C	15.	C

Review Questions

1. Give a brief description of the different measures of central tendency. Why is arithmetic mean so popular?
2. How would you explain the choice of arithmetic mean as the best measure of central tendency? Under what circumstances would you deem fit the use of median or mode?

3. Suppose the average amount of cash (in pocket, wallet, purse, etc.) possessed by 60 students attending a class is Rs 125. The median amount carried is Rs 90.
    a. What characteristics of the distribution of cash carried by the students can be explained. Why is mean larger than the median?
    b. Identify the process or population to which inferences based on these results might apply.
4. What is a statistical average? What are the desirable properties for an average to possess? Mention the different types of averages and state why arithmetic mean is most commonly used amongst them.
5. Given below is the distribution of profits (in ′000 rupees) earned by 94 per cent of the retail grocery shops in a city.

Profits	Number of Shops
0-10	0
10-20	5
20-30	14
30-40	27
40-50	48
50-60	68
60-70	83
70-80	91
80-90	94

6. The management of Doordarshan holds a preview of a new programme and asks viewers for their reaction. The following results by age groups, were obtained.

Age group	Liked the program	Dislike the program
Under 20	140	60
20-39	75	50
40-59	50	50
60 and above	40	20

7. The following are the profit figures earned by 50 companies in the country

Profit (Rs in lakh)	Number of Companies
10 or less	4
20 or less	10
30 or less	30
40 or less	40

50 or less	47
60 or less	50

Calculate

a. the median, and
b. the range of profit earned by the middle 80 per cent of the companies. Also verify your results by graphical method.

8. Write a short criticism of the following statement: 'Median is more representative than mean because it is relatively less affected by extreme values'.

Further Readings

- Pagano, R. (2004). Understanding Statistics in the Behavioural Sciences (7th edition). Pacific grove, ca: brooks/cole publishing co.
- Guilford, J.P...(1956). Fundamental Statistics in Psychology and Education. Mcgraw-hill book company. NY

Web Links

- Black, Thomas R. 1999. Doing Quantitative Research in the SocialSciences.AnIntegratedApproachto Research Design, Measurement and Statistics
- Nachmias,David and ChavaNachmias 1981.Research Methods in Social Sciences.St. Martin Press:New York.

Unit 09: Dispersion

CONTENTS

Objectives

- Describethe main properties limitation and uses of the range, quartile deviation, average deviation and standard deviation; and
- Explainvariance and coefficient of variance.

Introduction

Just as central tendency can be measured by a number in the form of an average, the amount of variation (dispersion, spread, or scatter) among the values in the data set can also be measured. The measures of central tendency describe that the major part of values in the data set appears to concentrate (cluster) around a central value called average with the remaining values scattered (spread or distributed) on either side of that value. But these measures do not reveal how these values are dispersed (spread or scatter) on each side of the central value. The dispersion of values is indicated by the extent to which these values tend to spread over an interval rather than cluster closely around an average.

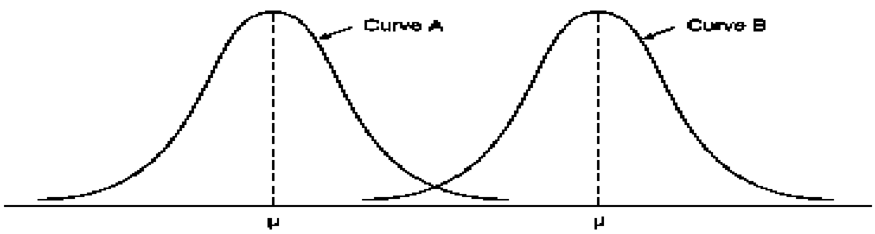
The statistical techniques to measure such dispersion are of two types:

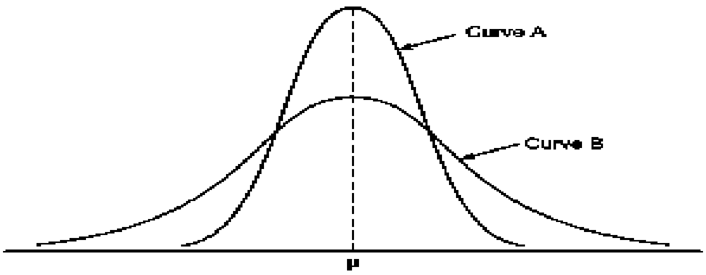
(a) Techniques that are used to measure the extent of variation or the deviation (also called degree of variation) of each value in the data set from a measure of central tendency usually the mean or median. Such statistical techniques are called measures of dispersion (or variation).

(b) Techniques that are used to measure the direction (away from uniformity or symmetry) of variation in the distribution of values in the data set. Such statistical techniques are called measures of skewness.

To measure the dispersion, understand it, and identify its causes is very important in statistical inference (estimation of parameter, hypothesis testing, forecasting, and so on). A small dispersion among values in the data set indicates that data are cluster closely around the mean. The mean is therefore considered representative of the data, i.e. mean is a reliable average. Conversely, a large dispersion among values in the data set indicates that the mean is not reliable, i.e. it is not representative of the data.

Fig. 1: Symmetrical Distributions with Unequal Mean and Equal Standard Deviation & Symmetrical Distributions with Equal Mean and Unequal Standard Deviation.





The symmetrical distribution of values in two or more sets of data may have same variation but differ greatly in terms of A.M. On the other hand, two or more sets of data may have the same A.M. values but differ in variation.

Meaning of Dispersion

The word dispersion is used to denote the degree of heterogeneity in the data. It is an important characteristic indicating the extent to which observations vary amongst themselves. The dispersion of a given set of observations will be zero when all of them-are equal (as in Set B given above). The wider the discrepancy from one observation to another, the larger would be the dispersion. (Thus dispersion in Set A should be larger than that in Set C.) A measure of dispersion is designed to state numerically the extent to which individual observations vary on the average.

According to Spiegel, the degree to which numerical data tend to spread about an average value is called the variation or dispersion of data. Actually, there are two basic kinds of a measure of dispersion (i) Absolute measures and (ii) Relative measures. The absolute measures of dispersion are used to measure the variability of a given data expressed in the same unit, while the relative measures are used to compare the variability of two or more sets of observations. Following are the different measures of dispersion

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard Deviation and Variance

9.1 Significance of Measuring Dispersion (Variation)

Following are some of the purposes for which measures of variation are needed.

1. *Test the reliability of an average*: Measures of variation are used to test to what extent andaverage represents the characteristic of a data set. If the variation is small, that is, extent of dispersion or scatter is less on each side of an average, then it indicates high uniformity of values in the distribution and the average represents an individual value in the data set. On the other hand, if the variation is large, then it indicates a lower degree of uniformity in values in the data set, and the average may be unreliable. No variation indicates perfect uniformity and, therefore, values in the data set are identical.

2. *Control the variability*: Measuring of variation helps to identify the nature and causes of variation. Such information is useful in controlling the variations. According to Spurr and Bonini, 'In matters of health, variations, in body temperature, pulse beat and blood pressure are the basic guides to diagnosis. Prescribed treatment is designed to control their variation. In industrial production efficient operation requires control of quality variation, the causes of which are sought through inspection and quality control programmes.' In social science, the measurement of 'inequality' of distribution of income and wealth requires the measurement of variability.

3. Compare two or more sets of data with respect to their variability: Measures of variation help in the comparison of the spread in two or more sets of data with respect to their uniformity or consistency. For example, (i) the measurement of variation in share prices and their comparison with respect to different companies over a period of time requires the measurement of variation, (ii) the measurement of variation in the length of stay of patients in a hospital every month may be used to set staffing levels, number of beds, number of doctors, and other trained staff, patient admission rates, and so on.

4. Facilitate the use of other statistical techniques: Measures of variation facilitate the use of other statistical techniques such as correlation and regression analysis, hypothesis testing, forecasting, quality control, and so on.

Characteristics for An Ideal Measure of Dispersion.

The characteristics, for an ideal measure of dispersion arc the same as those for all ideal measure of central tendency, viz.

1. It should he rigidly defined.
2. It should be easy to calculate and easy to understand.
3. It should be based on all the observations.
4. It should be amenable to further mathematical treatment.
5. It should be affected as little as possible by fluctuations of sampling.

Range

The range is the simplest measure of dispersion and is based on the location of the largest and the smallest values in the data. Thus, the range is defined to be the difference between the largest and lowest observed values in a data set. In other words, it is the length of an interval which covers the highest and lowest observed values in a data set and thus measures the dispersion or spread within the interval in the most direct possible way.

Range (R) = Highest value of an observation – Lowest value of an observation = H – L

For example,if the smallest value of an observation in the data set is 160 and largest value is 250, then the range is 250 – 160 = 90.

For grouped frequency distributions of values in the data set, the range is the difference between the upper-class limit of the last class and the lower-class limit of first class. In this case the range obtained may be higher than as compared to ungrouped data because of the fact that the class limits are extended slightly beyond the extreme values in the data set.

Example 1: Find the range of the distribution 6, 8, 2, 10, 15, 5, 1, 13.

Solution: For the given distribution, the maximum value of variable is 15 and the minimum value of variable is 1. Hence range = 15 -1 = 14.

It is intuitive that, because of central tendency, if one selects a small sample, observations are more likely to be around its mode than away from it. Less 4 likely or extreme values will be included in the sample when its size is large. This, in other words, implies that range will increase with increase in sample size. Also, it is known that in repeated sampling with same sample size, range varies considerably making it a less suitable measure for comparisons. However, range is a measure which is easy to understand and can be computed quickly.

Merits and Demerits of Range

Merits of Range

1. It is the simplest to understand;

2. It can be visually obtained since one can detect the largest and the smallest observations easily and can take the difference without involving much calculations; and

3. Though it is crude, it has useful applications in areas like order statistics and statistical quality control.

9.2 Demerits of Range

1. It utilizes only the maximum and the minimum values of variable in the series and gives no importance to other observations;

2. It is affected by fluctuations of sampling;

3. It is not very suitable for algebraic treatment;

4. If a single value lower than the minimum or higher than the maximum is added or if the maximum or minimum value is deleted range is seriously affected; and

5. Range is the measure having unit of the variable and is not a pure number. That's why sometimes coefficient of range is calculated by

Coefficient of Range = $\frac{Xmax-Xmin}{Xmax+Xmin}$

Interquartile Range or Deviation

The limitations or disadvantages of the range can partially be overcome by using another measure of variation which measures the spread over the middle half of the values in the data set so as to minimize the influence of outliers (extreme values) in the calculation of range. Since a large number of values in the data set lie in the central part of the frequency distribution, therefore it is necessary to study the Interquartile Range (also called midspread). To compute this value, the entire data set is divided into four parts each of which contains 25 per cent of the observed values. The quartiles are the highest values in each of these four parts. The interquartile range is a measure of dispersion or spread of values in the data set between the third quartile, Q3 and the first quartile, Q1. In other words, the interquartile range or deviation (IQR) is the range for the middle 50 per cent of the data.

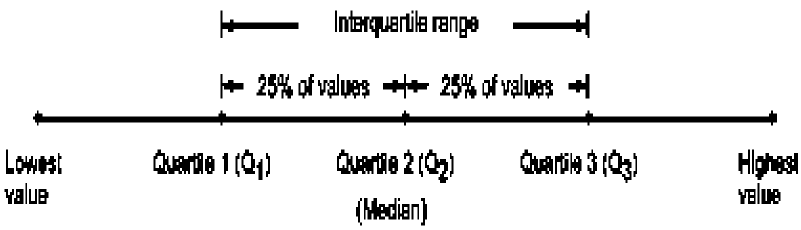
Interquartile range (IQR) = Q3 – Q1

Half the distance between Q1 and Q3 is called the semi-interquartile range or the quartile deviation (QD).

Quartile déviation (QD) = (Q3-Q1)/2

The median is not necessarily midway between Q1 and Q3, although this will be so for a symmetrical distribution. The median and quartiles divide the data into equal numbers of values but do not necessarily divide the data into equally wide intervals.

As shown above the quartile deviation measures the average range of 25 per cent of the values in the data set. It represents the spread of all observed values because its value is computed by taking an average of the middle 50 per cent of the observed values rather than of the 25 per cent part of the values in the data set.



In a non-symmetrical distribution, the two quartiles Q1 and Q3 are at equal distance from the median, that is, Median – Q1 = Q3 – Median. Thus, Median ± Quartile Deviation covers exactly 50 per cent of the observed values in the data set.

A smaller value of quartile deviation indicates high uniformity or less variation among the middle 50 per cent observed values around the median value. On the other hand, a high value of quartile deviation indicates large variation among the middle 50 per cent observed values.

Example 2:Calculate quartile deviation and its coefficient from the following data:

Weights (in kg)	Frequency
60	1

61	3
62	5
63	7
65	10
70	3
75	1
80	1

Solution:

Weights (in kg)	Frequency	Cumulative Frequency
60	1	1
61	3	4
62	5	9
63	7	16
65	10	26
70	3	29
75	1	30
80	1	31=n

Q, = Size of ¼ (n+1) or 8th observation

= 62 Kgs (because 8th observation falls in this category)

Q3 = Size of 3/4(n + 1) or 24th observation

= 65 Kgs. (because 24th observation falls in this category)

Therefore, Quartile Deviation= (Q3-Q1)/2

= (65-62)/2

= 1.5 Kgs.

Coefficient of Quartile Deviation= (Q3-Q1)/(Q3+Q1)

= (65-62)/(65+62)

=3/127

=0.024.

Example 3: Calculate semi-interquartile range and its coefficient from the following data

Marks	Frequency
0-10	11

10-20	18
20-30	25
30-40	28
40-50	30
50-60	33
60-70	22
70-80	15
80-90	22

Solution: To compute quartile deviation, we need the values of the first quartile and the third quartile which can be obtained from the following table:

Marks	Frequency	Cumulative Frequency
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	22	204

Q1 has N/4 observations i.e., 204/4 = 51 observations, below it. So Q1 lies in the 20-30 class.

$$Q_1 = l + \left(\frac{\frac{N}{4} - Cf}{f}\right) \times i$$

Where,

l = lower limit of quartile class

c = cumulated frequency preceding the quartile class

f = simple frequency of the quartile class

i = class-interval of quartile class

$$Q_1 = 20 + \left(\frac{51 - 29}{25}\right) \times 10$$

=28.8

Q1 has 3N/4 observations i.e., (204/4)*3 = 153observations, below it. So Q3 lies in the 60-70 class.

$$Q_3 = l + \left(\frac{\frac{3N}{4} - Cf}{f}\right) \times i$$

$$Q_3 = 60 + \left(\frac{153 - 145}{22}\right) \times 10$$

= 63.64

Semi-inter Quartile Range or Quartile Deviation is given by

Q.D= (Q3-Q1)/2

= (63.64-28.8)/2

=34.82/2

= 17.42 marks

The relative measure corresponding to quartile deviation, called the coefficient of quartile deviation, it calculated as follows:

Coefficient of Q .D.= (Q3-Q1)/(Q3+Q1)

= (63.64-28.8)/(63.64+28.8)

= 0.37

Advantages and Disadvantages of Quartile Deviation

The major advantages and disadvantages of quartile deviation are summarized as follows:

Advantages

i. It is not difficult to calculate but can only be used to evaluate variation among observed values within the middle of the data set. Its value is not affected by the extreme (highest and lowest) values in the data set.

ii. It is an appropriate measure of variation for a data set summarized in open-end class intervals.

iii. Since it is a positional measure of variation, therefore it is useful in case of erratic or highly skewed distributions, where other measures of variation get affected by extreme values in the data set.

Disadvantages

i. The value of Q.D. is based on the middle 50 per cent observed values in the data set, therefore it cannot be considered as a good measure of variation as it is not based on all the observations.

ii. The value of Q.D. is very much affected by sampling fluctuations.

iii. The Q.D. has no relationship to any particular value or an average in the data set for measuring the variation. Its value is not affected by the distribution of the individual values within the interval of the middle 50 per cent observed values.

9.3 Mean Deviation

Since two measures of variation, range and quartile deviation, discussed earlier do not show how values in a data set are scattered about a central value or disperse themselves throughout the range, therefore it is quite reasonable to measure the variation as a degree (amount) to which values within a data set deviate from either mean or median.

The mean of deviations of individual values in the data set from their actual mean is always zero so such a measure (zero) would be useless as an indicator of variation. This problem can be solved in two ways:

1. Ignore the signs of the deviations by taking their absolute value, or

2. Square the deviations because the square of a negative number is positive

Since the absolute difference between a value xi of an observation from A.M. is always a positive number, whether it is less than or more than the A.M., therefore we take the absolute value of each such deviation from the A.M. (or median). Taking the average of these deviations from the A.M., we get a measure of variation called the mean absolute deviation (MAD). In general, the mean absolute deviation is given by

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|x-\mu|\,, for a population$$

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|x-\bar{x}|\,, for a sample$$

Where | | indicates the absolute value. That is, the signs of deviations from the mean are disregarded.

For a grouped frequency distribution, MAD is given by

$$MAD = \frac{\sum_{i=1}^{n} f_1|x-\bar{x}|}{\Sigma f\text{i}}$$

It indicates that the MAD provides a useful method of comparing the relative tendency of values in the distribution to scatter around a central value or to disperse themselves throughout the range.

While calculating the mean absolute deviation, the median is also considered for computing because the sum of the absolute values of the deviations from the median is smaller than that from any other value. However, in general, arithmetic mean is used for this purpose.

If a frequency distribution is symmetrical, then A.M. and median values concide and the same MAD value is obtained. In such a case x ± MAD provides a range in which 57.5 per cent of the observations are included. Even if the frequency distribution is moderately skewed, the interval x ± MAD includes the same percentage of observations. This shows that more than half of the observations are scattered within one unit of the MAD around the arithmetic mean.

The MAD is useful in situations where occasional large and erratic deviations are likely to occur. The standard deviation, which uses the squares of these large deviations, tends to over-emphasize them.

Example 4: The number of patients seen in the emergency ward of a hospital for a sample of 5 days in the last month were: 153, 147, 151, 156 and 153. Determine the mean deviation and interpret.

Solution: The mean number of patients is, x = (153 + 147 + 151 + 156 + 153)/5 = 152. Below are the details of the calculations of MAD

Number of Patients (X)	$x - \bar{x}$	Absolute DDeviation $\lvert x-\bar{x}\rvert$ $x - \bar{x}\rvert$
153	153-152=1	1
147	147-152=-5	5
151	151-152=-1	1
156	156-152=4	4
153	153-152=1	1
		12

MAD = $\frac{1}{N}\sum_{i=1}^{N}|x-\mu|$

=12/5= 2.4=3 patients (approx.)

The mean absolute deviation is 3 patients per day. The number of patients deviate on the average by 3 patients from the mean of 152 patients per day.

9.4 Standard Deviation

The term standard deviation was first used in writing by Karl Pearson in 1894. The standard deviation of population is denoted by 'σ' (Greek letter sigma) and that for a sample is 's'. A useful property of SD is that unlike variance it is expressed in the same unit as the data. This is most widely used method of variability. The standard deviation indicates the average of distance of all the scores around the mean. It is the positive square root of the mean of squared deviations of all the scores from the mean. It is the positive square root of variance. It is also called as 'root mean square deviation.

Standard deviation shows how much variation there is, from the mean. SD is calculated from the mean only. If standard deviation is low, it means that the data is close to the mean. A high standard deviation indicates that the data is spread out over a large range of values. Standard deviation may serve as a measure of uncertainty. If you want to test the theory or in other word, want to decide whether measurements agree with a theoretical prediction, the standard deviation provides the information. If the difference between mean and standard deviation is very large then the theory being tested probably needs to be revised. The mean with smaller standard deviation is more reliable than mean with large standard deviation. A smaller SD shows the homogeneity of the data. The value of standard deviation is based on every observation in a set of data. It is the only measure of dispersion capable of algebraic treatment therefore, SD is used in further statistical analysis.

A. Ungrouped Data

Population Standard Deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\Sigma(x - \mu^2)} = \sqrt{\frac{1}{N}\Sigma x^2 - \mu^2}$

$= \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{N}\right)^2}$

Sample standard deviation, $s = \sqrt{\frac{\Sigma x^2}{n-1} - \frac{nx\hat{}2}{n-1}}$

B. Grouped Data

Population standard deviation, $\sigma = \sqrt{\frac{\Sigma f d^2}{n} - \left(\frac{\Sigma f d}{N}\right)^2}$

Where f = frequency of each class interval

N = ∑ f = total number of observations (or elements) in the population

h = width of class interval

d = (m− A), where A is any constant (also called assumed A.M.)

m = mid-value of each class interval

Sample standard deviation, $s = \sqrt{s^2}$

$= \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n-1}}$

$= \sqrt{\frac{\Sigma f x^2}{n-1} - \frac{(\Sigma f x)\hat{}2}{n-1}}$

Example 4: The wholesale prices of a commodity for seven consecutive days in a month is as follows:

Days	Commodity price/Quintal
1	240
2	260
3	270
4	245

5	255
6	286
7	264

Calculate the variance and standard deviation.

Solutions: The computations for variance and standard deviation are shown in below table 4.1

Observations	standard $x-\bar{x}$	'e shown i $(x-\bar{x})^2$
240	-20	400
260	0	0
270	10	100
245	-15	225
255	-5	25
286	26	676
26	4	16
1820		1442

$\bar{x} = \sum x/N = 1820/7 = 260$

Variance $\sigma^2 = \sum(x-\bar{x}) = \frac{1442}{7}$

Standard deviation $\sigma = \sqrt{\sigma^2} = \sqrt{206} = 14.352$

Example 5: A study of 100 engineering companies gives the following information

Profit (Rs in crore)	Number of companies
0-10	8
10-20	12
20-30	20
30-40	30
40-50	20
50-60	10

Calculate the standard deviation of the profit earned.

Solution: Let assumed mean, A be 35 and the value of h be 10. Calculations for standard deviation are shown in Table 5.1

Table 5.1 Calculations of Standard Deviation

Profit (Rs in crorr)	Mid-Value (m)	d=(m-a)/h	f	fd	fd^2

0-10	5	-3	8	-24	72
10-20	15	-2	12	-24	48
20-30	25	-1	20	-20	20
30-40	35=A	0	30	0	0
40-50	45	1	20	20	20
50-60	55	2	10	20	40
				-28	200

Standard deviation, σ = $\sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2} * h$

= $\sqrt{\frac{200}{100} - \left(\frac{-28}{100}\right)^2} * 10$

= √2-0.0.78

=13.683

Merits and Demerits of Standard Deviation

Merits

1) It is widely used because it is the best measure of variation by virtue of its mathematical characteristics.

2) It is based on all the observations of the data.

3) It gives an accurate estimate of population parameter when compared with other measures of variation.

4) SD is least affected by sample fluctuations

5) It is also possible to calculate combined SD that is not possible with other measures.

6) Further statistics can be applied on the basis of SD like, correlation, regression, tests of significance, etc.

7) Coefficient of variation is based on mean and SD. It is the most appropriate method to compare variability of two or more distributions.
The limitations of SD are as follows:

1) While calculating standard deviation more weight is given to extreme values and less to those, near the means. When we calculate SD, we take deviation from mean (X-M) and square these obtained deviations. Therefore, large deviations, when squared are proportionally more than small deviations. For example, the deviations 2 and 10 are in the ratio of 1:5 but their square 4 and l00 are in the ratio 1:25.

2) It is difficult to compute as compared to other measures of dispersion.

Variance

The term variance was used to describe the square of the standard deviation by R.A. Fisher in 1913. The concept of variance is of great importance in advanced work where it is possible to split the total into several parts, each attributable to one of the factors causing variations in their original series. Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean.

The variance (s^2) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it is the mean of the squared deviation of scores.Variance is expressed as $V = SD^2$.

The variance and the closely related standard deviation are measures that indicate how the scores are spread out in a distribution. In other words, they are measures of variability. The variance is computed as the average squared deviation of each number from its mean.

Calculating the variance is an important part of many statistical applications and analysis. It is a good absolute measure of variability and is useful in computation of Analysis of Variance (ANOVA) to find out the significance of differences between sample means

Merits and Demerits of Variance

The main merits of variance are listed as follows:

1) It is rigidly defined and based on all observations.

2) It is amenable to further algebraic treatment.

3) It is not affected by sampling fluctuations.

4) It is less erratic.

The main demerits of variance are listed as follows:

1) It is difficult to understand and calculate.

2) It gives greater weight to extreme values

Coefficient of Variation

Standard deviation is an absolute measure of variation and expresses variation in the same unit of measurement as the arithmetic mean or the original data. A relative measure called the coefficient of variation (CV), developed by Karl Pearson is very useful measure for (i) comparing two or more data sets expressed in different units of measurement (ii) comparing data sets that are in same unit of measurement but the mean values of data sets in a comparable field are widely dissimilar (such as mean wages received per month by the top management personnel and labor class personnel of a large organization).

Thus, in view of this limitation we need to convert absolute measure of variation, that is, S.D. into a relative measure, which can be helpful in comparing the variability of two or more sets of data. The new measure, coefficient of variation (CV) measures the standard deviation relative to the mean in percentages. In other words, CV indicates how large the standard deviation is in relation to the mean and is computed as follows:

Coefficient of variation (CV)= (Standard Deviation/Mean)*100= $\sigma/\bar{x} * 100$

Multiplying by 100 converts the decimal to a percent. The set of data for which the coefficient of variation is low is said to be more uniform (consistent) or more homogeneous (stable)

Example 6: The weekly sales of two products A and B were recorded as given below:

Product A	*Product B*
59	*150*
75	*200*
27	*125*
63	*310*
27	*330*
28	*250*
56	*225*

Find out which of the two shows greater fluctuation in sales.

Solution: For comparing the fluctuation in sales of two products we will prefer to calculate coefficient of variation for both the products.

Product A: Let A = 56 be the assumed mean of sales for product A.

Table 6.1 Calculations of the Mean and Standard Deviation

Sales (x)	Frequency (f)	d=x-A	fd	fd^2
27	2	-29	-58	1682
28	1	-28	-28	784
56	1	0	0	0
59	1	3	3	9
63	1	7	7	49
75	1	19	19	361
	7		-57	2885

$\bar{x} = A + \frac{\Sigma fd}{\Sigma f}$= 56-57/7=47.86

$s^2A = \frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2$ = 2885/7-(-57/7)2

=412.14-66.30=345.84

Then, CV(A)= $s_A / \bar{x}$

= 18.59/47.86*100= 38.84 percent

Summary

To summarize, the measures of central tendency are not sufficient to describe data. Thus, to describe distribution adequately, we must provide a measure of variability or dispersion. The measures of variability are summary figures that express quantitatively, the extent to which, scores in a distribution scatter around or cluster together. The measures of variability are range, quartile deviation, average deviation, standard deviation and variance. Range is easy to calculate and useful for preliminary work. But this is based on extreme items only, and does not consider intermediate scores. Thus, it is not useful as a descriptive measure. Quartile deviation is related to the median in its properties. It takes into consideration the number of scores lying above or below the outer quartile point but not to their magnitude. This is useful with open ended distribution. The average deviation takes into account the exact position of each score in the distribution. The means deviation gives a more precise measure of the spread of scores but is mathematically inadequate. The average deviation is less affected by sampling fluctuation. The standard deviation is the most stable measure of variability. Standard deviation shows how much the score departs from the mean. It is expressed in original scores unit. Thus, it is most widely used measure of variability in descriptive statistics. The variance (s 2) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it the mean of the squared deviation of scores. The relative measure corresponding to SD is the coefficient of variation. It is a useful measure of relative variation.

Keywords

Average Deviation or Mean Deviation: A measure of dispersion that gives the average difference (ignoring plus and minus sign) between each item and the mean.

Dispersion: The spread or variability is a set of data.

Deviation: The difference between raw score and mean.

Quartile Deviation: A measure of dispersion that can be obtained by dividing the difference between Q3 and Q1 by two.

Range: Difference between the largest and smallest value in a data.

Standard deviation: The square root of the variance in a series.

Variance: Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean.

SelfAssessment

1. Which of the following are characteristics of a good measure of dispersion?

A. It should be easy to calculate
B. It should be based on all the observations within a series
C. It should not be affected by the fluctuations within the sampling
D. All of the above

2. The scatter within a distribution that is high on each side indicates _________.

A. High uniformity of data
B. Outliers of data
C. Low uniformity of data
D. None of the above

3. Which of the following is false regarding Dispersion?

A. Reveals how items are spread out on either side of the centre
B. Indicates high or low uniformity of the items
C. It serve to locate the distribution
D. Difference or variation among the values

4. The measures used to calculate the variation present among the observations in the unit of the variable is called:

A. Relative measures of dispersion
B. Coefficient of skewness
C. Absolute measures of dispersion
D. Coefficient of variation

5. The measures of dispersion can never be:

A. Positive
B. Zero
C. Negative
D. Equal to 2

6. The measure of dispersion which uses only two observations is called:

A. Range
B. Quartile deviation
C. Mean deviation
D. Standard deviation

7. Given below the four sets of observations. Which set has the minimum variation?

A. 46, 48, 50, 52, 54
B. 30, 40, 50, 60, 70
C. 40, 50, 60, 70, 80
D. 48, 49, 50, 51, 52

8. An example of the application of range in a real-world scenario would be __________.
A. Fluctuation in share prices
B. Weather forecasts
C. Quality control
D. All of the above

9. Which measure of dispersion can be computed in case of open-end classes?
A. Standard deviation
B. Range
C. Quartile deviation
D. Coefficient of variation

10. Mean deviation computed from a set of data is always:
A. Negative
B. Equal to standard deviation
C. More than standard deviation
D. Less than standard deviation

11. The square of standard deviation is ________.
A. Square deviation
B. Mean square deviation
C. Variance
D. None of the above

12. Which of the following is an absolute measure of dispersion?
A. Coefficient of variation
B. Coefficient of dispersion
C. Standard deviation
D. Coefficient of skewness

13. Which of the following measures of dispersion is expressed in the same units as the units of observation?
A. Variance
B. Standard deviation
C. Coefficient of variation
D. Coefficient of standard deviation

14. Which of the following is a unit free quantity?
A. Range
B. Standard deviation
C. Coefficient of variation
D. Arithmetic mean

15. If the dispersion is small, the standard deviation is:
A. Large
B. Zero
C. Small
D. Negative

Answers for Self Assessment

1.	D	2.	B	3.	C	4.	C	5.	C
6.	A	7.	D	8.	C	9.	D	10.	C
11.	C	12.	C	13.	B	14.	C	15.	C

Review Questions

1. Explain the term variation. What does a measure of variation serve? In the light of these, comment on some of the well-known measures of variation.
2. What are the requisites of a good measure of variation?
3. Explain how measures of central tendency and measures of variation are complementary to each other in the context of analysis of data.
4. Distinguish between absolute and relative measures of variation. Give a broad classification of the measures of variation.
5. What do you understand by 'coefficient of variation'? Discuss its importance in business problems.
6. When is the variance equal to the standard deviation? Under what circumstances can variance be less than the standard deviation? Explain.
7. What advantages are associated with variance and standard deviation relative to range as the measure of variability?
8. Suppose you read a published statement that the average amount of food consumption in this country is adequate; the overall conclusion based upon the statement is that everyone is properly fed. Criticize the conclusion in terms of the concept of variability as it relates to the use of averages.
9. A retailer uses two different formulas for predicting monthly sales. The first formula has an average miss of 700 records, and a standard deviation of 35 records. The second formula has an average miss of 300 records, and a standard deviation of 16. Which formula is relatively less accurate?
10. Find the average deviation from mean for the following distribution:

Quantity demand	60	61	62	63	64	65	66	67	68
Frequency	2	0	15	29	25	12	10	4	3

11. The share prices of a company in Mumbai and Kolkata markets during the last ten months are recorded below:

Month	Mumbai	Kolkata

January	105	108
February	120	117
March	115	120
April	118	130
May	130	100
June	127	125
July	109	125
August	110	120
September	104	110
October	112	135

Determine the arithmetic mean and standard deviation of prices of shares. In which market are the share prices more stable?

12. Blood serum cholesterol levels of 10 persons are as under:

240 260 290 245 255 288 272 263 277 250

Calculate the standard deviation with the help of assumed mean.

Further Readings

- Pagano, R. (2004). Understanding Statistics in the Behavioural Sciences (7th edition). Pacific grove, ca: brooks/cole publishing co.
- Guilford, J.P...(1956). Fundamental Statistics in Psychology and Education. Mcgraw-hill book company. NY

Web Links

- Garrett, H.E. (1981), Statistics in Psychology and Education, (Tenth edition), Bombay, Vakils Feffer and Simons Ltd.
- McBride, Dawn M. (2018). The Process of Statistical Analysis in Psychology. Sage. USA
- Minium, E.W., King, B.M. & Bear. G (2001). Statistical Reasoning in Psychology and Education (3rd edition), Singapore, John Wiley & Sons, Inc.
- Mohanty, B. & Misra, Santa (2016). Statistics for Behavioural and Social Sciences. Sage. New Delhi.

Unit 10: Skewness

CONTENTS

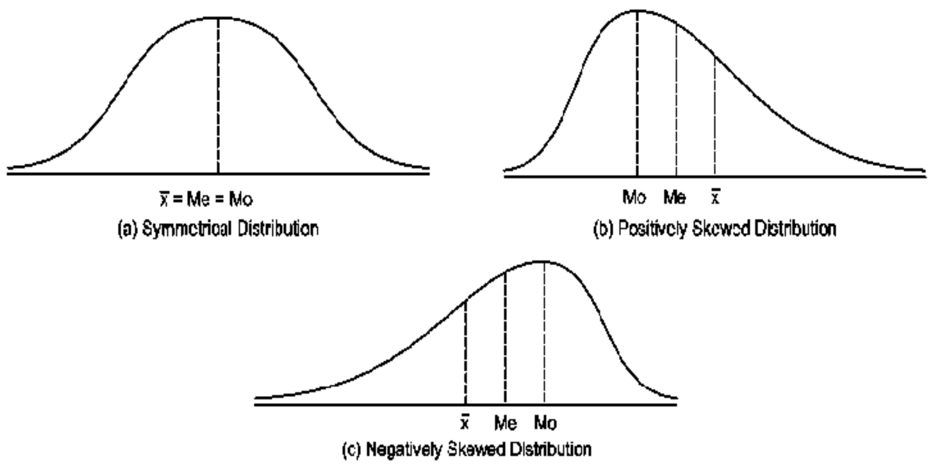
Objectives

- Understand the concept of skewness and Kurtosis and its importance
- Practice the computation of coefficient of skewness and kurtosis

Introduction

In Chapter 5 we discussed measures of variation (or dispersion) to describe the spread of individual values in a data set around a central value. Such descriptive analysis of a frequency distribution remains incomplete until we measure the degree to which these individual values in the data set deviate from symmetry on both sides of the central value and the direction in which these are distributed. This analysis is important due to the fact that data sets may have the same mean and standard deviation but the frequency curves may differ in their shape. A frequency distribution of the set of values that is not 'symmetrical (normal)' is called asymmetrical or skewed. In a skewed distribution, extreme values in a data set move towards one side or tail of a distribution, thereby lengthening that tail. When extreme values move towards the upper or right tail, the distribution is positively skewed. When such values move towards the lower or left tail, the distribution is negatively skewed. As discussed, the mean, median, and mode are affected by the highvalued observations in any data set. Among these measures of central tendency, the mean value gets affected largely due to the presence of high-valued observations in one tail of a distribution. The mean value shifted substantially in the direction of high-values. The mode value is unaffected, while the median value, which is affected by the numbers but not the values of such observations, is also shifted in the direction of high-valued observations, but not as far as the mean. The median value changes about 2/3 as far as the mean value in the direction of high-valued observations (called extremes).

For a positively skewed distribution A.M. > Median > Mode, and for a negatively skewed distribution A.M. < Median < Mode. The relationship between these measures of central tendency is used to develop a measure of skewness called the coefficient of skewness to understand the degree to which these three measures differ.



From the above discussion two points of difference emerge between variation and skewness:

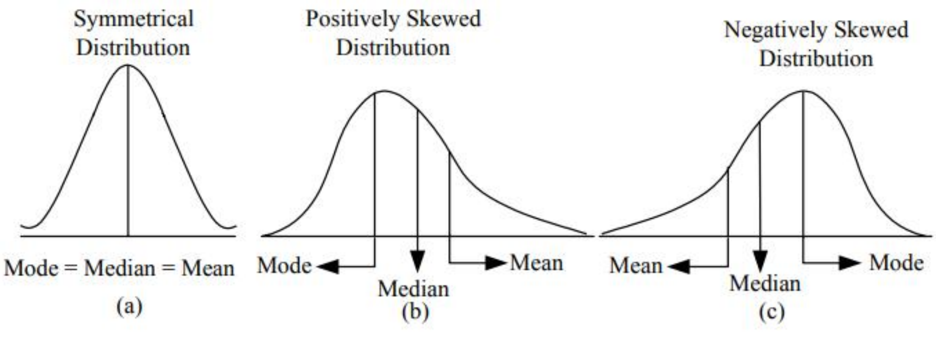
i. Variation indicates the amount of spread or dispersion of individual values in a data set around a central value, while skewness indicates the direction of dispersion, that is, away from symmetry.

ii. Variation is helpful in finding out the extent of variation among individual values in a data set, while skewness gives an understanding about the concentration of higher or lower values around the mean value.

Skewness

The measure of skewness tells us the direction of dispersion about the centre of the distribution. Measures of central tendency indicate only the single representative figure of the distribution while measures of variation, indicate only the spread of the individual values around the means. They do not give any idea of the direction of spread. Two distributions may have the same mean and variation but may differ widely in the shape of their distribution. A distribution is often found skewed on either side of its average, which is termed as asymmetrical distribution. Thus, skewness refers to the lack of symmetry in distribution. Symmetry signifies that the value of variables is equidistant from the average on both sides. In other words, a balanced pattern of a distribution is called symmetrical distribution; where as unbalanced pattern of distribution is called asymmetrical distribution.

A simple method of finding the direction of skewness is to consider the tails of a frequency polygon. The concept of skewness will be clear from the following three figures showing symmetrical, positively skewed and negatively skewed distributions.



Carefully observe the figures presented above and try to understand the following rules governing them.

It is clear from Figure (a) that the data are symmetrical when the spread of the frequencies is the same on both sides of the middle point of the frequency polygon. In this case the value of mean, median, and mode coincide i.e., Mean = Median = Mode.

When the distribution is not symmetrical, it is said to be a skewed distribution. Such a distribution could be either positively skewed or negatively skewed. In Figure (b), when there is a longer tail towards the righthand side of the centre of distribution, the skewness is said to be Positively Skewed.

In such a situation, Mean > Median > Mode. In Figure (c), when there is a longer tail towards the lefthand side of the centre, then the skewness is said to be Negatively Skewed. In such a case, Mean < Median < Mode.

It is seen that, in positively skewed distribution, dispersal of individual observations is greater towards the right of the central value. Where as in a negatively skewed distribution, a greater dispersal of individual observations is towards the left of the central value. We can say, therefore, the concept of Skewness not only refers to lack of symmetry in a distribution but also indicates the magnitude as well as the direction of skewness in a distribution. The relationship of mean, median and mode in measuring the degree of skewness is that, for a moderately symmetrical distribution the interval between the mean and the median is approximately 1/3rd of the interval between the mean and mode.

10.1 Measures of Skewness

Measures of skewness help us to know to what degree and in which direction (positive or negative) the frequency distribution has a departure from symmetry. Although positive or negative skewness can be detected graphically depending on whether the right tail or the left tail is longer but, we don't get idea of the magnitude. Besides, borderline cases between symmetry and asymmetry may be difficult to detect graphically. Hence some statistical measures are required to find the magnitude of lack of symmetry. A good measure of skewness should possess three criteria:

1. It should be a unit free number so that the shapes of different distributions, so far as symmetry is concerned, can be compared even if the unit of the underlying variables are different;

2. If the distribution is symmetric, the value of the measure should be zero. Similarly, the measure should give positive or negative values according as the distribution has positive or negative skewness respectively; and

3. As we move from extreme negative skewness to extreme positive skewness, the value of the measure should vary accordingly.

Measures of skewness can be both absolute as well as relative. Since in a symmetrical distribution mean, median and mode are identical more the mean moves away from the mode, the larger the asymmetry or skewness. An absolute measure of skewness cannot be used for purposes of comparison because of the same amount of skewness has different meanings in distribution with small variation and in distribution with large variation.

Absolute Measures of Skewness

The degree of skewness in a distribution can be measured both in the absolute and relative sense. For an asymmetrical distribution, the distance between mean and mode may be used to measure the degree of skewness because the mean is equal to mode in a symmetrical distribution. Thus,

Absolute Sk = Mean –Mode

= Q3 + Q1 – 2 Median (if measured in terms ofquartiles).

For comparing to series, we do not calculate these absolute measures we calculate the relative measures which are called coefficient of skewness. Coefficient of skewness is pure numbers independent of units of measurements.

Relative Measures of Skewness

The relative measure of skewness is termed as Coefficient of Skewness, It is useful in making a comparison between the skewness in two or more sets of data. There are two important methods for measuring the coefficient of skewness. They are:

1) Karl Pearson's coefficient of skewness.

2) Bowley's coefficient of skewness.

Karl Pearson's coefficient of skewness

The measure suggested by Karl Pearson for measuring coefficient of skewness is given by:

$$Sk_p = \frac{Mean - Mode}{StandardDeviation} = \frac{\bar{x} - M_0}{\sigma}$$

Where Skp = Karl Pearson's coefficient of skewness.

Since a mode does not always exist uniquely in a distribution, therefore it is convenient to define this measure using median. For a moderately skewed distribution, the following relationship holds:

Mean – Mode =3(Mean – Median) or Mode = 3 Median – 2 Mean

When this value of mode is substituted in the eqn., we get

$$Sk_p == \frac{[illegible]\bar{x} - M_0)}{\sigma}$$

Theoretically, the value of Skp varies between ± 3. But for a moderately skewed distribution, value of Skp = ± 1. Karl Pearson's method of determining coefficient of skewness is particularly useful in open-end distributions.

Example 1: Data of rejected items during a production process is as follows:

No. of rejects (per operator)	21-25	26-30	31-35	36-40	41-45	46-50	51-55
No. of operators	5	15	28	42	15	12	3

Calculate the mean, standard deviation, and coefficient of skewness and comment on the results.

Solution: The calculations for mean, mode, and standard deviation are shown in Table 1

Table 1.1 Calculations for Mean, Mode and Standard Deviation

Class	Mid-Value	Frequency (f)	D=(m-A)/h	fd	fd^2
21-25	23	5	-3	-15	45
26-30	28	15	-2	-30	60
31-35	33	28	-1	-28	28
36-40	38	42	0	0	0
41-45	43	15	1	15	15
46-50	48	12	2	24	48
51-55	53	3	3	9	27
		N=120		-25	223

Let assumed mean, A = 38. Then

$$\bar{x} = A + \frac{\Sigma fd}{N} * h$$

=38-(25/120)*5

=36.96 rejects per operator

Standard deviation:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{n}\right)^2} * h$$

$= \sqrt{\frac{223}{120} - \left(\frac{-25}{120}\right)^2} * 5$

= 6.736 rejects per operator

By inspection, mode lies in the class 36–40. Thus

$M_0 = L + \frac{[f1-f0]}{[f1-f0]+[f1-f2]} * h$

$= 36 + \frac{[42-28]}{(2*42-28-15)} * 5$

=37.70

$$Sk_p = \frac{Mean - Mode}{StandardDeviation} = \frac{\bar{x} - M_0}{\sigma}$$

= (36.96-37.70)/6.73= -0.74/6.73= -0.109

Since the coefficient of skewness, Sk = – 0.109, the distribution is skewed to left (negatively skewed). Thus, the concentration of the rejects per operator is more on the lower values of the distribution to the extent of 10.9 per cent.

Example 2:From the following data on age of employees, calculate the coefficient of skewness and comment on the result

Age below	25	30	35	40	45	50	55
Number of employees	8	20	40	65	80	92	100

Solution: The data are given in a cumulative frequency distribution form. So to calculate the coefficient of skewness, convert this data into a simple frequency distribution as shown in Table 2.1

Table 2.1 Calculations for Coefficient of Skewness

Age	Mid-Value	Frequency (f)	D=(m-A)/h	fd	fd^2
20-25	22.5	8	-3	-24	72
25-30	27.5	12	-2	-24	48
30-35	32.5	20	-1	-20	20
35-40	37.5	25	0	0	0
40-45	42.5	1	1	15	15
45-50	47.5	12	2	24	48
50-55	52.5	8	3	24	72
		N=120		-25	275

Let assumed mean, A = 37.5. Then

$$\bar{x} = A + \frac{\Sigma fd}{N} * h$$

= 37.5-(5/100)*5

= 37.25

Mode value lies in the class interval 35–40. Thus,

$M_0 = L + \frac{[f1-f0]}{[f1-f0]+[f1-f2]} * h$

$= 35 + \frac{[25-20]}{(2*25-20-15)} * 5$

=36.67

Standard deviation:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{n}\right)^2} * h$$

$= \sqrt{\frac{275}{100} - \left(\frac{-5}{100}\right)^2} * 5$

= 8.29

Karl Pearson's coefficient of skewness:

$Sk_p = \frac{Mean-Mode}{StandardDeviation} = \frac{37.25-36.67}{8.29}$=0.07

The positive value of Skp indicates that the distribution is slightly positively skewed.

Bowley's Coefficients of Skewness

The method suggested by Prof. Bowley is based on the relative positions of the median and the quartiles in a distribution. If a distribution is symmetrical, then Q1 and Q3 would be at equal distances from the value of the median, that is,

Median – Q_1 = Q_3 – Median

OrQ_3 + Q_1 – 2 Median = 0 or Median= (Q_3+Q_1)/2

This shows that the value of median is the mean value of Q_1 and Q_3. Obviously in such a case the absolute value of the coefficient of skewness will be zero.

When a distribution is asymmetrical, quartiles are not at equal distance from the median. The distribution is positively skewed, if Q_1 – Me > Q_3 – Me, otherwise negatively skewed.

The absolute measure of skewness is converted into a relative measure for comparing distributions expressed in different units of measurement. For this, absolute measure is divided by the inter-quartile-range. That is,

Relative Sk_b= (Q_3 + Q_1 – 2 Median)/(Q_3-Q_1)

= (Q_3 – Median)-(Median – Q_1)/(Q_3 – Median)+(Median – Q_1)

In a distribution, if Med = Q1, then Skb= ±1, but of Med = Q3 then Skb = – 1. This shows that the value of Skb varies between ± 1 for moderately skewed distribution. This method of measuring skewness is quite useful in those cases where (i) mode is ill-defined and extreme observations are present in the data, (ii) the distribution has open-end classes. These two advantages of Bowley's coefficient of skewness indicate that it is not affected by extreme observations in the data set.

Notes: The values of Skb obtained by Karl Pearson's and Bowley's methods cannot be compared. On certain occasions it is possible that one of them gives a positive value while the other gives a negative value.

Example 3: The data on the profits (in Rs lakh) earned by 60 companies is as follows:

Profits	Below 10	10-20	20-30	30-40	40-50	50-above
No. of Companies	5	12	20	16	5	2

a. Obtain the limits of profits of the central 50 per cent companies
b. Calculate Bowley's coefficient of skewness.

Solution: (a) Calculations for different quartiles are shown in Table 3.1

Profits (Rs in lakh)	Frequency (f)	Cumulative Frequency (C.f)
Below 10	5	5
10-20	12	17= Q1 class
20-30	20	37
30-40	16	53=Q3 class
40-50	5	58
50-above	2	60

Q1 = size of (N/4)th observation = (60 ÷ 4)th = 15th observation. Thus Q1 lies in the class 10–20, and

$$Q_1 = l + \left(\frac{\frac{N}{4} - Cf}{f}\right) \times i$$

$$Q_1 = 10 + \left(\frac{15 - 5}{12}\right) \times 10$$

= 10+8.33= 18.33 lakh

Q3 = size of (3N/4)th observation = 45th observation. Thus Q3 lies in the class 30–40, and

$$Q_3 = l + \left(\frac{\frac{3N}{4} - Cf}{f}\right) \times i$$

$$Q_3 = 30 + \left(\frac{45 - 37}{16}\right) \times 10$$

= 30+5= 35 lakh

Hence the profit of central 50 per cent companies lies between Rs 35 lakhs and Rs 10.83 lakh

Coefficient of quartile deviation, Q.D.= (Q3-Q1)/(Q3+Q1)

= (35-18.33)/(35+18.33)= 0.313

(a) Median = size of (N/2)th observation = 30th observation. Thus Median lies in the class 20–30, and

Median = L+ (N/2-C)/f *h

= 20+(30-17)/(20)*10

= 26.5 lakh

Coefficient of skewness, Sk_b= (Q_3 + Q_1 – 2 Median)/(Q_3-Q_1)

= (35+18.33-2(26.5))/36-18.33

= 0.02

The positive value of Sk_b indicates that the distribution is positively skewed and therefore there is a concentration of larger values on the right side of the distribution.

10.2 Kurtosis

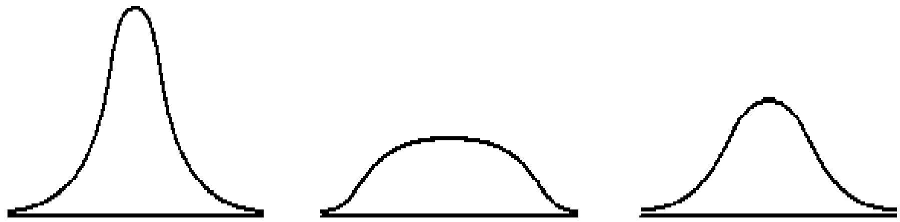
Kurtosis is another measure of the shape of a distribution. Whereas skewness measures the lack of symmetry of the frequency curve of a distribution, kurtosis is a measure of the relative peakedness of its frequency curve.

The measure of kurtosis describes the degree of concentration of frequencies (observations) in a given distribution. That is, whether the observed values are concentrated more around the mode (a peaked curve) or away from the mode towards both tails of the frequency curve.

The word 'kurtosis' comes from a Greek word meaning 'humped'. In statistics, it refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. A few definitions of kurtosis are as follows:

- The degree of kurtosis of a distribution is measured relative to the peakedness of a normal curve. —Simpson and Kafka
- A measure of kurtosis indicates the degree to which a curve of a frequency distribution is peaked or flat-topped. —Croxten and Cowden
- Kurtosis refers to the degree of peakedness of hump of the distribution. —C. H. Meyers

Two or more distributions may have identical average, variation, and skewness, but they may show different degrees of concentration of values of observations around the mode, and hence may show different degrees of peakedness of the hump of the distributions as shown in Fig. 1



(a) Leptokurtic (b) Platykurtic (c) Mesokurtic

Measures of Kurtosis

A measure of Kurtosis is given by $\beta_2 \cong \frac{\mu_4}{\mu_2^2}$, a coefficient given by Karl Pearson. The value of β_2= 3 for a mesokurtic curve. When β_2> 3, the curve is more peaked than the mesokurtic curve and is termed as lepotokurtic. Similarly, when β_2< 3, the curve is less peaked than the mesokurtic curve and is called as platykuritc curve.

Example 4: The first four moments of a distribution about the value 5 of the variable are 2, 20, 40, and 50. Show that the mean is 7. Also find the other moments, β1 and β2, and comment upon the nature of the distribution.

Solution: From the data of the problem, we have

$\mu' = 2, 2\ \mu' = 20, 3\ \mu' = 40, 4\ \mu' = 40$ and A = 5

Now the moments about the arbitrary point 5 are calculated as follows

Mean, x = $1\mu' + A = 2 + 5 = 7$

Variance, $\mu_2 = 2\ \mu' - (\mu'_1)\ 2 = 20 - (2)2 = 16$

$\mu_3 = 3\ \mu' - 3\mu'_1\ \mu'_2 + 2\ (\mu'_1)\ 3 = 40 - 3\ (2)\ (20) + 2\ (2)3 = -64$

$\mu 4 = 4\ \mu' - 4\ \mu'_1\mu'3 + 6\mu'2\ (\mu'_1\ )\ 2 - 3\ (\mu'\ _1)4 = 50 - 4\ (2)\ (40) + 6\ (20)\ (2)2 - 3\ (2)4 = 162$

The two constants, β_1 and β_2, calculated from central moments are as follows:

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-64)^2}{(16)^3} = \frac{4096}{4096} = 1$

$\beta_2 = \frac{\mu_4}{\mu_2^2} \cong \frac{162}{(16)^2} = 0 \cdot 63$

$$\gamma_1 = \frac{1u_3}{\mu_2^{3/2}} = \frac{-64}{(16)\frac{3}{2}} = -1 < 0, distributionisnegativelyskewed$$

γ_2= β_2 -3= 0.63-3=-2.37 (<0), distribution is platykurtic.

Example 5: Find the standard deviation and kurtosis of the following set of data pertaining to kilowatt hours (kwh) of electricity consumed by 100 persons in a city.

Consumption (in kwh): 0–10 10–20 20–30 30–40 40–50

Number of users: 10 20 40 20 10

Solution: The calculations for standard deviation and kurtosis are shown Table

Consumption (in kwh)	Number of users	Mid-Value(m)	D=(m-A)/10	-fd	fd²
0–10	10	5	-2	-20	40
10–20	20	15	-1	-20	20
20–30	40	25=A	0	0	0
30–40	20	35	1	20	20
40–50	10	45	2	20	40
	100			0	120

$$\bar{x} = A + \frac{\Sigma fd}{N} * h$$

Since x = 25 is an integer value, therefore we may calculate moments about the actual mean

$$\mu_r = \frac{1}{n}\Sigma f(x-\bar{x})^r = \frac{1}{n}\Sigma f(m-\bar{x})^r$$

Let d= $\frac{m-\bar{x}}{h}$ $or(m-\bar{x}) = hd$. Therefore

$\mu_r = \frac{1}{n}\Sigma fd^r$; h= width of class intervals

Mid-Value(m)	D=(m-A)/10	-fd	fd²	fd³	fd⁴
5	-2	-20	40	-80	160
15	-1	-20	20	-20	20
25=A	0	0	0	0	0
35	1	20	20	20	20
45	2	20	40	80	160
		0	120	0	360

Moments about the origin A = 25 are:

$\mu_1 = h\frac{1}{N}\Sigma fd = 10 \times \frac{1}{100} = 0$

$$\mu_3 = h^3\frac{1}{N}\Sigma fd^3 = (10)^3 \cdot \frac{1}{100} \cdot 0 = 0$$

$$\mu_2 = h^2\frac{1}{N}\Sigma fd^2 = (10)^2 * \frac{1}{100} * 120 = 120$$

$$\mu_4 = h^4\frac{1}{N}\Sigma fd^4 = (10)^4 * \frac{1}{100} * 360 = 36000$$

$$S \cdot D(\sigma) = \sqrt{\mu_2} = \sqrt{120} = 10.95$$

Karl Pearson's measure of kurtosis is givenby

$$\beta_2 \equiv \frac{\mu_4}{\mu_2^2} = \frac{36000}{(120)^2} = 2.5$$

And therefore

$\gamma_2 = \beta_2 - 3 = 2.5 - 3 = -0.50$

Since $\beta 2 < 3$ (or $\gamma 2 < 0$), distribution curve is platykurtic.

Summary

In this Unit you have learned about the measures of skewedness and kurtosis. These two concepts are used to get an idea about the shape of the frequency curve of a distribution. Skewness is a measure of the lack of symmetry whereas kurtosis is a measure of the relative peaked Ness of the top of a frequency curve.

Keywords

Asymmetry: A characteristic of a distribution in which the values of variables are not equidistant from the average on both sides.

Co-efficient of Skewness: It makes comparison between the skewness in two or more data sets.

Skewness: It refers to the lack of symmetry in distribution.

Symmetry: A characteristic of a distribution in which the values of variables are equidistant from the average on both sides.

Kurtosis:The degree of kurtosis of a distribution is measured relative to the peakedness of a normal curve.

Moment:It represents a convenient and unifying method for summarizing certain descriptive statistical measures.

Leptokurtic:Leptokurtic distributions are statistical distributions with kurtosis greater than three. It can be described as having a wider or flatter shape with fatter tails resulting in a greater chance of extreme positive or negative events.

Self Assessment

1. The moments about mean are called:

A. Raw moments
B. Central moments
C. Moments about origin
D. All of the above

2. For a negatively skewed distribution, the relationship between mean, media and mode is_________

A. Mean> median > mode
B. Mean> mode> median
C. Mean = mode = median
D. Mean< median< mode

3. For a positively skewed distribution the relationship between mean, median and mode is_______

A. Mean> median=mode

B. Mode< median < mean

C. Mean = mode < median

D. Mean< mode<median

4. Bowely's coefficient of skewness lies between

A. -3 to +3

B. -1 to +1

C. $-\infty$ to $+\infty$

D. 0 to ∞

5. For a negatively skewed distribution

A. Left tail is elongated

B. Right tail is elongated

C. Both tails are equally elongated

D. Nothing can be said about tail

6. For a positively skewed distribution

A. Left tail is elongated

B. Right tail is elongated

C. Both tails are equally elongated

D. Nothing can be said about tail

7. Kurtosis is _______

A. Spread of the frequency curve.

B. Central tendency of frequency distribution.

C. Symmetry of the frequency distribution.

D. Relative height of the frequency curve.

8. For mesokurtic distribution, the value of $\beta 2$ is

A. Equal to 3

B. Greater than 3

C. Less than 3

D. Equal to 0

9. The kurtosis defines the peakedness of the curve in the region which is

A. around the mode

B. around the mean

C. around the median

D. around the variance

10. The distribution is considered leptokurtic if

A. beta three is less than three

B. beta two is greater than two

C. beta three is greater than three
D. beta two is greater than three

11. The three times of difference between mean and median is divided by standard deviation to calculate coefficient of skewness by method of
A. Professor Keller
B. Professor Bowley
C. Karl Pearson
D. Professor Kelly

12. In kurtosis, the frequency curve that has flatten top than normal curve of bell-shaped distribution is classified as
A. leptokurtic
B. platykurtic
C. mega curve
D. mesokurtic

13. The method of calculating coefficient of skewness by Karl Pearson method is useful for the type of distributions that are
A. non concentrated
B. open ended
C. close ended
D. concentrated

14. The distribution whose mode is not well defined and the classes of distribution are open ended uses the coefficient of skewness by
A. Karl Pearson
B. Professor Kelly
C. Professor Keller
D. Professor Bowley

15. The degree or extent to which the frequency of the observations in data set are concentrated in given frequency distribution is classified as
A. alpha system
B. gamma system
C. beta system
D. kurtosis

16. As compared to measures of variation, the measure of skewness is used to understand concentration of
A. values around mean
B. upper tail only
C. lower tail only
D. median coefficients

Answers for Self Assessment

1. B	2. D	3. B	4. A	5. B
6. A	7. D	8. A	9. A	10. D
11. C	12. B	13. B	14. D	15. D
16. A				

Review Questions

1. Explain the meaning of skewness using sketches of frequency curves. State the different measures of skewness that are commonly used. How does skewness differ from dispersion?
2. Explain the term 'skewness'. What purpose does a measure of skewness serve? Comment on some of the well-known measures of skewness.
3. Distinguish between Karl Pearson's and Bowley's coefficient of skewness. Which one of these would you prefer and why?
4. Explain the terms 'skewness' and 'kurtosis' used in connection with the frequency distribution of a continuous variable. Give the different measures of skewness (any two of the measures to be given) and kurtosis.
5. How do measures of central tendency, dispersion, skewness, and kurtosis help in analyzing a frequency distribution? Explain with the help of an example
6. Calculate Bowley's coefficient of skewness from the following data

Sales	Below 50	60	70	80	90
No. of Companies	8	20	40	65	80

7. From the following information, calculate Karl Pearson's coefficient of skewness.

Measure	Place A	Place b
Mean	256.5	240.8
Median	201.0	201.6
S.D	215.0	181.0

8. The central moments of a distribution are given by $\mu_2 = 140$, $\mu_3 = 148$, $\mu_4 = 6030$. Calculate the moment measures of skewness and kurtosis and comment on the shape of the distribution.
9. Find the kurtosis for the following distribution
   Class interval: 0–10, 10–20, 20–30, 30–40
   Frequency: 1,3,4,2
   Comment on the nature of the distribution.
10. What do you mean by 'kurtosis' in statistics? Explain one of the methods of measuring it.

Further Readings

- Pagano, R. (2004). Understanding Statistics in the Behavioural Sciences (7th edition).

Pacific grove, ca: brooks/cole publishing co.

- Guilford, J.P...(1956). Fundamental Statistics in Psychology and Education. Mcgraw-hill book company. NY

Web Links

- Garrett, H.E. (1981), Statistics in Psychology and Education, (Tenth edition), Bombay, Vakils Feffer and Simons Ltd.
- McBride, Dawn M. (2018). The Process of Statistical Analysis in Psychology. Sage. USA
- Minium, E.W., King, B.M. & Bear. G (2001). Statistical Reasoning in Psychology and Education (3rd edition), Singapore, John Wiley & Sons, Inc.
- Mohanty, B. &Misra, Santa (2016). Statistics for Behavioural and Social Sciences. Sage. New Delhi.

Unit 11: Probability

CONTENTS

Objectives

- Help yourself to understand the amount of uncertainty that is involved before making important decisions.
- Understand fundamentals of probability and various probability rules that help to you measure uncertainty involving uncertainty.

Introduction

Probability theory plays a very important role in many areas of physical, social, biological, engineering and management sciences. It lays the foundations for a systematic study of mathematical statistics. Games of chance, number of accidents, birth and death rates, system reliability and expected gain in a business venture are some examples where the probability concepts are used.

In probability theory, we are usually interested in the occurrence or non-occurrence of some events. There are several ways of defining the probability of an event. Various definitions of probability are, in general, consistent with one another. In this unit, you will learn about these definitions. You will also learn about the ways, of calculating the probability of events in simple cases and methods of building some simple probability models.

11.1 History of Probability

The history of probability theory dates back to the 17th century. During that period, the classical theory of probability was propounded by several distinguished scientists. Among others, Pascal, Fermat, Huygens and Jakob Bernoulli applied it to games of chance and obtained numerical values of probability of various events by using the classical theory.

The relative frequency definition of probability of events which show statistical regularity in repeated experiments is due to Richard von Mises. He developed this theory around the year 1921. At that time, the classical theory was prevalent and the relative frequency approach provided a new dimension. This definition is quite popular among engineers and experimental scientists as it gives a physical interpretation to the occurrence of an event.

Both classical and relative frequency definitions give a method of assigning a numerical value to the probability of an event. This can also be done simply by interpreting probability as the degree of belief according to the subjective opinion of a person. For example, the event that a particular team will win a tournament may be assigned a probability 0.65 by an expert of the game. Such probabilities when assigned are called subjective probabilities.

The modem theory of probability owes a great deal to the work of several Russian mathematicians, notably A. Kolmogorov, who gave the axiomatic definition of probability in the year 1933. According to this, the probability of an event satisfies three axioms without any concern for its physical interpretation. As you will find out shortly it provides a solid foundation for a deeper study of probability and statistics.

11.2 Probability

Initial studies in probability theory originated from calculation of gambling odds. Slowly over a period of years, it found its applications in other areas where the outcome of an individual random experiment cannot be predicted with certainty. In many cases some events occur more often than others and it is desirable to attach a quantitative measure to various events of a random experiment. For example, if you toss five coins, you may like to assign a quantitative measure to the following two events:

1. The total number of heads observed is more than the total number of tails, and

2. Either two or three heads are observed.

Similarly, before the birth of a baby, a doctor may like to assign a quantitative measure to the event that the weight of baby is less than 1.5 kgms. Probability is such a quantitative measure. It is measured in a unit of "unity", although sometimes in everyday life we also express it as a percentage after multiplying it by 100. Thus, you may say that the chances that a team will win a particular tournament are 35%. The probability "p" of an event also reflects the degree of belief, which you have in the occurrence of that event. A high value of "p" indicates that you are almost certain that the event will occur, whereas a low value of "p" indicates that the event is almost impossible. Using appropriate analysis, if you find that the probability that a given dam will develop a major structural defect in the next 50 years is 0.001, then you are almost certain that this event will not occur and that the design is a safe one. You have thus seen that the probability of an event is a quantitative measure showing the degree of belief which one has in the occurrence or non-occurrence of the event under consideration.

11.3 Definition of Probability

A general definition of probability states that probability is a numerical measure (between 0 and 1 inclusively) of the likelihood or chance of occurrence of an uncertain event. However, it does not tell us how to compute the probability. In this section, we shall discuss different conceptual approaches of calculating the probability of an event.

11.4 Kolmogorov's Axiomatic Approach Probability

Axiomatic probability is a unifying probability theory. It sets down a set of axioms (rules) that apply to all of types of probability, including frequents probability and classical probability. These rules, based on Kolmogorov's Three Axioms, set starting points for mathematical probability.

The three axioms are:

1. For any event A, P(A) ≥ 0. In English, that's "For any event A, the probability of A is greater or equal to 0".
2. When S is the sample space of an experiment; i.e., the set of all possible outcomes, P(S) = 1. In English, that's "The probability of any of the outcomes happening is one hundred percent", or—paraphrasing— "anytime this experiment is performed, something happens".
3. If A and B are mutually exclusive outcomes, P(A ∪ B) = P(A) + P(B).

Here ∪ stands for 'union'. We can read this by saying "If A and B are mutually exclusive outcomes, the probability of either A or B happening is the probability of A happening plus the probability of B happening"

Many important laws are derived from Kolmogorov's three axioms. For example, the Law of Large Numbers can be deduced from the laws by logical reasoning (Tijms, 2004).Just because these axioms are universal, doesn't mean they provide all the answers. For example, any function that satisfies all three axioms is called a probability function. However, the axioms don't tell you which function to choose; it merely states that the probability function you choose must satisfy the rules.

Broadly Probability can be understood as following define approaches:

11.5 Classical Approach

This approach of defining the probability is based on the assumption that all the possible outcomes (finite in number) of an experiment are mutually exclusive and equally likely. It states that, during a random experiment, if there is 'a' possible outcome where the favorable event A occurs and 'b' possible outcomes where the event A does not occur, and all these possible outcomes are mutually exclusive, exhaustive, and equiprobable, then the probability that event A will occur is defined as

$$P(A)=\frac{a}{a+b}=\frac{NumberofFavourableOutcomes}{Totalnumberofoutcomes}=\frac{c(A)}{c(S)}=$$

For example, if a fair die is rolled, then on any trial each event (face or number) is equally likely to occur since there are six equally likely exhaustive events, each will occur 1/6 of the time, and therefore the probability of any one event occurring is 1/6. Similarly for the process of selecting a card at random, each event or card is mutually exclusive, exhaustive, and equiprobable. The probability of selecting any one card on a trial is equal to 1/52, since there are 52 cards. Hence, in general, for a random experiment with n mutually exclusive, exhaustive, equiprobable events, the probability of any of the events is equal to 1/n.

Since the probability of occurrence of an event is based on prior knowledge of the process involved, therefore this approach is often called a priori classical probability approach. This means, we do not have to perform random experiments to find the probability of occurrence of an event. This also implies that no experimental data are required for computation of probability. Since the assumption of equally likely simple events can rarely be verified with certainty, therefore this approach is not used often other than in games of chance.

The assumption that all possible outcomes are equally likely may lead to a wrong calculation of probability in case some outcomes are more or less frequent in occurrence. For example, if we classify two children in a family according to their sex, then the possible outcomes in terms of number of boys in the family are 0, 1, 2. Thus according to the classical approach, the probability for each of the outcomes should be 1/3. However, it has been calculated that the probabilities are approximately 1/4, 1/2, and 1/4 for 0, 1, 2 boys respectively. Similarly, we cannot apply this approach to find the probability of a defective unit being produced by a stable manufacturing process as there are only two possible outcomes, defective or non-defective.

11.6 Relative Frequency Approach

In situations where the outcomes of a random experiment are not all equally likely or when it is not known whether outcomes are equally likely, application of the classical approach is not desirable to quantify the possible occurrence of a random event. For example, it is not possible to state in advance, without repetitive trials of the experiment, the probabilities in cases like (i) whether a number greater than 3 will appear when die is rolled or (ii) if a lot of 100 items will include 10 defective items.

This approach of computing probability is based on the assumption that a random experiment can be repeated a large number of times under identical conditions where trials are independent to each other. While conducting a random experiment, we may or may not observe the desired event. But as the experiment is repeated many times, that event may occur some proportion of time. Thus, the approach calculates the proportion of the time (i.e. the relative frequency) with which the event occurs over an infinite number of repetitions of the experiment under identical conditions. Since no experiment can be repeated an infinite number of times, therefore a probability can never be exactly determined. However, we can approximate the probability of an event by recording the relative frequency with which the event has occurred over a finite number of repetitions of the experiment under identical conditions. For example, if a die is tossed n times and s denotes the number of times the event A (i.e., number 4, 5, or 6) occurs, then the ratio P(A) = c(s)/n gives the proportions of times the event A occurs in n trials, and are also called relative frequencies of the event in n trials. Although our estimate about P(A) may change after every trial, yet we will find that the proportion c(s)/n tends to cluster around a unique central value as the number of trials n becomes even larger. This unique central value (also called probability of event A) is defined as:

$$p(A) = \lim_{n\to\infty} \left\{ \frac{C(s)}{n} \right\}$$

where c(s) represents the number of times that an event s occurs in n trials of an experiment.

Since the probability of an event is determined objectively by repetitive empirical observations of experimental outcomes, it is also known as empirical probability. Few situations to which this approach can be applied are follows:

1. Buying lottery tickets regularly and observing how often you win
2. Commuting to work daily and observing whether or not a certain traffic signal is red when cross it.
3. Observing births and noting how often the baby is a female
4. Surveying many adults and determine what proportion smokes.

11.7 Subjective Approach

The subjective approach of calculating probability is always based on the degree of beliefs, convictions, and experience concerning the likelihood of occurrence of a random event. It is thus a way to quantify an individual’s beliefs, assessment, and judgment about a random phenomenon. Probability assigned for the occurrence of an event may be based on just guess or on having some idea about the relative frequency of past occurrences of the event. This approach must be used when either sufficient data are not available or sources of information giving different results are not known.

11.8 Basic Concepts of Probability

Probability, in common parlance, connotes the chance of occurrence of an event or happening. In order that we are able to measure it, a more formal definition is required. This is achieved through the study of certain basic concepts in probability theory, like experiment, sample space and event. In this section we explore these concepts.

Experiment

The term experiment is used in probability theory in a much broader sense than in physics or chemistry. Any action, whether it is the tossing of a coin, or measurement of a product's dimension to ascertain quality, or the launching of a new product in the market, constitute an experiment in the probability theory terminology.

These experiments have three things in common:

1. There is two or more outcomes of each experiment.

2. It is possible to specify the outcomes in advance.

3. There is uncertainty about the outcomes.

For example, a coin tossing may result in two outcomes, in head or tail, which we know in advance, and we are not sure whether a head or a tail will come up when we toss the coin. Similarly, the product we are measuring may turn out to be undersize or right size or oversize, and we are not certain which way it will be when we measure it. Also, launching a new product involves uncertain outcome of meeting with a success or failure in the market.

Sample Space

The set of all possible outcomes of an experiment is defined as the sample space. Each outcome is thus visualized as a sample point in the sample space. Thus, the set (head, tail) defines the sample space of a coin tossing experiment. Similarly, (success, failure) defines the sample space for the launching experiment. You may note here, that given any experiment, the sample space is fully determined by listing down all the possible outcomes of the experiment.

Event

An event, in probability theory, constitutes one or more possible outcomes of an experiment. Thus, an event can be defined as a subset of the sample space. Unlike the common usage of the term, where an event refers to a particular happening or incident, here, we use an event to refer to a single outcome or a combination of outcomes. Suppose, as a result of a market study experiment of a product, we find that the demand for the product for the next month is uncertain, and may take values from 100, 101, 102... 150. We can obtain different event like:

The event that demands is exactly 100.

The event that demands lies between 101 to 120.

The event that demands is 101 or 102.

In the first case, out of the 51 sample points that constitute the sample space, only one sample point or outcome defines the event, whereas the number of outcomes used in the second and third case are 20 and 2 respectively.

With this background on the above concepts, we are now in a position to formalize the definition of probability of an event. In the next section, we will look at the different approaches to probability that have been developed, and present the axioms for the definition of probability.

Example 1: Suppose we are interested in the following Event A in the above experiment: The number of defectives is exactly two. How many sample' points does this event correspond to?

Solution:

We can see from the sample space that there are three outcomes where D occurs twice, viz, DDG, DGD and GDD, thus the Event A corresponds to 3 sample point.

Exhaustive Cases

The total number of possible outcomes in a random experiment is called the exhaustive cases. In other words, the number of elements in the sample space is known as number of exhaustive cases, e.g.

i. If we toss a coin, then the number of exhaustive cases is 2 and the sample space in this case is {H, T}.

ii. If we throw a die then number of exhaustive cases is 6 and the sample space in this case is {1, 2, 3, 4, 5, 6}.

Favorable Cases

The cases which favor to the happening of an event are called favorable cases. e.g.

i. For the event of drawing a card of spade from a pack of 52 cards, the number of favorable cases is 13.

ii. For the event of getting an even number in throwing a die, the number of favourable cases is 3 and the event in this case is {2, 4, 6}.

Mutually Exclusive Cases

Cases are said to be mutually exclusive if the happening of any one of them prevents the happening of all others in a single experiment, e.g.

i. In a coin tossing experiment head and tail are mutually exclusive as there cannot be simultaneous occurrence of head and tail.

Equally Likely Cases

Cases are said to be equally likely if we do not have any reason to expect one in preference to others. If there is some reason to expect one in preference to others, then the cases will not be equally likely, For example,

i. Head and tail are equally likely in an experiment of tossing an unbiased coin. This is because if someone is expecting say head, he/she does not have any reason as to why he/she is expecting it.
ii. All the six faces in an experiment of throwing an unbiased die are equally likely.

You will become more familiar with the concept of "equally likely cases" from the following examples, where the non-equally likely cases have been taken into consideration.

i. Cases of "passing" and "not passing" a candidate in a test are not equally likely. This is because a candidate has some reason(s) to expect "passing" or "not passing" the test. If he/she prepares well for the test, he/she will pass the test and if he/she does not prepare for the test, he/she will not pass. So, here the cases are not equally likely.
ii. Cases of "falling a ceiling fan" and "not falling" are not equally likely. This is because, we can give some reason(s) for not falling if the bolts and other parts are in good condition.

11.9 Algebra of Events

Let C be a fixed sample space. We have already defined an event as a collection of sample points from C. Imagine that the (conceptual) experiment underlying 6 is being performed. The phrase "the event E occurs" would mean that the experiment results in an outcome that is included in the event E. Similarly, non-occurrence of the event E would mean that the experiment results into an outcome that is not an element of the event E. Thus, the collection of all sample points that are not included in the event E is also an event which is complementary to E and is denoted as EC. The event EC is therefore the event which contains all those sample points of < which are not in E. As such, it is easy to see that the event E occurs if and only if the event EC does not take place. The events E and EC are complementary events and taken together they comprise the entire sample space, i.e., E U E^C= C

You may recall that C is an event which consists of all the sample points. Hence, its complement is an empty set in the sense that it does not contain any sample point and is called the null event, usually denoted as 4 so that CC = Φ.

Example 1: Consider the event E that the three tosses produce at least one head.

Solution: Thus, E = {sl , s2, s3, s4, ss, s6, s7) so that the complementary event E^C = {ss}, which is the event of not scoring a head at all.

Example 2: n the case of selection without replacement, event that the white marble is picked up at least once is defined as E = {(r_l , w), (r_2, w.) , (w, r_2) (w , r_l)). Hence. E^C = {(r_l, r_2), (r_z, r_l)) i.e. the event of not picking the white marble at all.

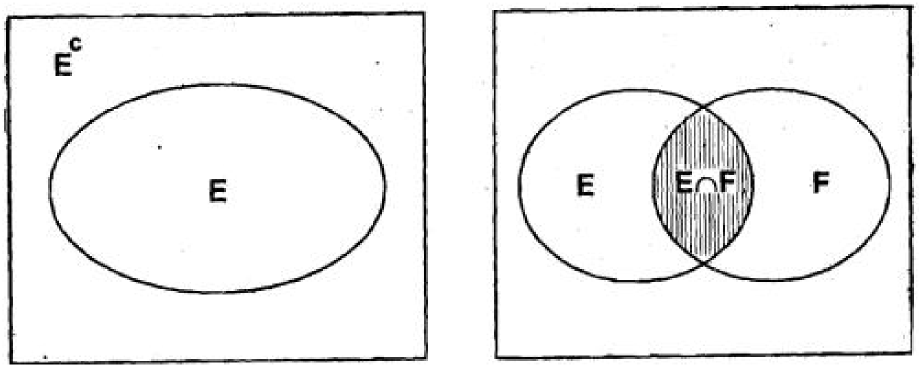
Let us now consider two events E and F. We write E U F, read as E "union" F, to denote the collection of sample points, which are responsible for occurrence of either E or F or both. Thus, E U

F is a new event and it occurs if and only if either E or F or both occur i.e. if and only if at least one of the events E or F occurs. Generalizing this idea, we can define a new event $\bigcup_{j=1}^{k} E_j$ read as "union" of the k events El, E2,. . . , Ek, as the event which consists of all sample points that are in at least one of the events El, E2, . . . , Ek and it occurs if and only if at least one of the events El, E2, . . . , Ek occurs.

Again, let E and F be two given events. We write E n F, read as E "intersection" F, to denote the collection of sample points any of whose occurrence implies the occurrence of both E and F, Thus, E fl F is a new event and it occurs if and only if both the events E and F occur. Generalizing this idea, we can define a new event $\bigcap_{j=1}^{k} E_j$ read as "intersection" of the k events E_l, E_2,. . . , E_k, as the event which consists of sample points that are common to each of the events E_l, E_2, . . . , E_k. and it occurs only if all the k events E_l, E_2, . . . , E_k occur simultaneously.

Further, two events E and F are said to be mutually exclusive or disjoint if they do not have a common sample point i.e. E n F = Φ. Two mutually exclusive events then cannot occur simultaneously. In the coin-tossing experiment for instance, the two events, heads and tails, are mutually exclusive: if one occurs, the other cannot occur.

Let E be the event of scoring an odd number of heads and F be the event that tail appears in the first two tosses, so that E = {sl, ss, s6, s7) and F = {ss, ss). Now E n F = {ss), the event that only the third toss yields a head. Thus events E and F are not mutually exclusive.



a) b)

The above relations between events can be best viewed through a Venn diagram. A rectangle is drawn to represent the sample space c. All the sample points are represented within the rectangle by means of points. An event is represented by the region enclosed by a closed curve containing a!l the sample points leading to that event. The space inside the rectangle but outside the closed curve representing E represents the complementary event ECin (a). Similarly, in Fig.l(b), the space inside the curve represented by the broken line represents the event E U F and the shaded portion represents E ∩ F.

11.10 Probability Rules

In probability we use set theory notations to simplify the presentation of ideas. As discussed earlier in this chapter, the probability of the occurrence of an event A is expressed as:

P(A) = probability of event A occurrence

Such single probabilities are called marginal (or unconditional) probabilities because it is the probability of a single event occurring. In the coin tossing example, the marginal probability of a tail or head in a toss can be stated as P(T) or P(H).

Before considering various probability laws, let us be familiar with certain notations.

a) If A and B are two events, then P(A∪B) or P(A+ B) denotes the probability that either A occurs or B occurs or both occur simultaneously. It can also be interpreted as the probability of the occurrence of at least one of the two events A and B. The symbol u above represents 'union' between two events. (Read A∪B as 'A union By).

b) P(A∩B) or P(AB) denotes the probability of thesimultaneous occurrence of both A and B. (Read A∩B as 'A intersection B').

c) P(A / B) denotes the conditional probability of the occurrence of A given. that B has already occurred.

11.11 Rules of Addition

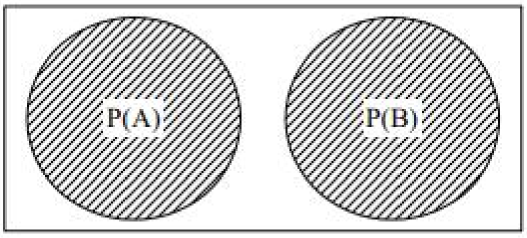
The addition rules are helpful when we have two events and are interested in knowing the probability that at least one of the events occurs.

1. *Addition Rule for Mutually Exclusive Events*

If two events, A and B, are mutually exclusive, then the probability of occurrence of either A or B is given by the following formula:

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If A and B are mutually exclusive events, this rule is depicted in Figure 8.1, below



The essential requirement for any two events to be mutually exclusive is that there are no outcomes common to the occurrence of both. This condition is satisfied when sample space does not contain any outcome favorable to the occurrence of both A and B means $A \cap B = \varphi$

Example 2: In a game of cards, where a pack contains 52 cards, 4 categories exist namely spade, club, diamond, and heart. If you are asked to draw a card from this pack, what is the probability that the card drawn belongs to either spade or club category

Solution: Here, P (Spade or club) = 13/52=1/4, P (Club)= 13/52=1/4

P (Spades) + P (Club)=1/4+1/4=1/2

There is an important special case for any event E, either E happens or it does not. So, the events E and not E are exhaustive and exclusive.

So, P (E) + P (not E) = 1

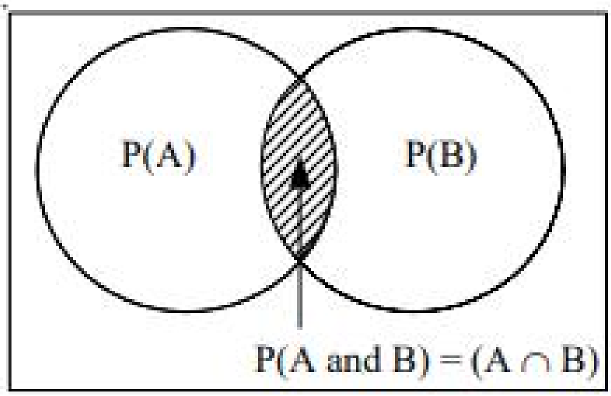
or, P (E) = 1–P (not E)

Sometimes P (not E) is also written as either P(E) = 1- P($\bar{E}$)

So, P (E)= 1-P(E)=1-P($\bar{E}$)

2. *Addition Rule for Non-Mutually Exclusive Events*

Non-mutually exclusive (overlapping) events present another significant variant of the additive rule. Two events (A and B) are not mutually exclusive if they have some outcomes common to the occurrence of both, then the above rule has to be modified in order to account for the overlapping areas, as it is clear from Figure 8.2. below.



In this situation, the probability of occurrence of event A or event B is given by the formula

P (A or B) = P (A ∪ B) = P (A) + P (B) – P (A and B)

where P (A and B) is the joint probability of events A and B, i.e., both occurring together and is usually written as P (A ∩ B).

Thus, it is clear that the probability of outcomes that are common to both the events is to be subtracted from the sum of their simple probability.

Example 3: The event of drawing either a Jack or a spade from a well shuffled deck of playing cards. Find the probability.

Solution: These events are not mutually exclusive, so the required probability of drawing a Jack or a spade is given by:

P (Jack or Spade) = P (Jack) + P (Spade) – P (Jack and Spade)

=4/52+13/52-1/52=16/52=4/13

11.12 Rules of Multiplication

Statistically Independent Event:

When the occurrence of an event does not affect and is not affected by the probability of occurrence of any other event, the event is said to be a statistically independent event. There are three types of probabilities under statistical independence: marginal, joint, and conditional.

Probability Under Statistical Independence

The two or more events are termed as statistically independent events, if the occurrence of any one event does not have any effect on the occurrence of any other event. For example, if a fair coin is tossed once and suppose head comes, then this event has no effect in any way on the outcome of second toss of that same coin. Similarly, the results obtained by drawing hearts from a pack has no effect in any way on the results obtained by throwing a dice. These events thus are being termed as statistically independent events. There are three types of probability under statistically independent case.

a) Marginal Probability;

b) Joint Probability;

c) Conditional Probability

a) Marginal Probability Under Statistical Independence

A Marginal/Simple/Unconditional probability is the probability of the occurrence of an event. For example, in a fair coin toss, probability of having a head is:

P (H) = ½ = 0.5

Therefore, the marginal probability of an event (i.e. having a head) is 0.5. Since, the subsequent tosses are independent of each other, therefore, it is a case of statistical independence.

Another example can be given in a throw of a fair die, the marginal probability of the face bearing number 3, is:

P(3) = 1/6 = 0.166

Since, the tosses of the die are independent of each other, this is a case of statistical independence.

b) Joint Probability Under Statistical Independence

This is also termed as "Multiplication Rule of Probability". In many situations we are interested in finding out the probability of two or more events either occurring together or in quick succession to each other, for this purpose the concept of joint probability is used.

This joint probability of two or more statistically independent events occurring together is determined by the product of their marginal probability. The corresponding formula may be expressed as:

P (A and B) = P (A) × P (B)

Similarly, it can be extended to more than two events also as:

P (A and B and C) = P (A) × P (B) × P (C) and so on.

i.e. P (A and B and C and …) = P (A) × P (B) × P (C) × …

For instance, when a fair coin is tossed twice in quick succession, the probability of head occurring in both the tosses is:

P(H_1 and H_2) = P (H_1) × P (H_2)

= 0.5 × 0.5 = 0.25

Where, H_1 is the occurrence of head in 1st toss, and H_2 is the occurrence of head in 2nd toss

Take another example: When a fair die is thrown twice in quick succession, then to find the probability of having 2 in the 1st throw and 4 in second throw is, given as:

P (2 in 1st throw and 4 in 2nd throw)

= P (2 in the 1st throw) × P (4 in the 2nd throw)

= 1/6*1/6=1/36=0.028

c) Conditional Probability under the Condition of Statistical Independence

The third type of probability under the condition of statistical independence is the Conditional Probability. It is symbolically written as P (A/B), i.e., the conditional probability of occurrence of event A, on the condition that event B has already occurred.

In case of statistical independence, the conditional probability of any event is akin to its marginal probability, when both the events are independent of each other.

Therefore, P (A/B) = P (A), and

P (B/A) = P (B).

For Example, if we want to find out what is the probability of heads coming up in the second toss of a fair coin, given that the first toss has already resulted in head. Symbolically, we can write it as:

P (H_2 /H_1)

As, the two tosses are statistically independent of each other

so, P (H_2 /H_1) = P (H_2)

The following table 8.1 summarizes these three types of probabilities, their symbols and their mathematical formulae under statistical independence.

Probability's type	Symbol	Formula
Marginal	P(A)	P(A)
Joint	P(AB)	P(A)* P(B)
Conditional	P(B/A)	P(B)

Probability Under Statistical Dependence

Two or more events are said to be statistically dependent, if the occurrence of any one event affects the probability of occurrence of the other event.

There are three types of probability under statistical dependence case. They are:

a) Conditional Probability;

b) Joint Probability;

c) Marginal Probability

a) Conditional Probability under Condition of Statistical Dependence

The conditional probability of event A, given that the event B has already occurred, can be calculated as follows:

$$P(A / B) = \frac{P(AB)}{P(B)}$$

Where, P (AB) is the joint probability of events A and B.

Example 4: (i) A box containing 10 balls which have the following distribution on the basis of colour and pattern.

a. 3 are colored and dotted.
b. 1 is colored and stripped.
a. Suppose someone draws a colored ball from the box. Find what is the probability that it is (i) dotted and (ii) it is stripped?

Solution: The problem can be expressed as P (D/C) i.e., the conditional probability that the ball drawn is dotted given that it is colored.

Now from the information given in the question.

i) P(CD) = 3/10 = Joint Probability of drawn ball becoming a colored as well as a dotted one.

Similarly, P (CS) = 1/10, P (GD) = 2/10, and P (GS) = 4/10

$$P(D / C) = \frac{P(DC)}{P(C)}$$

Where, P(C) = Probability of drawing a colored ball from the box = 4/10 (4 colored balls out of 10 balls).

$P(D/C) = \frac{\frac{3}{10}}{\frac{4}{10}} = 0.75$

ii) Similarly, P(S/C) = Conditional probability of drawing a stripped ball on the condition of knowing that it is a colored one.

$P(D/C) = \frac{\frac{1}{10}}{\frac{4}{10}} = 0.25$

Thus, the probability of colored and dotted ball is 0.75. Similarly, the probability of colored and stripped ball is 0.25.

a. Continuing the same illustration, if we wish to find the probability of (i) P (D/G) and (ii) P (S/G)

Solution:

$P(D / G) = \frac{P(DG)}{P(GC)}$= 2/10/6/10= 1/3=0.33

Where, P(G) = Total probability of grey balls, i.e., 6/10 and

II) $P(S/G) = \frac{P(SG)}{P(G)}$= (4/10)/(6/10)=2/3=0.66

b) Joint Probability Under the Condition of Statistical Dependence

This is an extension of the multiplication rule of probability involving two or more events, which have been discussed in the previous section 13.6, for calculating joint probability of two or more events under the statistical independence condition.

The formula for calculating joint probability of two events under the condition of statistical independence is derived from the formula of Bayes' Theorem.

Therefore, the joint probability of two statistically dependent events A and B is given by the following formula:

P (AB) = P (A/B) × P (B)

or P (BA) = P (B/A) × P (A)

Depending upon whether order of occurrence of two events is B, A or A, B.

Since, P (A/B) = P (B/A), So the product on the RHS of the formula must also be equal to each other.

∴ P (A/B) × P (B) = P (B/A) × P (A)

Notice that this formula is not the same under conditions of statistical independence, i.e., P (BA) = P(B) × P (A). Continuing with our previous illustration 4, of a box containing 10 balls, the value of different joint probabilities can be calculated as follows:

Converting the above general formula i.e., P (AB) = P (A/B) × P (B) into our illustration and to the terms colored, dotted, stripped, and grey, we would have calculated the joint probabilities of P (CD), P (GS), P (GD), and P (CS) as follows:

i. P (CD) = P (C/D) × P (D) = 0.6 × 0.5 = 0.3
ii. P (GS) = P (G/S) × P (S) = 0.8 × 0.5 = 0.4

c) Marginal Probability Under the Condition of Statistical Dependence

Finally, we discuss the concept of marginal probability under the condition of statistical dependence. It can be computed by summing up all the probabilities of those joint events in which that event occurs whose marginal probability we want to calculate.

Example 5: Consider the previous illustration 4, to compute the marginal probability under statistical dependence of the event: i) dotted balls occurred, ii) colored balls occurred, iii) grey balls occurred, and iv) stripped balls occurred.

Solution: We can obtain the marginal probability of the event dotted balls by adding the probabilities of all the joint events in which dotted balls occurred.

P (D) = P (CD) + P (GD) = 3/10 + 2/10 = 0.5

In the same manner, we can compute the joint probabilities of the remaining events as follows:

i. P (C) = P (CD) + P(CS) = 3/10 + 1/10 = 0.4
ii. P (G) = P (GD) + P (GS) = 2/10 + 4/10 = 0.6

The following table 8.2 summarizes three types of probabilities, their symbols and their mathematical formulae under statistical dependence.

Table 8.2: Probabilities under Statistical Dependence

Probability's type	Symbol	Formula
Marginal	P(A)	Sum of the probabilities of joint events in which 'A' occurs
Joint	P(AB) or P(BA)	P(A/B)* P(B) OR P(B/A)* P(A)
Conditional	P(B/A) or P(A/B)	P(B/A)/P(A) OR P(A/B)/P(B)

Summary

At the beginning of this unit the historical evolution and the meaning of probability has been discussed. Contribution of leading mathematicians has been highlighted. Fundamental concepts and approaches to determining probability have been explained. The three approaches namely; the classical, the relative frequency, and the subjective approaches are used to determine the probability in case of risky and uncertain situation have been discussed. Probability rules for calculating probabilities of different types of events have been explained. Further the condition of statistical independence and statistical dependence has been defined. Three types of probabilities namely: marginal, joint and conditional under statistical independence and statistical dependence have been explained. Finally, the Bayesian approach to the revision of a priori probability in the light of additional information has been undertaken.

Keywords

1. Classical/Logical Approach: An objective way of assessing probabilistic value based on logic.
2. Collectively Exclusive Event: This is the collection of all possible outcomes of an experiment.
3. Conditional Probability: The probability of the happening of an event on the condition that another event has already occurred.
4. Dependent Event: This is the situation in which the occurrence of one event affects the happening of another event.
5. Independent Event: This is the situation in which the occurrence of an event has no effect on the probability of the occurrence of any other event.
6. Joint Probability: The probability of occurring of events together or in quick succession
7. Marginal/Simple Probability: As the name suggests, it is the simple probability of occurrence of an event.
8. Mutually Exclusive Events: A situation in which only one event can occur on any given trial/experiment. It means events that cannot occur together.
9. Conditional Probability: If A and B are not mutually exclusive events then the probability of B given that A has already occurred is known as the conditional probability of B given A. It is denoted by P(B/A).

Self Assessment

1. ______________ is a mechanism that produces a definite outcome that _________.

A. Dependent experiment; can be predicted with certainty.
B. Simple experiment; cannot be predicted with certainty.
C. Random experiment; can be predicted with certainty.
D. Random experiment; cannot be predicted with certainty.

2. Consider the experiment of tossing three coins simultaneously. The sample space is given by

A. S = {HHH, HHT, HTH, HHH, HTT, THT, TTH, TTT}
B. S = {HHH, HHT, HTH, THH, TTT, THT, TTH, TTT}
C. S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}
D. S = {HHH, HHT, HTH, THH, THT, THT, TTH, TTT}

3. A set of events {A1, A2, . . ., An}is collectively exhaustive if

A. Union of A1, A2, . . ., An is identical with the sample space, S = {A1∪A2∪. . .∪An}.
B. Union of A1, A2, . . ., An is not identical with the sample space, S = {A1∪A2∪. . .∪An}.
C. Union of A1, A2, . . ., An is equal to one.
D. Union of A1, A2, . . ., Anlie between zero and one.

4. Compound events

A. Are dependent only.
B. are independent only
C. May be dependent or independent.
D. Neither dependent nor independent.

5. ____________ is a way to quantify an individual's belief, assessment, and judgement about a random experiment.
A. Relative frequency approach
B. Subjective approach
C. Classical approach
D. Fundamental approach

6. If A and B are any two events then the probability of happening of at least one of the events is defined as
A. P(AUB) = P(A)/P(B)
B. P(AUB) = P(A) - P(B)
C. P(AUB) = P(A) + P(B)
D. P(AUB) = P(A) * P(B)

7. Partially overlapping events are those events which
A. Are mutually exclusive.
B. Are not mutually exclusive.
C. Have no sample points in common.
D. Are mutually exclusive and have no point in common.

8. Which of the following is correct?
A. $P(A\cup B\cup C) = P(A)+ P(B) – P(A\cap B) + P(C)– [P(A\cap C) + P(B\cap C) – P((A\cap C)\cap(B\cap C))]$
B. $P(A\cup B\cup C) = P(A)+ P(B) + P(A\cap B) + P(C)– [P(A\cap C) + P(B\cap C) – P((A\cap C)\cap(B\cap C))]$
C. $P(A\cup B\cup C) = P(A)+ P(B) – P(A\cap B) - P(C)– [P(A\cap C) + P(B\cap C) – P((A\cap C)\cap(B\cap C))]$
D. $P(A\cup B\cup C) = P(A)+ P(B) – P(A\cap B) + P(C)– [P(A\cap C) + P(B\cap C) + P((A\cap C)\cap(B\cap C))]$

9. Joint Probability (two or more independent events)
A. = Product of marginal probabilities of only two independent events
B. = Product of marginal probabilities of two or more independent events
C. = Product of marginal probabilities of two or more dependent events
D. = Product of marginal probabilities of only two dependent events

10. The conditional probability of B, given that event A has already occurred, is given by
A. $P(B \mid A) = P(A)/P(A\cap B)$
B. $P(A \mid B) = P(A\cap B)/P(B)$
C. $P(B \mid A) = P(A\cap B)/P(B)$
D. $P(B \mid A) = P(A\cap B)/P(A)$

11. If two events (both with probability greater than 0) are mutually exclusive, then:
A. They also must be independent.
B. They also could be independent.
C. They cannot be independent.
D. none of the above

12. Suppose that the probability of event A is 0.2 and the probability of event B is 0.4. Also, suppose that the two events are independent. Then $P(A|B)$ is:
A. P(A) = 0.2

B. P(A)/P(B) = 0.2/0.4 = ½
C. P(A) × P(B) = (0.2)(0.4) = 0.08
D. None of the above.

13. The range of probability is
A. any value greater than zero
B. any value less than one
C. zero to one
D. any value between -1 to 1

14. Two events A and B are statistically independent when
A. $P(A \cap B) = P(A) \times P(B)$
B. $P(A|B) = P(A)$
C. $P(A \cup B) = P(A) + P(B)$
D. both (a) and (b)

15. Which of the following pairs of events are mutually exclusive?
A. A contractor loss a major contract, and he increases his work force by 50 per cent.
B. A man is older than his uncle and he is younger than his cousins.
C. A football team loses its last game of the year, and it wins the world cup.
D. None of these

Answers for Self Assessment

1.	D	2.	C	3.	A	4.	C	5.	B
6.	C	7.	B	8.	A	9.	B	10.	D
11.	C	12.	C	13.	C	14.	D	15.	C

Review Questions

1. Explain whether or not each of the following claims could be correct:
    a. A businessman claims the probability that he will get contract A is 0.15 and that he will get contract B is 0.20. Furthermore, he claims that the probability of getting A or B is 0.50.
    b. A market analyst claims that the probability of selling ten million rupees of plastic A or five million rupees of plastic B is 0.60. He also claims that the probability of selling ten million rupees of A and five million rupees of B is 0.45.
2. Explain what you understand by the term probability. Discuss its importance in busines decision-making.
3. Define independent and mutually exclusive events. Can two events be mutually exclusive and independent simultaneously? Support your answer with an example.
4. Explain the meaning of each of the following terms:

(a) Random phenomenon (b) Statistical experiment (c) Random event (d) Sample space

5. Distinguish between the two concepts in each of the following pairs: (a) Elementary event and compound events (b) Mutually exclusive events and overlapping events (c) Sample space and sample point

6. Suppose an entire shipment of 1000 items is inspected and 50 items are found to be defective. Assume the defective items are not removed from the shipment before being sent to a retail out let for sale. If you purchase one item from this shipment, what is the probability that it will be one of the defective items?
7. Life insurance premiums are higher for older people, but auto insurance premiums are generally higher for younger people. What does this suggest about the risks and probabilities associated with these two areas of insurance business?
8. A problem in business statistics is given to five students, A, B, C, D, and E. Their chances of solving it are 1/2, 1/3, 1/4, 1/5, and 1/6 respectively. What is the probability that the problem will be solved?
9. There are three brands, say X, Y, and Z of an item available in the market. A consumer chooses exactly one of them for his use. He never buys two or more brands simultaneously. The probabilities that he buys brands X, Y, and Z are 0.20, 0.16, and 0.45. (a) What is the probability that he does not buy any of the brands? (b) Given that a customer buys some brand, what is the probability that he buys brand X?
10. Two sets of candidates are competing for positions on the board of directors of a company. The probability that the first and second sets will win are 0.6 and 0.4 respectively. If the first set wins, the probability of introducing a new product is 0.8 and the corresponding probability if the second set wins is 0.3. (a) What is the probability that the new product will be introduced? (b) If the new product was introduced, what is the probability that the first set won as directors?

Further Readings

- Levin, R.I. and Rubin, D.S., 1991, Statistics for Management, PHI, : New Delhi.
- Feller, W., 1957, An Introduction to Probability Theory and Its Applications, John Wiley & Sons Inc.: New York.
- Hooda, R.P. 2001, Statistics for Business and Economics. MacMillan India Limited, Delhi.
- Gupta S.C., and V.K. Kapoor, 2005, Fundamentals of Mathematical Statistics, Sultan Chand & Sons, Delhi.

Unit 12: Probability Distribution

CONTENTS

Objective

- understand the random variables and how they are inseparable to probability distributions;
- Solve the problems of probability, which fit into binomial and normal distributions.

Introduction

In our study of Probability Theory, we have so far been interested in specific outcomes of an experiment and the chances of occurrence of these outcomes. In the last unit, we have explored different ways of computing the probability of an outcome. For example, we know how to calculate the probability of getting all heads in a toss of three coins. We recognize that this information on probability is helpful in our decisions. In this case, a mere 0.125 chance of all heads may dissuade you from betting on the event of "all heads". It is easy to see that it would have been further helpful, if all the possible outcomes of the experiment together with their chances of occurrence were made available. Thus, given your interest in betting on head's, you find that a toss of three coins may result in zero, one, two or three heads with the respective probabilities of 1/8, 3/8, 3/8, and 1/8. The wealth of information, presented in this way, helps you in drawing many different inferences. Looking at this information, you may be more ready to bet on the event that either one or two heads occur in a toss of three coins. This representation of all possible outcomes and their probabilities is known as a probability distribution. Thus, we refer to this as the probability distribution of "number of heads" in the experiment of tossing of three coins. While we see that our previous knowledge on computation of probabilities helps us in arriving at such representations, we recognize that the calculations may be quite tedious. This is apparent, if you try to calculate the probabilities of different number of heads in a tossing of twelve coins. Developments in Probability

Theory help us in specifying the probability distribution in such cases with relative ease. The theory also gives certain standard probability distributions and provides the conditions under which they can be applied. We will study the probability distributions and their applications in this and the subsequent unit. The objective of this unit is to look into a type of probability distribution, viz., a discrete probability distribution. Accordingly, after the initial presentation on the basic concepts and definitions, we will discuss as to how discrete probability distributions can be used in decision-making.

12.1 Concepts Of Probability Distributions

Before we attempt a formal definition of probability distribution, the concept of 'random variable' which is central to the theme, needs to be elaborated.

In the example given in the Introduction, we have seen that the outcomes of the experiment of a toss of three coins were expressed in terms of the "number of heads" Denoting this "number of heads" by the letter H, we find that in the example, H can assume values of 0, 1, 2 and 3 and corresponding to each value, a probability is associated. This uncertain real variable H, which assumes different numerical values depending on the outcomes of an experiment, and to each of whose values a probability assignment can be made, is known as a random variable. The resulting representation of all the values with their probabilities is termed as the probability distribution of H. It is customary to present the distribution as follows:

Table 1: Probability Distribution of Number of Heads (H)	
H	P(H)
0	0.125
1	0.375
2	0.375
3	0.125

In this case, as we find that H takes only discrete values, the variable H is called a discrete random variable and the resulting distribution is a discrete probability distribution.

In the above situation, we have seen that the random variable takes a limited number of values. There are certain situations where the variable of interest may take infinitely many values. Consider for example that you are interested in ascertaining the probability distribution of the weight of the one kilogram tea pack, that is produced by your company. You have reasons to believe that the packing process is such that the machine produces a certain percentage of the packs slightly below one kilogram and some above one kilogram. It is easy to see that there is essentially to chance that the pack will weigh exactly 1.000000 kg., and there are infinite number of values that the random variable "weight" can take. In such cases, it makes sense to talk of the probability that the weight will be between two values, rather than the probability of the weight will be between two values, rather than the probability of the weight taking any specific value. These types of random variables which can take an infinitely large number of values are called continuous random variables, and the resulting distribution is called a continuous probability distribution. Sometimes, for the sake of convenience, a discrete situation with a large number of outcomes is approximated by a continuous distribution: Thus, if we find that the demand of a product is a random variable taking values of 1, 2, 3... to 1000, it may be worthwhile to treat it as a continuous variable. Obviously, the representation of the probability distribution for a continuous random variable is quite different from the discrete case that we have seen. We will be discussing this in a later unit when we take up continuous probability distributions.

Coming back to our example on the tossing of three coins, you must have noted the presence of another random variable in the experiment, namely, the number of tails (say T). T has got the same distribution as H. In fact, in the same experiment, it is possible to have some more random variables, with a slight extension of the experiment. Supposing a friend comes and tells you that he will toss 3 coins, and will pay you Rs. 100 for each head and Rs. 200 for each tail that turns up. However, he will allow you this privilege only if you pay him Rs. 500 to start with.

You may like to know whether it is worthwhile to pay him Rs. 500. In this situation, over and above the random variables H and T, we find that the money that you may get is also a random variable. Thus,

if H =number of heads in any outcome, then 3 - H = number of tails in any outcome (as the total number of heads and tails that can occur in a toss of three coins is 3) The money you get in any outcome = 100H + 200 (3 - H) = 600 -100H = x (say)

We find that x which is a function of the random variable H, is also a random variable.

We can see that the different values x will take in any outcome are

(600 -100 x 0) =600

(600-1010 x 1) =500

(600-100 x 2 =400

(600-100 x 3) =300

Hence the distribution of x is:

X	P(X)
600	1/8
500	3/8
400	3/8
300	1/8

The above gives you the probability of your getting different sums of money. This may help you in deciding whether you should utilize this opportunity by paying Rs. 500.

12.2 Probability Distribution Function (PDF)

Probability distribution functions can be classified into two categories:

1. Discrete probability distributions
2. Continuous probability distributions

a. A discrete probability distribution assumes that the outcomes of a random variable under study can take on only integer values, such as:
   - A book shop has only 0, 1, 2, ... copies of a particular title of a book
   - A consumer can buy 0, 1, 2, ... shirts, pants, etc.

If the random variable x is discrete, its probability distribution called probability mass function (pmf) must satisfy following two conditions:

i. The probability of a any specific outcome for a discrete random variable must be between 0 and 1. Stated mathematically, $0 \le f\ (x{=}k) \le 1$, for all value of k

ii. The sum of the probabilities over all possible values of a discrete random variable must equal 1. Stated mathematically, all $\sum_{all\ k} f(x = k) = 1$

   a. A continuous probability distribution assumes that the outcomes of a random variable can take on only value in an interval such as:

      If the random variable x is continuous, then its probability density function must satisfy following two conditions:

i) $P(x) \ge 0$; $-\infty < x < \infty$ (non-negativity condition)

ii) $\int_{-\infty}^{\infty} (P\ x\, dx) = 1$ (Area under the continuous curve must total 1

12.3 Discrete Probability Distributions

Binomial Probability Function

Binomial distribution which was discovered by J. Bernoulli (1654-1705) and was first published eight years after his death i.e. in 1713 and is also known as "Bernoulli distribution for n trials". Binomial distribution is applicable for a random experiment comprising a finite number (n) of independent Bernoulli trials having the constant probability of success for each trial.

Before defining binomial distribution, let us consider the following example: Suppose a man fires 3 times independently to hit a target. Let p be the probability of hitting the target (success) for each trial and p (=1–q) be the probability of his failure.

Let S denote the success and F the failure. Let X be the number of successes in 3 trials,

P[X = 0] = Probability that target is not hit at all in any trial

= P [Failure in each of the three trials]

=P (F ∩F∩F)

=P (F) .P (F) .P (F) [trials are independent]

=q.q.q

= q^3

This can be written as

P[X=3]=$3C_3p^3q^{3-3}$ [$b: 3c_3 = 1, q^{3-3} = 1$]

From the above four rectangle results, we can write

P[X=r]=$3C_rp^rq^{3-r}$

which is the probability of r successes in 3 trials. 3Cr, here, is the number of ways in which r successes can happen in 3 trials.

The result can be generalized for n trials in the similar fashion and is given as P[X=r]=$nC_rp^rq^{n-r}$

r=0, 1, 2,..., n. 1)

This distribution is called the binomial probability distribution. The reason behind giving the name binomial probability distribution for this probability distribution is that the probabilities for x = 0, 1, 2, …, n are the respective probabilities $n_{C_0}p^0q^{n-0}$, $n_{C_1}p^1q^{n-1}$, $n_{C_n}p^nq^{n-n}$which are the successive terms of the binomial expansion (q + p)n.

Example 1: An unbiased coin is tossed six times. Find the probability of obtaining

i. exactly 3 heads
ii. less than 3 heads
iii. more than 3 heads

Solution: Let p be the probability of getting head (success) in a toss of the coin and n be the number of trials.

∴ n = 6, p = 1 2 and hence q = 1 – p = 1 – 1/2 = 1/2.

Let X be the number of successes in n trials,

∴ by binomial distribution, we have

P[X=r]=$nC_rp^rq^{n-r}$, r=0, 1, 2,..., n.

$= 6C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{6-x}$ $; x = 1,2, \ldots 6$

$= 6C_x\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{6-0}$ $; x = 1,2, \ldots 6$

= 1/64.$6C_x x = 1,2, \ldots 6$

Therefore,

i) P[exactly 3 heads] = P [X = 3]

= 1/64 ($6C_3$)

= 1/64 (6*5*4/3*2)= 5/16

ii) P[less than 3 heads] = P[X < 3]

=P[X= 2 or X =1 or X =0]

=P[X= 2] + P[X =1]+ P[X =0]

=1/64[$6C_2 + 6C_1 + 6C_0$]

= 1/64[(6*5)/2 +6+1]

=22/64=11/32

iii) P[more than 3 heads] = P[X > 3]

=P[X= 4 or X =5 or X =6]

=P[X= 4] + P[X =5]+ P[X =6]

=1/64[$6C_4 + 6C_5 + 6C_6$]

=1/64[(6*5)/2 +6+1]

= 22/64=11/32

Example 2: The chances of catching cold by workers working in an ice factory during winter are 25%. What is the probability that out of 5 workers 4 or more will catch cold?

Solution: Let catching cold be the success and p be the probability of success for each worker.

∴ Here, n = 5, p = 0.25, q = 0.75 and by binomial distribution

P[X=r]=$nC_r p^r q^{n-r}$, r=0, 1, 2,..., n.

= $5C_r 0.25^r 0.75^{n-r}$,r=0, 1, 2,..., n.

Therefore, the required probability = P[X ≥ 4]

= p [X=4 or X= 5]

= P [X= 4]+ P [X= 5]

= $5C_4 0.25^4 0.75^1 + 5C_4 0.25^4 0.75^0$

=(5)(0.00293)+1 (0.000977)

= 0.014650+ 0.000977

= 0.015627

Characteristics of the Binomial Distribution

The expression (1) is known as binomial distribution with parameters n and p. Different values of n and p identify different binomial distributions which lead to different probabilities of r-values. The mean and standard deviation of a binomial distribution are computed in a shortcut manner as follows

Mean, μ = np,

Standard deviation, σ = $\sqrt{npq}$

Knowing the values of first two central moments μ0 = 1 and μ1 = 1, other central moments are given by

Second moment, μ_2 = npq

Third moment, μ^3 = npq (q – p)

Fourth moment, $\mu^4 = 3n^2p^2q^2$ + npq (1 – 6pq)

Sothat $y_1 = \sqrt{\beta_1} = \frac{u_3}{\mu_2^{3/2}} = \frac{qp}{\sqrt{npq}}$, where $\beta_1 = \frac{n^2p^2q^2(9-p)^2}{n^3p^3q^3}$

and $y_2 = \beta_2 - 3 = \frac{\mu_4}{u_2^2} - 3 = \frac{1-6pq}{npq}$, where $\beta_2 = \frac{3n^2p^2q^2 + npq(1-6pq)^1}{n^2p^2q^2}$

For a binomial distribution, variance < mean. This distribution is unimodal when np is a whole number, and mean = mode = np.

A binomial distribution satisfies both the conditions of pdf, because

$P(x = r) \geq 0$ for all r = 0, 1, 2, . . ., n

$$\sum_{r=0}^{n} p(x = r) = \sum_{r=0}^{n} [n_{C_r} p^r q^{n-r}] = (p+q)^n = 1$$

Fitting a Binomial Distribution

A binomial distribution can be fitted to the observed values in the data set as follows:

i) Find the value of p and q. If one of these is known, the other can be obtained by using the relationship p + q = 1.

ii) Expand $(p + q)^n = p^n + {}^nC_1p^{n-1}q + {}^nC_2p^{n-2}q^2 + \ldots + {}^nC_rp^{n-r}q_r + \ldots + {}^nC_nq^n$ using the concept of binomial theorem.

iii) Multiply each term in the expansion by the total number of frequencies, N, to obtain the expected frequency for each of the random variable value.

The following recurrence relation can be used for fitting of a binomial distribution:

$f(r) = {}^nC_rp^rq^{n-r}$

$f(r+1) = {}^nC_{r+1}p^{r+1}q^{n-r-1}$

Therefore, $\frac{f(r+1)}{f(r)} = \frac{p}{q}\frac{n-r}{r+1}$ or $f(r+1) = \frac{p}{q}\frac{n-r}{r+1} * f(r)$

Example 3: A brokerage survey reports that 30 per cent of individual investors have used a discount broker, i.e. one which does not charge the full commission. In a random sample of 9 individuals, what is the probability that

(a) exactly two of the sampled individuals have used a discount broker?

(b) not more than three have used a discount broker

(c) at least three of them have used a discount broker

Solution: The probability that individual investors have used a discount broker is, p = 0.30, and therefore q = 1 – p = 0.70

a) Probability that exactly 2 of the 9 individual have used a discount broker is given by

$P(x = 2) = {}^9C_2 (0.30)2 (0.70)7 = \frac{9!}{(9-2)!2!}(0 \cdot 30)^2(0.70)^7$

=(9*8)/2*0.09*0.082=0.2656

b) Probability that out of 9 randomly selected individuals not more than three have used a discount broker is given by

$P(x \leq 3) = P(x = 0) + P(x = 1) + P(x=2) + P(x=3)$

$= {}^9C_0 (0.30)^0 (0.70)^9 + {}^9C_1 (0.30) (0.70)^8 + {}^9C_2 (0.30)^2 (0.70)^7 + {}^9C_3 (0.30)^3 (0.70)^6$

= 0.040 + 9× 0.30× 0.058 + 36×0.09×0.082 + 84×0.027×0.118

= 0.040 + 0.157 + 0.266 + 0.268 = 0.731

c) Probability that out of 9 randomly selected individuals at least three have a discount broker is given by

$P(x \geq 3) = 1 - P(x < 3) = 1 - [P(x=0) + P(x=1) + P(x= 2)]$

= 1 – [0.040 + 0.157 + 0.266] = 0.537

Example 4: Mr Gupta applies for a personal loan of Rs 1,50,000 from a nationalized bank to repair his house. The loan offer informed him that over the years bank has received about 2920 loan applications per year and that the probability of approval was, on average, above 0.85.

(a) Mr Gupta wants to know the average and standard deviation of the number of loans approved per year.

(b) Suppose bank actually received 2654 loan applications per year with an approval probability of 0.82. What are the mean and standard deviation now?

Solution: (a) Assuming that approvals are independent from loan to loan, and that all loans have the same 0.85 probability of approval. Then

Mean, $\mu = np = 2920 \times 0.85 = 2482$

Standard deviation, $\sigma = \sqrt{npq} = \sqrt{2920 * 0.85 * 0.15} = 19.295$

(b) Mean, $\mu = np = 2654 \times 0.82 = 2176.28$

Standard deviation, $\sigma = \sqrt{npq} = \sqrt{2654 * 0.82 * 0.18} = 19.792$

Example 5: The incidence of occupational disease in an industry is such that the workers have 20 per cent chance of suffering from it. What is the probability that out of six workers 4 or more will come in contact of the disease?

Solution: The probability of a worker suffering from the disease is, p = 20/100 = 1/5. Therefore q = 1 – p = 1 – (1/5) = 4/5.

The probability of 4 or more, that is, 4, 5, or 6 coming in contact of the disease is given by

$P(x \geq 4) = P(x = 4) + P(x = 5) + P(x = 6)$

$= {}^{6}C_4 \left(\frac{1}{5}\right)^4 \left(\frac{4}{5}\right)^2 + 6C_5 \left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right) + 6C_5 \left(\frac{1}{5}\right)^6$

$= \frac{15*16}{15625} + \frac{6*4}{15625} + \frac{1}{15625} = \frac{1}{15625}(240 + 24 + 1)$

$= \frac{265}{15625}$

= 0.01695

12.4 Poisson Distribution

Poisson distribution is named after the French Mathematician S. Poisson (1781–1840), The Poisson process measures the number of occurrences of a particular outcome of a discrete random variable in a predetermined time interval, space, or volume, for which an average number of occurrences of the outcome is known or can be determined. In the Poisson process, the random variable values need counting. Such a count might be (i) number of telephone calls per hour coming into the switchboard, (ii) number of fatal traffic accidents per week in a city/state, (iii) number of patients arriving at a health centre every hour, (iv) number of organisms per unit volume of some fluid, (v) number of cars waiting for service in a workshop, (vi) number of flaws per unit length of some wire, and so on. The Poisson probability distribution provides a simple, easy-to compute and accurate approximation to a binomial distribution when the probability of success, p is very small and n is large, so that $\mu = np$ is small, preferably np > 7. It is often called the 'law of improbable' events meaning that the probability, p, of a particular event's happening is very small. As mentioned above Poisson distribution occurs in business situations in which there are a few successes against a large number of failures or vice-versa (i.e. few successes in an interval) and has single independent events that are mutually exclusive. Because of this, the probability of success, p is very small in relation to the number of trials n, so we consider only the probability of success.

Definition:

A random variable X is said to follow Poisson distribution if it assumes indefinite number of non-negative integer values and its probability mass function is given by:

$P(x) = P(X=x)=\begin{cases}\frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2,3 \dots. \; and \lambda > 0 \\ \text{or } 0; \text{elsewhere}\end{cases}$

Where e = base of natural logarithm, whose value is approximately equal to 2.7183 corrected to four decimal places.

12.5 Conditions for Poisson Process

The use of Poisson distribution to compute the probability of the occurrence of an outcome during a specific time period is based on the following conditions:

(i) The outcomes within any interval occur randomly and independently of one another.

(ii) The probability of one occurrence in a small time interval is proportional to the length of the interval and independent of the specific time interval.

(iii) The probability of more than one occurrence in a small time interval is negligible when compared to the probability of just one occurrence in the same time interval.

(iv) The average number of occurrences is constant for all time intervals of the same size.

12.6 Characteristics of Poisson Distribution

Since Poisson probability distribution is specified by a process rate λ and the time period t, its mean and variance are identical and are expressed in terms of the parameters: n and p as shown below:

1. The arithmetic mean, $\mu = E(x)$ of Poisson distribution is given by

   $\mu = \sum xP(x) = \sum x\frac{-\lambda\lambda^x}{x}$, $\{x = 0,1,2,3 \dots \; and xP(x) = 0 \; for x = 0.$

   $= \lambda e^{-\lambda} + \lambda^2 e^{-\lambda} + \lambda^2 e^{-\lambda}/2! + \dots\dots\dots \lambda^r e^{-\lambda}/(x-1)! + \dots\dots$

   $= \lambda e^{-\lambda}[1 + \lambda + \lambda^2/2! + \dots\dots + \lambda^{r-1}/(x-1)! + \dots\dots]$

   $= \lambda e^{-\lambda} e^{\lambda}$

   $= \lambda$

   Thus the mean of the distribution is $\mu = \lambda = np$.

2. The variance σ^2 of Poisson distribution is given by

   $\sigma^2 = E(x^2) - [E(x)]^2 = E(x^2) - \lambda^2$

   $\sum x\frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2 = e^{-\lambda}\sum\frac{x(x-1)+x}{x!}\lambda^x - \lambda^2$

   $= \lambda^2 e^{-\lambda}\sum(\lambda^{x-2})/(x-2)! + \lambda e^{-\lambda}\sum(\lambda^{x-1})/(x-1)! - \lambda^2$

   $= \lambda^2 e^{-\lambda}e^{\lambda} + \lambda e^{-\lambda}e^{\lambda} - \lambda^2$

   $= \lambda^2 + \lambda - \lambda^2 = \lambda$

   Thus the variance of the distribution is $\sigma^2 = \lambda = np$.

   The central moments of Poisson distribution can also be determined by the following recursion relation:

   $\mu_r = E(x-\lambda)^r = \sum(x-\lambda)^r = \sum(x-\lambda)^r e^{-\lambda}\lambda^x/x!$

   Differentially μr with respect to λ, we have

   $\frac{d\mu_r}{d\lambda} = -r\mu_{r-1} + \frac{\mu_{r+1}}{\lambda}$ or $\mu_{r+1} = \lambda[r\mu_{r-1} + \frac{\mu_{r+1}}{\lambda}]$

   Substituting $\mu 0 = 1$ and $\mu 1 = 0$ and putting r = 1, 2 and 3 in

   $\mu_2 = \mu_3 = \lambda$

   $\mu_4 = \lambda + 3\lambda^2$

   so that $\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}$

$\gamma_2 = \beta_2 - 3 = 1/\lambda$

Hence Poisson distribution is defined by the parameter λ and is positively skewed and leptokurtic. This implies that there is a possibility of infinitely large number of occurrences in a particular time interval, even though the average rate of occurrences is very small. However, as $\lambda \rightarrow \infty$, the distribution tends to be symmetrical and mesokurtic.

It is very rare for more than one event to occur during a short interval of time. The shorter the duration of interval, the occurrence of two or more events becomes also rare. The probability that exactly one event will occur in such an interval is approximately λ times its duration.

If λ is not an integer and m = [λ], the largest integer contained in it, then m is the unique mode of the distribution. But if λ is an integer, the distribution would be bimodal.

The typical application of Poisson distribution is for analyzing queuing (or waiting line) problems in which arriving customers during an interval of time arrive independently and the number of arrivals depends on the length of the time interval. While applying Poisson distribution if we consider a time period of different length, the distribution of number of events remains Poisson with the mean proportional to the length of the time period.

12.7 Fitting a Poisson Distribution

Poisson distribution can be fitted to the observed values in the data set by simply obtaining values of λ and calculating the probability of zero occurrence. Other probabilities can be calculated by the recurrence relation as follows:

$f(r) = \frac{-\lambda \lambda^x}{x!}$

$f(r+1) = \frac{-\lambda \lambda^{x+1}}{(x+1)!}$

Or $f(x+1)/f(r) = \lambda/(r+1)$

Or $f(r+1) = \lambda/(r+1).f(r)$; r=0,1,2………

Thus, for r = 0, f (1) = λ f (0),

for r = 1,

$f(2) = \lambda/2f(1) = \lambda^2 = \lambda^2/2\ f(0)$

and so on, where $f(0) = e^{-\lambda}$

After obtaining the probability for each of the random variable values, multiply each of them by N (total frequency) to get the expected frequency for the respective values.

Example 6: What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model, what is the probability that a page observed at random will contain at least three misprints?

Solution: Since 100 misprints are distributed randomly throughout the 100 pages of a book, therefore on an average there is only one mistake on a page. This means, the probability of there being a misprint, p = 1/100, is very small and the number of words, n, in 100 pages are vary large. Hence, Poisson distribution is best suited in this case.

Average number of misprints in one page, λ = np = 100×(1/100) = 1. Therefore $e^{-\lambda} = e^{-1} = 0.3679$.

Probability of at least three misprints in a page is

$P(x \geq 3) = 1 - P(x < 3) = 1 - \{P(x = 0)+P(x = 1)+P(x = 2)\}$

$= 1- [e^{-\lambda} + \lambda e^{-\lambda} + 1/2!\, \lambda^\wedge 2e^{-\lambda}]$

$=1- [e^{-1} + e^{-1} + 1/2!\, \lambda^\wedge 2e^{-1}]$

$= 1-2.5e^{-1}$

$= 1-2.5(0.3679)$

$=0.0802$

Example 7: A new automated production process has had an average of 1.5 breakdowns per day. Because of the cost associated with a breakdown, management is concerned about the possibility of having three or more breakdowns during a day. Assume that breakdowns occur randomly, that the probability of a breakdown is the same for any two time intervals of equal length, and that breakdowns in one period are independent of breakdowns in other periods. What is the probability of having three or more breakdowns during a day?

Solution: Given that, $\lambda = np = 1.5$ breakdowns per day. Thus probability of having three or more breakdowns during a day is given by

$P(x \geq 3) = 1 - P(x < 3) = 1 - [P(x= 0) + P(x=1) + P(x=2)]$

$= 1- [e^{-\lambda} + \lambda e^{-\lambda} + 1/2!\, \lambda^2 e^{-\lambda}]$

$= 1- e^{-\lambda}[1 + \lambda + 1/2\lambda^2]$

$=1-.2231[1+1.5+1/2(1.5)^2]$

$=1-0.2231(3.625)$

$=1-0.8088$

$=0.1912$

Example 8: Suppose a life insurance company insures the lives of 5000 persons aged 42. If studies show the probability that any 42-years old person will die in a given year to be 0.001, find the probability that the company will have to pay at least two claims during a given year.

Solution: Given that, n = 5000, p = 0.001, so $\lambda = np = 5000\times0.001 = 5$. Thus the probability that the company will have to pay at least 2 claims during a given year is given by

$P(x \geq 2) = 1 - P(x < 2) = 1 - [P(x = 0) + P(x = 1)]$

$= 1 - [e^{-\lambda} + \lambda e^{-\lambda}] = 1 - [e^{-5} + 5e^{-5}] = 1 - 6e^{-5}$

$=1-6*0.0067=0.9598$

Example 9: A manufacturer who produces medicine bottles, finds that 0.1 per cent of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of bottles. Using Poisson distribution, find how many boxes will contain:

i. no defectives
ii. at least two defectives

Solution: Given that, p = 1 per cent = 0.001, n = 500, $\lambda = np = 500\times0.001 = 0.5$ (i) $P[x = 0] = e - \lambda = e - 0.5 = 0.6065$

Therefore, the required number of boxes are : $0.6065\times100 = 61$ (approx.)

ii) $P(x > 2) = 1 - P(x \leq 1) = 1 - [P(x = 0) + P(x = 1)]$

$= 1 - [e^{\lambda} + \lambda e^{-\lambda}]$

$= 1 - [0.6065 + 0.5(0.6065)]$

$= 1 - 0.6065 (1.5)$

$= 1 - 0.90975 = 0.09025.$

Therefore, the required number of boxes is $100\times0.09025 = 10$ (approx.)

12.8 Continuous Probability Distributions

If a random variable is discrete, then it is possible to assign a specific probability to each of its value and get the probability distribution for it. The sum of all the probabilities associated with the different values of the random variable is 1. However, not all experiments result in random variables that are discrete. Continuous random variables such as height, time, weight, monetary values, length of life of a particular product, etc. can take large number of observable values corresponding to points on a line interval much like the infinite number of gains of sand on a beach. The sum of probability to each of these infinitely large values is no longer sum to 1.

Unlike discrete random variables, continuous random variables do not have probability distribution functions specifying the exact probabilities of their specified values. Instead, probability distribution is created by distributing one unit of probability along the real line, much like distributing a handful of sand along a line. The probability of measurements (e.g. gains of sand) piles up in certain places resulting into a probability distribution called probability density function. Such distribution is used to find probabilities that the random variable falls into a specified interval of values. The depth or density of the probability that varies with the random variable (x) may be described by a mathematical formula.

The probability density function for a continuous random variable x is a curve such that the area under the curve over an interval equals the probability that x falls into that interval, i.e. the probability that x is in that interval can be found by summing the probabilities in that interval. Certain characteristics of probability density function for the continuous random variable, x are follows:

(i) The area under a continuous probability distribution is equal to 1.

(ii) The probability P $(a \leq x \geq b)$ that random variable x value will fall into a particular interval from a to b is equal to the area under the density curve between the points (values) a and b.

Nature seems to follows a predictable pattern for many kinds of measurements. Most numerical values of a random variable are spread around the center, and greater the distance a numerical value has from the center, the fewer numerical values have that specific value. A frequency distribution of values of random variable observed in nature which follows this pattern is approximately bell shaped. A special case of distribution of measurements is called a normal curve (or distribution).

If a population of numerical values follows a normal curve and x is the randomly selected numerical value from the population, then x is said to be normal random variable, which has a normal probability distribution.

12.9 Normal Distribution

Normal distribution is perhaps the most widely used distribution in Statistics and related subjects. It has found applications in inquiries concerning heights and weights of people, IQ scores, errors in measurement, rainfall studies and so on. Abraham de Moivre gave the mathematical equation for the normal distribution in 1733. Karl Friedrich Gauss also independently derived its equation from a study of errors in repeated measurements of the same quantity. Accordingly, sometimes it is also referred to as the Gaussian distribution. The distribution has provided the foundation for much of the subsequent development of mathematical statistics.

Normal distribution provides an adequate representation of a continuous phenomenon or process such as daily changes in the stock market index, frequency of arrivals of customers at a bank, frequency of telephone calls into a switch board, customer servicing times, and so on.

Normal Probability Distribution Function

The formula that generates normal probability distribution is as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{(-1/2)[(x-u)/\sigma]^2}$$

Where,

π = constant 3.1416

e = constant 2.7183

μ = mean of the normal distribution

σ = standard of normal distribution

The f (x) values represent the relative frequencies (height of the curve) within which values of random variable x occur. The graph of a normal probability distribution with mean μ and standard deviation σ is shown in Fig. 9.2. The distribution is symmetric about its mean μ that locates at the centre.

Since the total area under the normal probability distribution is equal to 1, the symmetry implies that the area on either side of μ is 50 per cent or 0.5. The shape of the distribution is determined by μ and σ values.

In symbols, if a random variable x follows normal probability distribution with mean μ and standard deviation σ, then it is also expressed as: $x \sim N(\mu, \sigma)$.

12.10 Normal Distribution/ Normal Probability Curve

Carefully look at the following hypothetical frequency distribution, which a teacher has obtained after examining 150 students of class IX on a Mathematics achievement test.

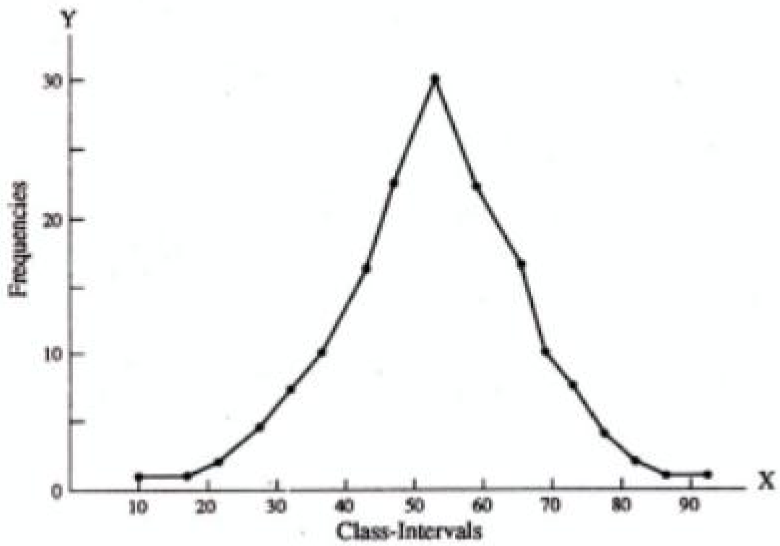
Table 9.1 Frequency distribution of the Mathematics achievement test scores

Class Intervals	Tallies	Frequency
85 – 89	I	1
80 – 84	II	2
75 – 79	IIII	4
70 – 74	𝍸 II	7
65 – 69	𝍸 𝍸	10
60 – 64	𝍸 𝍸 𝍸 I	16
55 – 59	𝍸 𝍸 𝍸 𝍸	20
50 – 54	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
45 – 49	𝍸 𝍸 𝍸 𝍸	20
40 – 44	𝍸 𝍸 𝍸 I	16
35 – 39	𝍸 𝍸	10
30 – 34	𝍸 II	7
25 – 29	IIII	4
20 – 24	II	2
15 – 19	I	1
	Total	**150**

Are you able to find some special trend in the frequencies shown in the column 3 of the above table? Probably yes! The concentration of maximum frequencies (f = 30) lies near a central value of distribution and frequencies gradually tapper off symmetrically on both the sides of this value

Concept of Normal Curve

Now, suppose if we draw a frequency polygon with the help of above distribution, we will have a curve as shown in the fig. 9.1



The shape of the curve in Fig. 9.1 is just like a 'Bell' and is symmetrical on both the sides.

If you compute the values of Mean, Median and Mode, you will find that these three are approximately the same (M = 52; Md = 52 and Mo = 52).

This Bell-shaped curve technically known as Normal Probability Curve or simply Normal Curve and the corresponding frequency distribution of scores, having just the same values of all three measures of central tendency (Mean, Median and Mode) is known as Normal Distribution.

Many variables in the physical (e.g. height, weight, temperature etc.) biological (e.g. age, longevity, blood sugar level and behavioral (e.g. Intelligence; Achievement; Adjustment; Anxiety; Socio-Economic-Status etc.) sciences are normally distributed in the nature. This normal curve has a great significance in mental measurement. Hence to measure such behavioral aspects, the Normal Probability Curve in simple terms Normal Curve worked as reference curve and the unit of measurement is described as σ (Sigma).

Theoretical Base of the Normal Probability Curve

The normal probability curve is based upon the law of Probability (the various games of chance) discovered by French Mathematician Abraham Demoiver (1667-1754). In the eighteenth century, he developed its mathematical equation and graphical representation also.

The law of probability and the normal curve that illustrates it is based upon the law of chance or the probable occurrence of certain events. When anybody of observations conforms to this mathematical form, it can be represented by a bell-shaped curve with definite characteristics.

Characteristics of A Normal Curve

The following are the characteristics of the normal curve.

1. Normal curves are of symmetrical distribution. It means that the left half of the normal curve is a mirror image of the right half. If we were to fold the curve at its highest point at the center, we would create two equal halves.
2. The first and third quartiles of a normal distribution are equidistance from the median.
3. For the curve the mean median and mode all have the same value.
4. In skewed distribution mean median and mode fall at different points.
5. The normal curve is unimodal, having only one peak or point of maximum frequency that point in the middle of the curve.
6. The curve is a asymptotic. It means starting at the centre of the curve and working outward, the height of the curve descends gradually at first then faster and finally slower. An important situation exists at the extreme of the curve. Although the curve descends promptly toward the horizontal axis it never actually touches it. It is therefore said to be asymptotic curve.
7. In the normal curve the highest ordinate is at the centre. All ordinate on both sides of the distribution are smaller than the highest ordinate.
8. A large number of scores fall relatively close to the mean on either side. As the distance from the mean increases, the scores become fewer.
9. The normal curve involves a continuous distribution.

Properties of Normal Distribution Curve (NPC)

The following are the properties of the normal curve:

1. It is a bell shaped curve which is bilaterally symmetrical and has continuous frequency distribution curve.
2. It is a continuous probability distribution for a random variable.
3. It has two halves (right and left) and the value of mean, median and mode are equal (mean = median = mode), that is, they coincide at same point at the middle of the curve.
4. The normal curve is asymptotic, that is, it approaches but never touches the x-axis, as it moves farther from mean.

5. The mean lies in the middle of the curve and divides the curve in to two equal halves. The total area of the normal curve is within z ± 3 σ below and above the mean.
6. The area of unit under the normal curve is said to be equal to one (N=1), standard deviation is one (σ =1), variance is one (σ^2=1) and mean is zero (μ=0).
7. At the points where the curve changes from curving upward to curving downward are called inflection points.
8. The z-scores or the standard scores in NPC towards the right from the mean are positive and towards the left from the mean are negative.
9. About 68% of the curve area falls within the limit of plus or minus one standard deviation (±1 σ) unit from the mean; about 95% of the curve area falls within the limit of plus or minus two standard deviations (±2 σ) unit from the mean and about 99.7% of the curve area falls within the limit of plus or minus three standard deviations (±3 σ) unit from the mean (refer to figure 8.2).
10. The normal distribution is free from skewness, that is, it's coefficient of skewness amounts to zero. λ The fractional areas in between any two given z-scores is identical in both halves of the normal curve, for example, the fractional area between the z-scores of +1 is identical to the z-scores of –1. Further, the height of the ordinates at a particular z-score in both the halves of the normal curve is same, for example, the height of an ordinate at +1z is equal to the height of an ordinate at –1z.

Concept of Standard Score (z-score)

The standard score is a score that informs about the value and also where the value lies in the distribution. Typically, for example, if the value is 5 standard deviations above the mean then it refers to five times the average distance above the mean. It is a transformed score of a raw score. A raw score or sample value is the unchanged score or the direct result of measurement. A raw score (X) or sample value cannot give any information of its position within a distribution. Therefore, these raw scores are transformed in to z-scores to know the location of the original scores in the distribution. The z-scores are also used to standardize an entire distribution.

These scores (z) help compare the results of a test with the "normal" population. Results from tests or surveys have thousands of possible results and units. These results might not be meaningful without getting transformed. For example, if a result shows that height of a particular person is 6.5 feet; such findings can only be meaningful if it is compared to the average height. In such a case, the z-score can provide an idea about where the height of that person is in comparison to the average height of the population.

Properties of z-score

Following are some of the properties of the Standard (z) Score:

1. The mean of the z-scores is always 0.
2. It is also important to note that the standard deviation of the z-scores is always 1.
3. Further, the graph of the z-score distribution always has the same shape as the original distribution of sample values.
4. The z-scores above the value of 0 represent sample values above the mean, while z-scores below the value of 0 represent sample values below the mean.
5. The shape of the distribution of the z-score will be similar or identical to the original distribution of the raw scores. Thus, if the original distribution is normal, then the distribution of the z-score will also be normal. Therefore, converting any data to z-score does not normalize the distribution of that data.

Uses of z-score

Z-scores are useful in the following ways:

1. It helps in identifying the position of observation(s) in a population distribution: As mentioned earlier, the z-scores help in determining the position/distance of a value or an observation from the mean in the units of standard deviations. Further, if the distribution of the scores is like the normal distribution, then we are able to estimate the proportion of the population falling above or below a particular value. Z-score has important implication in the studies related to diet and nutrition of children. It helps in estimating the values of height, weight and age of children with reference to nutrition.
2. It is used for standardizing the raw data: It helps in standardizing or converting the data to enable standard measurements. For example, if you wish to compare your scores on one test with the scores achieved in another test, comparison on the basis of raw score is not possible. In such a situation, comparisons across tests can only be done when you standardize both sets of test scores.
3. It helps in comparing scores that are from different normal distributions: As mentioned in the previous example, z-scores help in comparing scores from different normal distribution. Thus, z-scores can help in comparing the IQ scores received from two different tests.

Standard Normal Probability Distribution:

To deal with problems where the normal probability distribution is applicable more simply, it is necessary that a random variable x is standardized by expressing its value as the number of standard deviations (σ) it lies to the left or right of its mean (μ). The standardized normal random variable, z (also called z–statistic, z–score or normal variate) is defined as:

$z = x - \mu/\sigma$ 1

Or equivalently $x = \mu + z\sigma$

A z–score measures the number of standard deviations that a value of the random variable x falls from the mean. From formula (1) we may conclude that

i. When x is less than the mean (μ), the value of z is negative
ii. When x is more than the mean (μ), the value of z is positive
iii. When x = μ, the value of z = 0.

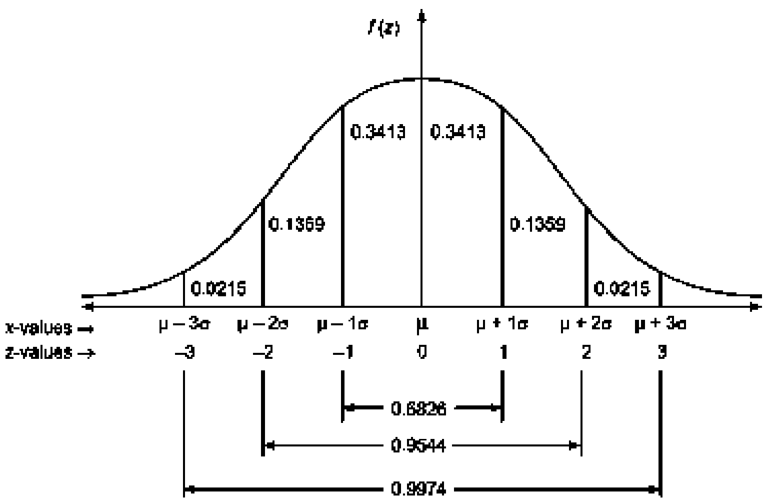


Fig. 9.2 Standard Normal Distribution

Any normal probability distribution with a set of μ and σ value with random variable can be converted into a distribution called standard normal probability distribution z, as shown in Fig. 9.2, with mean μz = 0 and standard deviation σz = 1 with the help of the formula (1).

A z-value measures the distance between a particular value of random variable x and the mean (μ) in units of the standard deviation (σ). With the value of z obtained by using the formula (1), we can find the area or probability of a random variable under the normal curve by referring to the standard distribution in Appendix. For example, z = ± 2 implies that the value of x is 2 standard deviations above or below the mean (μ).

Area Under the Normal Curve

Since the range of normal distribution is infinite in both the directions away from μ, the pdf function f (x) is never equal to zero. As x moves away from μ, f (x) approaches x-axis but never actually touches it.

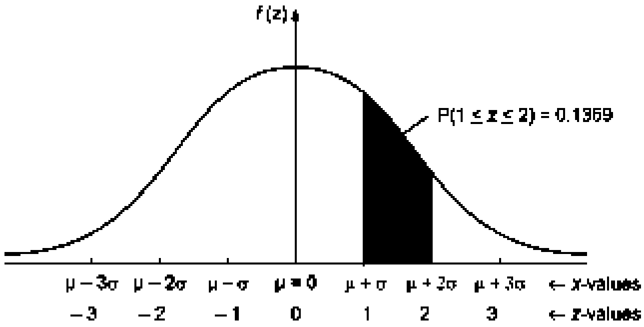


Fig. 9.3: Diagram for Finding P(1 < z < 2)

The area under the standard normal distribution between the mean z = 0 and a specified positive value of z, say z0 is the probability P(0 ≤ z ≤ z 0) and can be read off directly from standard normal (z) tables. For example, area between 1 ≤ z ≤ 2 is the proportion of the area under the curve which lies between the vertical lines erected at two points along the x–axis. For example, as shown in Fig. 9.3, if x is σ away from μ, that is, the distance between x and μ is one standard deviation or (x – μ)/σ = 1, then 34.134 per cent of the distribution lies between x and μ. Similarly, if x is at 2σ away from μ, that is, (x – μ)/σ = 2, then the area will include 47.725 per cent of the distribution, and so on, as shown in Table 9.2.

Table 9.2: Area Under the Normal Curve

z = x − μ/σ	Area Under Normal Curve Between x and μ
1.0	0.34134
2.0	0.47725
3.0	0.49875
4.0	0.49997

Since the normal distribution is symmetrical, Table 9.2 indicates that about 68.26 per cent of the normal distribution lies within the range μ – σ to μ + σ. The other relationships derived from Table 9.2 are shown in Table 9.3

Table 9.3: Percentage of the Area of the Normal Distribution Lying within the Given Range

Number of Standard Deviations from Mean	*Approximate Percentage of Area under Normal Curve*
$x \pm \sigma$	68.26
$x \pm 2\sigma$	95.45
$x \pm 3\sigma$	99.75

The standard normal distribution is a symmetrical distribution and therefore

P(0 ≤ z ≤ a) = P (– a ≤ z ≤ 0) for any value a.

For example, $P(1 \le z \le 2) = P(z \le 2) - P(z \le 1)$

$= 0.9772 - 0.8413 = 0.1359$

Example 5: 1000 light bulbs with a mean life of 120 days are installed in a new factory and their length of life is normally distributed with standard deviation of 20 days.

(a) How many bulbs will expire in less than 90 days?

(b) If it is decided to replace all the bulbs together, what interval should be allowed between replacements if not more than 10% should expire before replacement?

Solution: (a) Given, $\mu = 120$, $\sigma = 20$, and $x = 90$. Then

$z = x - \mu/\sigma$

$= (90\text{-}120)/20$

$= -1.5$

The area under the normal curve between z = 0 and z = – 1.5 is 0.4332. Therefore, area to the left of – 1.5 is 0.5 – 0.4332 = 0.0668. Thus, the expected number of bulbs to expire in less than 90 days will be 0.0668 × 1000 = 67 (approx.).

(b) The value of z corresponding to an area 0.4 (0.5 – 0.10). Under the normal curve is 1.28. Therefore

$z = x - \mu/\sigma$

$= (x\text{-}120)/20$

$= 94$

Hence, the bulbs will have to be replaced after 94 days.

Example 6: The lifetimes of certain kinds of electronic devices have a mean of 300 hours and standard deviation of 25 hours. Assuming that the distribution of these lifetimes, which are measured to the nearest hour, can be approximated closely with a normal curve.

(a) Find the probability that any one of these electronic devices will have a lifetime of more than 350 hours.

(b) What percentage will have lifetimes of 300 hours or less?

(c) What percentage will have lifetimes from 220 or 260 hours?

Solution: (a) Given, $\mu = 300$, $\sigma = 25$, and $x = 350$. Then

$z = x - \mu/\sigma$

$= (350\text{-}300)/25$

$=2$

The area under the normal curve between z = 0 and z = 2 is 0.9772. Thus the required probability is, 1 – 0.9772 = 0.0228.

b. $z = x - \mu/\sigma$

$= (300\text{-}300)/25$

$= 0$

Therefore, the required percentage is, 0.5000 × 100 = 50%.

a. Given, $x1 = 220$, $x2 = 260$, $\mu = 300$ and $\sigma = 25$. Thus

   $Z_1 = x - \mu/\sigma$

   $=(220\text{-}300)/25$

   $=-3.2$

$Z_2 = x - \mu/\sigma$

=(260-300)/25

=-1.6

From the normal table, we have

P(z = – 1.6) = 0.4452 and P(z = – 3.2) = 0.4903

Thus, the required probability is

P(z = – 3.2) – P(z = – 1.6) = 0.4903 – 0.4452 = 0.0541

Hence the required percentage = 0.0541 × 100 = 5.41 per cent

Summary

In this unit, we have discussed the meaning of frequency distribution and probability distribution, and the concepts of random variables and probability distribution. In any uncertain situation, we are often interested in the behavior of certain quantities that take different values in different outcomes of experiments. These quantities are called random variables and a representation that specifies the possible values a random variable can take, together with the associated probabilities, is called a probability distribution. The distribution of a discrete variable is called a discrete probability distribution and the function that specifies a discrete distribution is termed as a probability mass function (p.m.f.). In the discrete distribution we have considered the binomial and poison distributions and discussed how these distributions are helpful in decision-making. We have shown the fitting of such distributions to a given observed data. In the final section, we have examined situations involving continuous random variables and the resulting probability distributions. The random variable which can take an infinite number of values is called a continuous random variable and the probability distribution of such a variable is called a continuous probability distribution. The function that specifies such distribution is called the probability density function (p.d.f.). One such important distribution, viz., the normal distribution has been presented and we have seen how probability calculations can be done for this distribution.

Keywords

1. Binomial Distribution: It is a type of discrete probability distribution function that includes an event that has only two outcomes (success or failure) and all the trials are mutually independent.
2. Discrete Probability Distribution: A probability distribution in which the variable is allowed to take on only a limited number of values.
3. Probability Distribution: A curve that shows all the values that the random variable can take and the likelihood that each will occur.
4. Poisson Distribution: The Poisson distribution is a frequency distribution of a discrete event occurring rarely.
5. Mean of Poisson Distribution: The mean of the Poisson distribution is μ and standard deviation is$\sqrt{\mu}$.
6. Normal Probability Curve: A normal curve is a bell shaped curve, bilaterally symmetrical and continuous frequency distribution curve.
7. Normal Probability Distribution: A continuous probability distribution for a variable is called as normal probability distribution or simply normal
8. Standard score: Standard score or z-score is a transformed score which shows the number of standard deviation units by which the value of observation (the raw score) is above or below the mean.

Self Assessment

1. Which of the following is correct for a binomial distribution?
A. variance < mean
B. variance = mean
C. variance > mean
D. variance ≥ mean

2. When the value of p is ______, then the distribution is skewed to the left.
A. Less than 0.3
B. Less than 0.5
C. More than 0.3
D. More than 0.5

3. If q = 0.13 and n = 50 then the approximate values of mean and standard deviation are
A. 43; 2.38
B. 43; 2.36
C. 44; 2.38
D. 44;2.36

4. If the P (success) of an event is 0.4, then the P (success in at least one trail) out of five trails is
A. 0.6
B. 0.07
C. 0.26
D. 0.92

5. If p = 1/5, n = 5 then the value of P(x = 3) is
A. $10\ (4)^3(1/5)^2$
B. $10\ (4)^2\ (1/5)^5$
C. $10\ (4/5)^5$
D. $10\ (4/5)^3$

6. In binomial distributions, the formula for calculating standard deviation is:
A. square root of p
B. square root of np
C. square root of pq
D. square root of npq

7. For larger values of 'n', Binomial Distribution ___________
A. loses its discreteness
B. tends to Poisson Distribution
C. stays as it is
D. gives oscillatory values

8. Binomial Distribution is a ___________
A. Continuous distribution

B. Discrete distribution

C. Irregular distribution

D. Not a Probability distribution

9. If 'm' is the mean of Poisson Distribution, the P(0) is given by ___________

A. e^{-m}

B. e^{m}

C. e

D. m^{-e}

10. In a Poisson Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by?

A. m = np

B. $m = (np)^2$

C. m = np(1-p)

D. m = p

11. In a binomial distribution if n is fixed and p > 0.5, then

A. the distribution will be skewed to left

B. the distribution will be skewed to right

C. the distribution will be symmetric

D. cannot say anything

12. The standard deviation of the binomial distribution is:

A. np

B. $\sqrt{np}$

C. npq

D. (d)$\sqrt{npq}$

13. Which of the following a necessary condition for use of apoison distribution?

A. Probability of an event in a short interval of time isconstant

B. The number of events in any interval of time isindependent of successes in other intervals.

C. Probability of two or more events in the short intervalof time is zero

D. All of these

14. For a Poisson distribution $P(x) = (5)^2e^{-5}/2!$, the mean valueis:

A. 2

B. 5

C. 10

D. none of the above

15. For a binomial distribution $P(x) = {}^{10}C_r\,(0.5)^r\,(0.5)^{10-r}$,r = 0, 1, 2, . . ., 10, the mean value is:

A. 4

B. 5

C. 10

D. 15

16. A frequency distribution is said to be platykurtic if

A. -0.263 <Ku < 0.263

B. Ku < 0.263

C. Ku = 0.263

D. Ku > 0.263

17. Normal curve has significance in the __________.

1. Mental measurement

2. Educational evaluation

A. only option 1

B. only option 2

C. both options 1 and 2

D. neither option 1 nor 2

18. If N = 125, μ = 10, X = 24 and σ = 8, the z-value will be

A. -1.75

B. 1.70

C. 1.75

D. 1.74

19. If the total number of cases lie between 0 and+1 are 3413 and total cases lie between mean (0) and +0.5 = 1915, then the percentage of cases between+0.5 and +1 are

A. 53.28%

B. 14.98%

C. 13.28%

D. 14.89%

20. If there are an odd number of categories, then in the normal probability curve the middle category will be

A. Immediately on the right side of the mean

B. Immediately on the left side of the mean

C. Half on the immediate left and half on the immediate right side of the mean

D. Coincides with the mean

Answers for Self Assessment

1.	A	2.	D	3.	C	4.	D	5.	B
6.	D	7.	B	8.	B	9.	A	10.	A
11.	A	12.	D	13.	D	14.	B	15.	B
16.	D	17.	C	18.	C	19.	B	20.	C

Review Questions

1. Define binomial distribution stating its parameters, mean, and standard deviation, and give two examples where such a distribution is ideally suited.
2. What information is provided by the mean, standard deviation, and central moments of the binomial distribution?
3. The normal rate of infection of a certain disease in animals is known to be 25 per cent. In an experiment with 6 animals injected with a new vaccine it was observed that none of the animals caught the infection. Calculate the probability of the observed result.
4. The incidence of a certain disease is such that on an average 20 per cent of workers suffer from it. If 10 workers are selected at random, find the probability that (i) exactly 2 workers suffer from the disease, (ii) not more than 2 workers suffer from the disease.
   Calculate the probability upto fourth decimal place.
5. A supposed coffee connoisseur claims that he can distinguish between a cup of instant coffee and a cup of percolator coffee 75 per cent of the time. It is agreed that his claim will be accepted if he correctly identifies at least 5 out of 6 cups. Find (a) his chance of having the claim accepted if he is in fact only guessing, and (b) his chance of having the claim rejected when he does have the ability he claims.
6. What is Poisson distribution? Point out its role in business decision-making. Under what conditions will it tend to become a binomial distribution?
7. Discuss the distinctive features of Poisson distribution. When does a binomial distribution tend to become a Poisson distribution?
8. The following table shows the number of customers returning the products in a marketing territory. The data is for 100 stores:
   No. of returns : 0 1 2 3 4 5 6
   No. of stores : 4 14 23 23 18 9 9
   Fit a Poisson distribution.
9. One-fifth per cent of the blades produced by a blade manufacturing factory turn out to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective, and two defective blades respectively in a consignment of 1,00,000 packets
10. Suppose a population contains 10 elements, 6 of which are defective. A sample of 3 elements is selected. What is the probability that exactly 2 are defective?

Further Readings

- Hoel, P (1962), Introduction to Mathematical Statistics, Wiley John & Sons, New York.
- Hoel, Paul G. (1971), Introduction to Probability Theory, Universal Book Stall, New Delhi.
- Olkin, I., L.J. Gleser, and C. Derman (1980), Probability Models and Applications,Macmillan Publishing, New York.

Unit 13: Correlation

CONTENTS

Objective

- Understand the meaning of correlation
- Recognize the various types of correlation

Introduction

The statistical methods, discussed so far, are used to analyze the data involving only one variable. Often an analysis of data concerning two or more quantitative variables is needed to look for any statistical relationship or association between them that can describe specific numerical features of the association. The knowledge of such a relationship is important to make inferences from the relationship between variables in a given situation.Examples of correlation problems are found in the study of the relationship between IQ and aggregate percentage marks obtained by a person in SSC examination, blood pressure and metabolism or the relation between height and weight of individuals. In these examples both variables are observed as they naturally occur, since neither variable is fixed at predetermined levels.

Meaning of Correlation

Correlation refers to the associations between variables. When an association exists between two variables, it means that the average value of one variable changes as there is a change in the value of the other variable. A correlation is the simplest type of association. When a correlation is weak, it means that the average value of one variable change only slightly (only occasionally) in response to changes in the other variable. If there is no association, it means that there is nochange in the value of one variable in response to the changes in the other variable. In some cases, the correlation may be positive or it may be negative. A positive correlation means that as one variable increases the other variable increases, e.g. Height of a child and age of the child. Negative correlation implies as one variable increases the other variable decrease, e.g. value of a car and age of the car.

A statistical technique that is used to analyze the strength and direction of the relationship between two quantitative variables is called correlation analysis. A few definitions of correlation analysis are:

An analysis of the relationship of two or more variables is usually called correlation.

— A. M. Tuttle

When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation.

— Croxton and Cowden

The coefficient of correlation is a number that indicates the strength (magnitude) and direction of statistical relationship between two variables.

- The strength of the relationship is determined by the closeness of the points to a straight line when a pair of values of two variables is plotted on a graph. A straight line is used as the frame of reference for evaluating the relationship.
- The direction is determined by whether one variable generally increases or decreases when the other variable increases.

The importance of examining the statistical relationship between two or more variables can be divided into the following questions and accordingly requires the statistical methods to answer these questions:

i. Is there an association between two or more variables? If yes, what is form and degree of that relationship?

ii. Is the relationship strong or significant enough to be useful to arrive at a desirable conclusion?

iii. Can the relationship be used for predictive purposes, that is, to predict the most likely value of a dependent variable corresponding to the given value of independent variable or variables?

Significance of Measuring Correlation

The objective of any scientific and clinical research is to establish relationships between two or more sets of observations or variables to arrive at some conclusion which is also near to reality. Finding such relationships is often an initial step for identifying causal relationships. Few advantages of measuring an association (or correlation) between two or more variables are as under:

1. Correlation analysis contributes to the understanding of economic behavior, aids in locating the critically important variables on which others depend, may reveal to the economist the connections by which disturbances spread and suggest to him the paths through which stabilizing forces may become effective. —W. A. Neiswanger

2. The effect of correlation is to reduce the range of uncertainty of our prediction. The prediction based on correlation analysis will be more reliable and near to reality. — Tippett

3. In economic theory we come across several types of variables which show some kind of relationship. For example, there exists a relationship between price, supply, and quantity demanded; convenience, amenities, and service standards are related to customer retention; yield of a crop related to quantity of fertilizer applied, type of soil, quality of seeds, rainfall, and so on. Correlation analysis helps in quantifying precisely the degree of association and direction of such relationships.

4. Correlations are useful in the areas of healthcare such as determining the validity and reliability of clinical measures or in expressing how health problems are related to certain biological or environmental factors. For example, correlation coefficient can be used to determine the degree of inter-observer reliability for two doctors who are assessing a patient's disease.

13.1 Types of Correlations

There are three broad types of correlations:

1. Positive and negative,

2. Linear and non-linear,

3. Simple, partial, and multiple.

1. ***Positive Correlation***: If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales. Some other examples of series of positive correlation are:

i. Price and supply of commodities;

ii. Amount of rainfall and yield of crops.
iii. Heights and weights;
iv. Household income and expenditure.

1.1 Negative Correlation: If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations and cinema attendance.

Some other examples of series of negative correlation are:

i. Volume and pressure of perfect gas;
ii. Current and resistance [keeping the voltage constant]
iii. Price and demand for goods.

2. Linear Correlation: A linear correlation implies a constant change in one of the variable values with respect to a change in the corresponding values of another variable. In other words, a correlation is referred to as linear correlation when variations in the values of two variables have a constant ratio. The following example illustrates a linear correlation between two variables x and y.

x : 10 20 30 40 50
y : 40 60 80 100 120

When these pairs of values of x and y are plotted on a graph paper, the line joining these points would be a straight line.

In general, two variables x and y are said to be linearly related, if there exists a relationship of the form.

$y = a + bx$

Where 'a' and 'b' are real numbers. This is nothing but a straight line when plotted on a graph sheet with different values of x and y and for constant values of a and b. Such relations generally occur in physical sciences but are rarely encountered in economic and social sciences.

2.1 Non-linear Relationship: A non-linear (or curvi-linear) correlation implies an absolute change in one of the variable values with respect to changes in values of another variable. In other words, a correlation is referred to as a non-linear correlation when the amount of change in the values of one variable does not bear a constant ratio to the amount of change in the corresponding values of another variable. The following example illustrates a non-linear correlation between two variables x and y.

x: 8 9 9 10 10 28 29 30
y: 80 130 170 150 230 560 460 600

When these pair of values of x and y are plotted on a graph paper, the line joining these points would not be a straight line, rather it would be curvi-linear.

In other words, the relationship between two variables is said to be non – linear if corresponding to a unit change in one variable, the other variable does not change at a constant rate but changes at a fluctuating rate. In such cases, if the data is plotted on a graph sheet we will not get a straight line curve. For example, one may have a relation of the form.

$y = a + bx + cx2$

3. Simple Correlation: The distinction between simple, partial, and multiple correlations is based upon the number of variables involved in the correlation analysis. If only two variables are chosen to study correlation between them, then such a correlation is referred to as simple correlation. A study on the yield of a crop with respect to only amount of fertilizer, or sales revenue with respect to amount of money spent on advertisement, are a few examples of simple correlation.

3.1 Partial Correlation: In partial correlation, two variables are chosen to study the correlation between them, but the effect of other influencing variables is kept constant. For example (i) yield of a crop is influenced by the amount of fertilizer applied, rainfall, quality of seed, type of soil, and pesticides, (ii) sales revenue from a product is influenced by the level of advertising expenditure, quality of the product, price, competitors, distribution, and so on. In such cases an attempt to measure the correlation between yield and seed quality, assuming that the average values of other factors exist, becomes a problem of partial correlation.

3.2 Multiple Correlations: In multiple correlations, the relationship between more than three variables is considered simultaneously for study. For example, employer-employee relationship in any organization may be examined with reference to, training and development facilities; medical, housing, and education to children's facilities; salary structure; grievances handling system; and so on.

Degrees of Correlation

Through the coefficient of correlation, we can measure the degree or extent of the correlation between two variables. On the basis of the coefficient of correlation we can also determine whether the correlation is positive or negative and also its degree or extent.

1. Perfect correlation: If two variables change in the same direction and in the same proportion, the correlation between the two is perfect positive. According to Karl Pearson the coefficient of correlation in this case is +1. On the other hand, if the variables change in the opposite direction and in the same proportion, the correlation is perfect negative. Its coefficient of correlation is -1. In practice we rarely come across these types of correlations.

2. Absence of correlation: If two series of two variables exhibit no relations between them or change in one variable does not lead to a change in the other variable, then we can firmly say that there is no correlation or absurd correlation between the two variables. In such a case the coefficient of correlation is 0.

3. Limited degrees of correlation: If two variables are not perfectly correlated or there is a perfect absence of correlation, then we term the correlation as Limited correlation.

Thus Correlation may be positive, negative or zero but lies with the limits ± 1. i.e. the value of r is such that $-1 \le r \le +1$. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.

1. If x and y have a strong positive linear correlation, r is close to +1. An r value of exactly +1 indicates a perfect positive correlation.
2. If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative correlation.
3. If there is no linear correlation or a weak linear correlation, r is close to 0.

Table 1: Degree and Types of Correlation

Degrees	Positive	Negative
Absence of correlation →	Zero	Zero
Perfect correlation →	+ 1	–1
High degree →	+ 0.75 to + 1	– 0.75 to –1
Moderate degree →	+ 0.25 to + 0.75	– 0.25 to – 0.75
Low degree →	0 to 0.25	0 to – 0.25

Note that r is a dimensionless quantity; that is, it does not depend on the units employed.

13.2 Properties of Correlation

1. Coefficient of Correlation lies between -1 and +1:

The coefficient of correlation cannot take value less than -1 or more than one +1. Symbolically,

-1<=r<= + 1 or | r | <1

2. Coefficients of Correlation are independent of Change of Origin:

This property reveals that if we subtract any constant from all the values of X and Y, it will not affect the coefficient of correlation.

3. Coefficients of Correlation possess the property of symmetry: The degree of relationship between two variables is symmetric as shown below:

4. Coefficient of Correlation is independent of Change of Scale:

This property reveals that if we divide or multiply all the values of X and Y, it will not affect the coefficient of correlation.

5. Co-efficient of correlation measures only linear correlation between X and Y.

6. If two variables X and Y are independent, coefficient of correlation between them will be zero.

13.3 Methods of Correlation

The following methods of finding the correlation coefficient between two variables x and y are discussed:

1. Scatter Diagram method

2. Karl Pearson's Coefficient of Correlation method

3. Spearman's Rank Correlation method

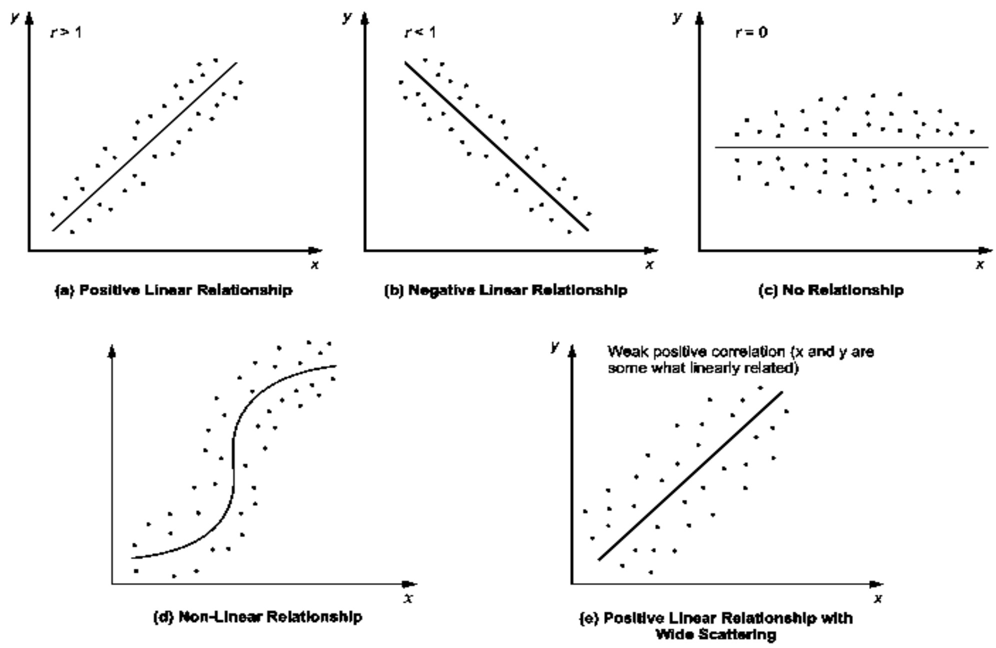
1. Scatter Diagram Method

The scatter diagram method is a quick at-a-glance method of determining of an apparent relationship between two variables, if any. A scatter diagram (or a graph) can be obtained on a graph paper by plotting observed (or known) pairs of values of variables x and y, taking the independent variable values on the x-axis and the dependent variable values on the y-axis.

It is common to try to draw a straight line through data points so that an equal number of points lie on either side of the line. The relationship between two variables x and y described by the data points is defined by this straight line.

In a scatter diagram the horizontal and vertical axes are scaled in units corresponding to the variables x and y, respectively. The pattern of data points in the diagram indicates that the variables are related. If the variables are related, then the dotted line appearing in each diagram describes relationship between the two variables.

Fig 1: Various types of Correlation Coefficient



The patterns depicted in Fig. 1 (a) and (b) represent linear relationships since the patterns are described by straight lines. The pattern in Fig. 1 (a) shows a positive relationship since the value of y tends to increase as the value of x increases, whereas pattern in Fig. 1(b) shows a negative relationship since the value of y tends to decrease as the value of x increases. The pattern depicted in Fig. 1(c) illustrates very low or no relationship between the values of x and y, whereas Fig. 1(d) represents a curvilinear relationship since it is described by a curve rather than a straight line. Figure 13.2(e) illustrates a positive linear relationship with a widely scattered pattern of points. The wider scattering indicates that there is a lower degree of association between the two variables x and y than there is in Fig. 1

2. Karl Pearson's coefficient of correlation

It gives the precise numerical expression for the measure of correlation. It is denoted by 'r'. The value of 'r' gives the magnitude of correlation and its sign denotes its direction. The mathematical formula for computing r is:

$$p = \frac{\Sigma xy}{N\sigma x\sigma y} \qquad 1)$$

Where x= (X -$\bar{X}$), y =(Y -$\bar{Y}$), σx = s.d. of X, $\sigma y = s.d\ of\ Y$

AndN number of pairs of observations

Since $\sigma x = \sqrt{\frac{\Sigma x^2}{N}}$ and $\sigma y = \sqrt{\frac{\Sigma y^2}{N}}$

So,Equation 1 can be written as

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

By Using Actual Mean

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

By Assumed Mean Method

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

By direct method

$$r = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2 \cdot N\sum Y^2 - (\sum Y)\text{^}2}}$$

Now covariance of X and Y is defined as

cov(X,Y)= $(\Sigma(X-\bar{X})(Y-\bar{Y}))/N$

Therefore, r= cov(X,y)/$\sigma x\sigma y$

Where N is the number of pairs of data.

d_x= X – A_X

d_y= Y – A_Y

Example 1: Example 1: Calculate the coefficient of correlation between the expenditure on advertising and sales of the company from the following data

Advertising Expenditure	165	166	167	168	167	169	170	172
Sales (in Lakh)	167	168	165	172	168	172	169	171

Solution: N = 8 (pairs of observations)

Advertising Expenditure	Sales (in Lakh)	x=Xi - $\bar{X}$	y= Yi - $\bar{Y}$	xy	x^2	y^2
165	167	-3	-2	6	9	4

166	168	-2	-1	2	4	1
167	165	-1	-4	4	1	16
167	168	-1	-1	1	1	1
168	172	0	3	0	0	9
169	172	1	3	3	1	9
170	169	2	0	0	4	0
172	171	4	2	8	16	4

Calculation:

$\bar{X} = \frac{\Sigma X_i}{N}$ = 1344/8= 168 cm

$$\sigma x = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{36}{8}}$$

$\bar{Y} = \frac{\Sigma Y_i}{N}$= 1352/8= 169 cm

$\sigma y = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{44}{8}}$

Now, r= cov(X,y)/$\sigma x \sigma y$

= 24/8 * $\sqrt{\frac{36}{8}} * \sqrt{\frac{44}{8}} = \frac{24}{\sqrt{36 \cdot 44}} =$ +0.6029

Since r is positive and 0.6. This shows that the correlation is positive and moderate (i.e. direct and reasonably good).

Example 2: The following data relates to the Cost and Sales of a Company for the past 10 months

Cost (in 000 rupees)	165	166	167	168	167	169	170	172
Sales (in 000 rupees)	167	168	165	172	168	172	169	171

Find the coefficient of correlation between the two

Solution: Here A = 60, h = 4, B = 60 and k = 3

Cost (in 000 rupees)	Sales (in 000 rupees)	u=(Xi - A)/h	v=(Yi - B)/d	uv	u^2	v^2
44	48	-4	-4	16	16	16
80	75	5	5	25	25	25
76	54	4	-2	-8	16	4
48	60	-3	0	-2	4	1
52	63	-2	1	-2	4	1
72	69	3	3	9	9	9

68	72	2	4	8	4	16
56	51	-1	-3	3	1	9
60	57	0	-1	0	0	1
64	66	1	2	2	4	4
		Σu = 5	Σv = 5	Σuv = 53	$\Sigma u^2 = 85$	$\Sigma v^2 = 85$

Calculation

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

$$r = \frac{53 - \frac{5*5}{10}}{\sqrt{85 - \frac{(5)^2}{10}}\sqrt{85 - \frac{(5)^2}{10}}}$$

$$r = \frac{53 - 2.5}{\sqrt{82.5}\sqrt{82.5}}$$

r= 50.5/82.5= 0.61

Assumptions of Using Pearson's Correlation Coefficient

i. Pearson's correlation coefficient is appropriate to calculate when both variables x and y are measured on an interval or a ratio scale.
ii. Both variables x and y are normally distributed, and that there is a linear relationship between these variables.
iii. The correlation coefficient is largely affected due to truncation of the range of values in one or both of the variables. This occurs when the distributions of both the variables greatly deviate from the normal shape.
iv. There is a cause-and-effect relationship between two variables that influences the distributions of both the variables. Otherwise, correlation coefficient might either be extremely low or even zero.

Advantage and Disadvantages of Pearson's Correlation Coefficient

The correlation coefficient is a numerical number between – 1 and 1 that summarizes the magnitude as wellas direction (positive or negative) of association between two variables. The chief limitations of Pearson's method are:

i. The correlation coefficient always assumes a linear relationship between two variables, whether it is true or not.
ii. Great care must be exercised in interpreting the value of this coefficient as very often its value is misinterpreted.
iii. The value of the coefficient is unduly affected by the extreme values of two variable values.
iv. As compared with other methods the computational time required to calculate the value of r using Pearson's method is lengthy.

3. Spearman's Rank Correlation Coefficient

This method is based on the ranks of the items rather than on their actual values. The advantage of this method over the others in that it can be used even when the actual values of items are unknown. For example, if you want to know the correlation between honesty and wisdom of the boys of your class, you can use this method by giving ranks to the boys. It can also be used to find the degree of agreements between the judgments of two examiners or two judges. The formula is:

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

Where,

R = Rank correlation coefficient

D = Difference between the ranks of two items

N = the number of observations.

Note: –1 ≤ R ≤ 1.

i. When R = +1 ⇒ Perfect positive correlation or complete agreement in the same direction
ii. When R = –1 ⇒ Perfect negative correlation or complete agreement in the opposite direction.
iii. When R = 0 ⇒ No Correlation.

Example 3: Calculate 'R' of 6 students from the following data

Student No.:	1	2	3	4	5	6	7	8	9	10
Rank in Maths	1	3	7	5	4	6	2	10	9	8
Rank in Stata	3	1	4	5	6	9	7	8	10	2

Solution:

Student no.	Rank in Maths (R_1)	Rank in Stats(R_2)	D= (R_1- R_2)	D^2
1	1	3	-2	4
2	3	1	2	4
3	7	4	3	9
4	5	5	0	0
5	4	6	-2	4
6	6	9	-3	9
7	2	7	-5	25
8	10	8	2	4
9	9	10	-1	1
10	8	2	6	36
N=10			ΣD = 0	ΣD^2 = 96

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

$$R = 1 - \frac{6 * 96}{100(100^2 - 1)}$$

$$R = 1 - \frac{6 * 96}{10 * 99}$$

= 0.4181

Example 4: The value of Spearman's rank correlation coefficient for a certain number of pairs of observations was found to be 2/3. The sum of the squares of difference between the corresponding marks was 55. Find the number of pairs.

Solution: We have

$R = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$ but R=2/3 and ΣD^2=55

Therefore: $2/3 = 1 - \frac{6*55}{N(N^2-1)}$

$-1/3 = -\frac{6*55}{N(N^2-1)}$

$N(N^2 - 1)\ 6 \times 55$

Now $N(N^2 - 1) = 990$

$\therefore N(N^2 - 1) = 10 \times 99 = 10(100 - 1)$

$\therefore N(N^2 - 1) = 10(102 - 1) \Rightarrow N = 10$

Therefore, there were 10 students.

Advantages and Disadvantages of Spearman's Correlation Coefficient Method

Advantages

i. This method is easy to understand and its application is simpler than Pearson's method.
ii. This method is useful for correlation analysis when variables are expressed in qualitative terms like beauty, intelligence, honesty, efficiency, and so on.
iii. This method is appropriate to measure the association between two variables if the data type is at least ordinal scaled (ranked)
iv. The sample data of values of two variables is converted into ranks either in ascending order or descending order for calculating degree of correlation between two variables.

Disadvantages

i. Values of both variables are assumed to be normally distributed and describing a linear relationship rather than non-linear relationship.
ii. A large computational time is required when number of pairs of values of two variables exceed 30.
iii. This method cannot be applied to measure the association between two variable grouped data.

Summary

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

Keywords

1. Correlation: Degree of association between two variables.
2. Correlation Coefficient: A number lying between -1 (Perfect negative correlation) and + i (perfect positive correlation) to quantify the association between two variables.
3. Scatter Diagram: An ungrouped plot of two variables, on the X and Y axes.

4. Rank Correlation: There happen to be many occasions when it may Coefficient not be convenient, economic or even possible to give values to variables. However, various items can be ranked. In such cases, a rank correlation coefficient may be used.
5. Positive Correlation: If two variables change in the same direction.
6. Multiple Correlations: In multiple correlation, the relationship between more than three variables is considered simultaneously for study.

SelfAssessment

1. Correlation analysis is a ....
A. Univariate analysis
B. Bivariate analysis
C. Multivariate analysis
D. Both b and c

2. If the ratio of change in one variable is equal to the ratio of change in the other variable, then the correlation is said to be ....
A. Linear
B. Non-linear
C. Curvilinear
D. None of these

3. Regression coefficient is independent of...........................
A. Origin
B. Scale
C. Both a and b
D. Neither origin nor scale

4. Study of correlation among three or more variables simultaneously is called.............
A. Partial correlation
B. Multiple correlation
C. Nonsense correlation
D. Simple correlation

5. The unit of Coefficient of correlation is ....
A. Percentage
B. Ratio
C. Same unit of the data
D. No unit

6. Correlation analysis between one dependent variable with one independent variable by keeping the other independent variables as constant is called......................
A. Partial correlation
B. Multiple correlation
C. Nonsense correlation
D. Simple correlation

7. In a correlation analysis, if r= 0, then we may say that there is .... between variables.

A. No correlation

B. Linear correlation

C. Perfect correlation

D. none of these

8. The coefficient of correlation is independent of

A. Change of scale only

B. Change of origin only

C. Both change of origin and scale

D. Neither change of origin nor change of scale

9. If the correlation coefficient between the variables X and Y is ρ, the correlation coefficient between X2 and Y2 is

A. ρ

B. ρ2

C. 0

D. 1

10. Spearman’s Rank Correlation Coefficient is usually denoted by....................

A. k

B. r

C. S

D. R

11. Pearsonian correlation coefficient if denoted by the symbol ....

A. K

B. r

C. R

D. None of these

12. If the dots in a scatter diagram fall on a narrow band, it indicates a ....degree of correlation.

A. Zero

B. High

C. Low

D. None of these

13. If all the dots of a scatter diagram lie on a straight line falling from left bottom corner to the right upper corner, the correlation is called..................

A. Zero correlation

B. High degree of positive correlation

C. Perfect negative correlation

D. Perfect positive correlation

14. Which of following is/ are characteristics of Karl Pearson's coefficient of correlation

A. Indication of degree
B. Indicators of the direction
C. A satisfactory measure
D. All of the above

15. Karl Pearson coefficient of Correlation between two variables is
A. The product of their standard deviation
B. The square root of the product of their regression coefficients
C. The co-variance between the variables
D. None of the above

Answers for Self Assessment

1.	D	2.	A	3.	A	4.	B	5.	D
6.	A	7.	A	8.	C	9.	B	10.	D
11.	B	12.	B	13.	D	14.	D	15.	B

Review Questions

1. The data relating to variable X and Y is given below:

X	72	73	75	76	77	78	79	80	80	81	82	83	84	85	86	88
Y	45	38	41	35	31	40	25	32	36	29	34	38	26	32	28	27

(a) Sketch a scatter plot.

(b) Compute the correlation coefficient, r.

2. Calculate and analyze the correlation coefficient between the number of study hours and the number of sleeping hours of different students.

Number of Study hours	2	4	6	8	10
Number of sleeping hours	10	9	8	7	6

3. A trainee manager wondered whether the length of time his trainees revised for an examination had any effect on the marks they scored in the examination. Before the exam, he asked a random sample of them to honestly estimate how long, to the nearest hour, they had spent revising. After the examination he investigated the relationship between the two variables.

Trainee	A	B	C	D	E	F	G	H	I	J
Revision time	4	9	10	14	4	7	12	22	1	17
Exam mark	31	58	65	73	37	44	60	91	21	84

(a) Plot the scatter diagram in order to inspect the data.

(b) Calculate the correlation coefficient

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

5. The coefficient of correlation between two variables x and y is 0.3. The covariance is 9. The variance of x is 16. Find the standard deviation of y series.

6. The correlation between the price of two commodities x and y in a sample of 60 is 0.68. Could the observed value have arisen?

(a) from an uncorrelated population?

(b) from a population in which true correlation was 0.8?

7. A small retail business has determined that the correlation coefficient between monthly expenses and profits for the past year, measured at the end of each month, is r = 0.56. Assuming that both expenses and profits are approximately normal, test at α = 0.05 level of significance the null hypothesis that there is no correlation between them.

8. Define correlation coefficient 'r' and give its limitations. What interpretation would you give if told that the correlation between the number of truck accidents per year and the age of the driver is (–) 0.60 if only drivers with at least one accident are considered?

9. Does correlation always signify a cause-and-effect relationship between the variables?

10. What is coefficient of rank correlation? Bring out its usefulness. How does this coefficient differ from the coefficient of correlation?

Further Readings

- Nagar, A.L. and R.K Das, 1989 : Basic Statistics, Oxford University Press, Delhi.
- Goon, A.M., M.K. Gupta and B.'Dasgupta, 19.87 : Basic Statistics, The WorldPress Pvt. Ltd., Calcutta.

Web Links

- Edwards, B. 1980. The Readable Maths and Statistics Book, George Allen and Unwin: London.
- Makridakis, S. and S. Wheelwright, 1978. Interactive Forecasting: Univariate and Multivariate Methods, Holden-Day: San Francisco.

Unit 14: Regression

CONTENTS

Objective

understand the meaning of correlation

Calculate a regression line that allows to predict the value of one of the variables if the value of the other variable is known

Introduction

We introduced the concept of statistical relationship between two variables such as: level of sales and amount of advertising; yield of a crop and the amount of fertilizer used; price of a product and its supply, and so on. The relationship between such variables indicates the degree and direction of their association, but fail to answer following question:

• Is there any functional (or algebraic) relationship between two variables? If yes, can it be used to estimate the most likely value of one variable, given the value of other variable? The statistical technique that expresses the relationship between two or more variables in the form of an equation to estimate the value of a variable, based on the given value of another variable, is called regression analysis. The variable whose value is estimated using the algebraic equation is called dependent (or response) variable and the variable whose value is used to estimate this value is called independent (regressor or predictor) variable.

The linear algebraic equation used for expressing a dependent variable in terms of independent variable is called linear regression equation. The term regression was used in 1877 by Sir Francis Galton while studying the relationship between the height of father and sons. He found that though 'tall father has tall sons', the average height of sons of tall father is x above the general height, the average height of sons is 2x/3 above the general height. Such a fall in the average height was described by Galton as 'regression to mediocrity'. However, the theory of Galton is not universally applicable and the term regression is applied to other types of variables in business and economics. The term regression in the literary sense is also referred as 'moving backward'.

The basic differences between correlation and regression analysis are summarized as follows:

1. Developing an algebraic equation between two variables from sample data and predicting the value of one variable, given the value of the other variable is referred to as regression analysis, while measuring the strength (or degree) of the relationship between two variables is referred as correlation analysis. The sign of correlation coefficient indicates the nature (direct or inverse) of relationship between two variables, while the absolute value of correlation coefficient indicates the extent of relationship.

2. Correlation analysis determines an association between two variables x and y but not that they have a cause-and-effect relationship. Regression analysis, in contrast to correlation, determines the cause-and-effect relationship between x and y, that is, a change in the value of independent variable x causes a corresponding change (effect) in the value of dependent variable y if all other factors that affect y remain unchanged.

3. In linear regression analysis one variable is considered as dependent variable and other as independent variable, while in correlation analysis both variables are considered to be independent.

4. The coefficient of determination r2 indicates the proportion of total variance in the dependent variable that is explained or accounted for by the variation in the independent variable. Since value of r 2 is determined from a sample, its value is subject to sampling error. Even if the value of r 2 is high, the assumption of a linear regression may be incorrect because it may represent a portion of the relationship that actually is in the form of a curve

Concept of Regression

The statistical technique that expresses the relationship between two or more variables in the form of an equation to estimate the value of a variable, based on the given value of another variable, is called regression analysis. The variable whose value is estimated using the algebraic equation is called dependent (or response) variable and the variable whose value is used to estimate this value is called independent (regressor or predictor) variable. The linear algebraic equation used for expressing a dependent variable in terms of independent variable is called linear regression equation.

The term regression was used in 1877 by Sir Francis Galton while studying the relationship between the height of father and sons. He found that though 'tall father has tall sons', the average height of sons of tall father is x above the general height, the average height of sons is 2x/3 above the general height. Such a fall in the average height was described by Galton as 'regression to mediocrity'. However, the theory of Galton is not universally applicable and the term regression is applied to other types of variables in business and economics. The term regression in the literary sense is also referred as 'moving backward'.

The regression equation can be written as

$Y = \alpha + \beta X + \varepsilon$ 1

Where,

Y = dependent variable or criterion variable

á = the population parameter for the y-intercept of the regression line, or regression coefficient (r*=óy/ óx)

â = population slope of the regression line or regression coefficient (r*óx/ óy)

ε = the error in the equation or residual

The value of á and â are not known, since they are values at the level of population. The population level value is called the parameter. It is virtually impossible to calculate parameter. So we have to estimate it. The two parameters estimated are á and â. The estimator of the á is 'a' and the estimator for â is 'b'. So at the sample level equation can be written as

$Y = a + \beta x + e$ 2

Where,

Y = the scores on Y variable

X = scores on X variable

a = the Y-intercept of the regression line for the sample or regression constant in sample

b = the slope of the regression line or regression coefficient in sample

e = error in prediction of the scores on Y variable, or residual

$\hat{Y}$= a +bX 3

Where, Y = predicted value of Y in sample. This value is not an actual value but the value of Y that is predicted using the equation $\hat{Y}$= a +bX. So, we can write error as by substituting the in the earlier equation.

Y-$\hat{Y}$=e

Advantages of Regression Analysis

The following are some important advantages of regression analysis:

1. Regression analysis helps in developing a regression equation by which the value of a dependent variable can be estimated given a value of an independent variable.

2. Regression analysis helps to determine standard error of estimate to measure the variability or spread of values of a dependent variable with respect to the regression line. Smaller the variance and error of estimate, the closer the pair of values (x, y) fall about the regression line and better the line fits the data, that is, a good estimate can be made of the value of variable y. When all the points fall on the line, the standard error of estimate equals zero.

3. When the sample size is large (df ≥ 29), the interval estimation for predicting the value of a dependent variable based on standard error of estimate is considered to be acceptable by changing the values of either x or y. The magnitude of r 2 remains the same regardless of the values of the two variables.

14.1 Types of Regression Models

The primary objective of regression analysis is the development of a regression model to explain the association between two or more variables in the given population. A regression model is the mathematical equation that provides prediction of value of dependent variable based on the known values of one or more independent variables.

The particular form of regression model depends upon the nature of the problem under study and the type of data available. However, each type of association or relationship can be described by an equation relating a dependent variable to one or more independent variables.

a. Simple and Multiple Regression Models

If a regression model characterizes the relationship between a dependent y and only one independent variable x, then such a regression model is called a simple regression model.

But if more than one independent variable is associated with a dependent variable, then such a regression model is called a multiple regression model. For example, sales turnover of a product (a dependent variable) is associated with multiple independent variables such as price of the product, expenditure on advertisement, quality of the product, competitors, and so on. Now if we want to estimate possible sales turnover with respect to only one of these independent variables, then it is an example of a simple regression model, otherwise multiple regression model is applicable.

b. Linear and Nonlinear Regression Models

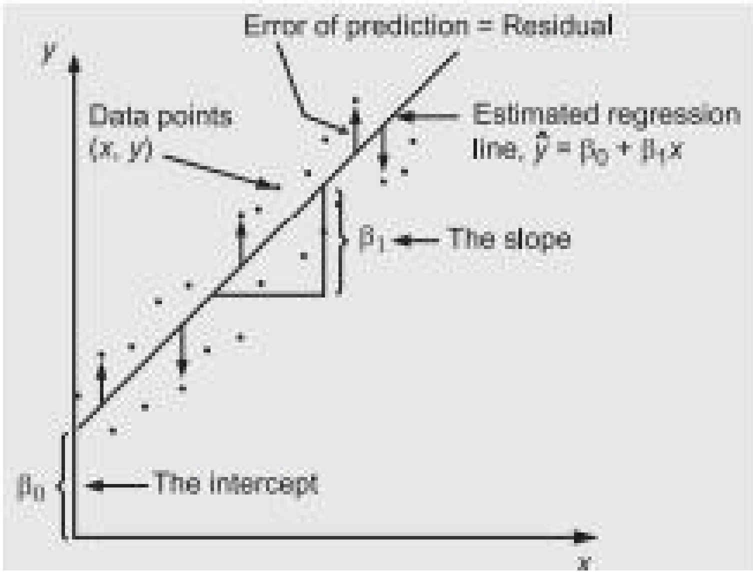
If the value of a dependent (response) variable y in a regression model tends to increase in direct proportion to an increase in the values of independent (predictor) variable x, then such a regression model is called a linear model. Thus, it can be assumed that the mean value of the variable y for a given value of x is related by a straight-line relationship. Such a relationship is called simple linear regression model expressed with respect to the population parameters β0 and β1 as:

E(y|x) = β0 + β1x

Where,

β0 = y-intercept that represents mean (or average) value of the dependent variable y when x = 0

β1 = slope of the regression line that represents the expected change in the value of y (either positive or negative) for a unit change in the value of x.



The intercept β0 and the slope β1 are unknown regression coefficients. The equation (1) requires to compute the values of β0 and β1 to predict average values of y for a given value of x. However, Fig. 14.1 presents a scatter diagram where each pair of values (xi, yi) represents a point in a two-dimensional coordinate system. Although the mean or average value of y is a linear function of x, but not all values of y fall exactly on the straight line rather fall around the line.

Since few points do not fall on the regression line, therefore values of y are not exactly equal to the values yielded by the equation: E(y|x) = β0 + β1x, also called line of mean deviations of observed y value from the regression line. This situation is responsible for random error (also called residual variation or residual error) in the prediction of y values for given values of x. In such a situation, it is likely that the variable x does not explain all the variability of the variable y. For instance, sales volume is related to advertising, but if other factors related to sales are ignored, then a regression equation to predict the sales volume (y) by using annual budget of advertising (x) as a predictor will probably involve some error. Thus, for a fixed value of x, the actual value of y is determined by the mean value function plus a random error term as follows:

Y= =Mean value function + Deviation

=β0 + β1x + e = E(y) + e

where e is the observed random error. This equation is also called simple probabilistic linear regression model.

14.2 Ordinary Least Squares (OLS)

In the previous section, we have discussed the simple regression equation with only one regressor variable X and the variable of interest Y. We have also discussed the simple linear regression model with a single regressor variable X. The simple linear regression model has two unknown parameters a and b, which are known as intercept and regression coefficient, respectively. Their values are unknown. Therefore, they must be estimated using sample data. The estimation of the parameters a and b is done by minimizing the error term e.

Let $(Y_1, X_1), (Y_2, X_2), \ldots, (Yn, Xn)$ be n pairs of values in the data. The equation of the simple linear regression model may be written as

$Y = a + bX + e$ 4)

where e represents the error term which arises due to the difference of the observed Y and the fitted line $\hat{y} = \hat{a} + \hat{b}x$. We use the method of least squares to minimize the error term e. From equation (4), we may write a simple regression model as

$Y_i = a + b X_i + e_i \qquad i = 1, 2, \ldots, n$ 5) for a sample data of n pairs of values given in terms of

(Y_i, X_i), (i = 1, 2, …n).

We estimate a and b so that the sum of the squares of the differences between the observed values (Yi) and the points lying on the straight line is minimum, i.e., the sum of squares of the error terms given by

$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \qquad 6)$$

is minimum. To find the values of a and b for which the sum of squares of the error terms, i.e., E is minimum, we differentiate it with respect to the parameters a and b and equate the results to zero:

$$\frac{\partial E}{\partial a} = -2\sum_{i=1}^{n}(Y_i - a - bX_i) = 0 \qquad 7)$$

And

$$\frac{\partial E}{\partial a} = -2\sum_{i=1}^{n}(Y_i - a - bX_i)X_i = 0 \qquad 8)$$

Simplifying equations (7) and (8), we get

$$na + b\sum_{i=1}^{n} X_i = \sum_{i=}^{n} Yi \qquad 9)$$

$$a\sum_{1=1}^{n} X_i + b\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i X_i \qquad 10)$$

Equations (10) and (9) are called the least-squares normal equations. The solution to these normal equations is

$$\hat{Y} - \hat{b}\bar{X} = a \qquad 11)$$

where Y and X are the averages of Yi and Xi , respectively. On putting the value of aˆ from equation (10) in equation (9), we get

$$\hat{b} = \frac{\sum_{i=1}^{n} y_i X_i - (\sum Y_i)(\sum X_i/n}{\sum_{i=1}^{n} X_1^2 - (\Sigma X_i)^2/n} \qquad 12)$$

Since the denominator of equation (11) is the corrected sum of squares of Xi, we may rewrite it as

$$Ss_x = \sum_{i=1} X_i^2 - \frac{(\Sigma Y_i)^2}{n} = \sum_{i=1}(X_i - \bar{X})^2 \qquad 13)$$

Similarly, the numerator is the corrected sum of the cross product of Xi and Yi and may be rewritten as:

$$Ss_{XY} = \sum_{i=1} X_i Y_i - \frac{\sum Yi \Sigma x_i}{n} = \sum_{i=1}(X_i - \bar{X})(Y_i - \bar{Y})$$

Therefore, the expression for $\hat{b}$may be rewritten as

$$\hat{b} = \frac{SS_{XY}}{SS_X}$$

Thus, aˆ and b ˆ are the least squares estimates of the intercept a and slope b, respectively. Therefore, the fitted simple linear regression model is given by

line $\hat{Y} = \hat{a} + \hat{b}x$

Equation (13) gives a point estimate of the mean of Y for a particular X. The difference between the fitted value Yi ˆ and Yi is known as the residual and is denoted by ri:

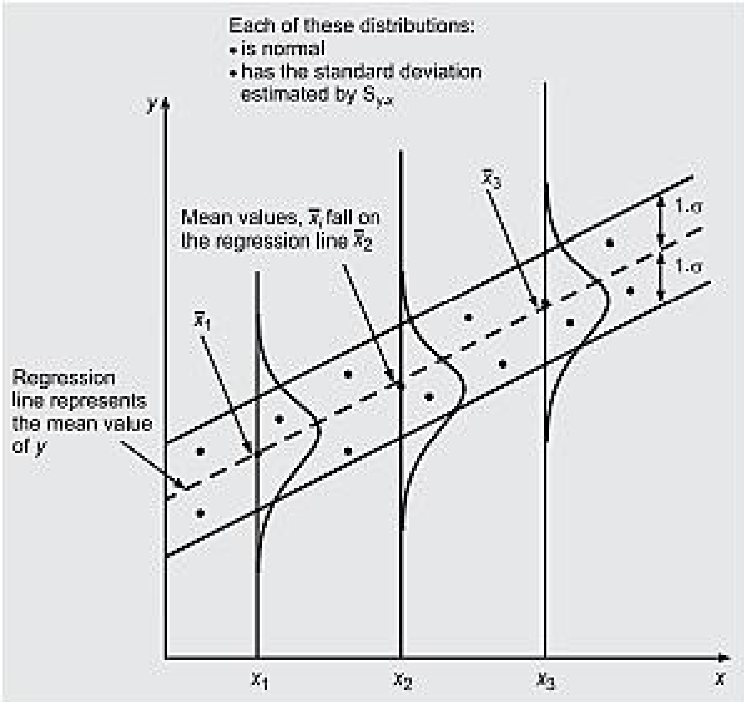
$r_i = Y_i - \hat{Y}i$, i = 1, 2, …, n

The role of the residuals and its analysis is very important in regression modelling.

14.3 Assumptions for A Simple Linear Regression Model

To make valid statistical inference using regression analysis, we make certain assumptions about the bivariate population from which a sample of paired observations is drawn and the manner in which observations are generated. These assumptions form the basis for application of simple linear regression models.

14.4 Assumptions



1. The relationship between the dependent variable y and independent variable x exists and is linear. The average relationship between x and y can be described by a simple linear regression equation y = a + bx + e, where e is the deviation of a particular value of y from its expected value for a given value of independent variable x.

2. For every value of the independent variable x, there is an expected (or mean) value of the dependent variable y and these values are normally distributed. The mean of these normally distributed values fall on the line of regression.

3. The dependent variable y is a continuous random variable, whereas values of the independent variable x are fixed values and are not random.

4. The sampling error associated with the expected value of the dependent variable y is assumed to be an independent random variable distributed normally with mean zero and constant standard deviation. The errors are not related with each other in successive observations.

5. The standard deviation and variance of expected values of the dependent variable y about the regression line are constant for all values of the independent variable x within the range of the sample data.

6. The value of the dependent variable cannot be estimated for a value of an independent variable lying outside the range of values in the sample data.

14.5 Regression coefficients

To estimate values of population parameter β0 and β1, under certain assumptions, the fitted or estimated regression equation representing the straight line regression model is written as:

$\hat{y} = a + bx$

$\hat{y}$
= estimated average (mean) value of dependent variable y for a given value of independent variable x.

a or b0 = y-intercept that represents average value of $\hat{y}$

b = slope of regression line that represents the expected change in the value of y for unit change in the value of x.

To determine the value of ˆy for a given value of x, this equation requires the determination of two unknown constants a (intercept) and b (also called regression coefficient). Once these constants are calculated, the regression line can be used to compute an estimated value of the dependent variable y for a given value of independent variable x.

The regression coefficient 'b' is also denoted as:

• byx (regression coefficient of y on x) in the regression line, y = a + bx

• bxy (regression coefficient of x on y) in the regression line, x = c + dy

14.6 Properties of Regression Coefficients

1. 1. The correlation coefficient is the geometric mean of two regression coefficients, that is, r = $\sqrt{byx^{*}bxy}$
2. If one regression coefficient is greater than one, then other regression coefficient must be less than one, because the value of correlation coefficient r cannot exceed one. However, both the regression coefficients may be less than one.
3. Both regression coefficients must have the same sign (either positive or negative). This property rules out the case of opposite sign of two regression coefficients.
4. The correlation coefficient will have the same sign (either positive or negative) as that of the two regression coefficients. For example, if by x = – 0.664 and bxy = – 0.234, then r = – $\sqrt{0.664 * 0.234}$ = – 0.394.
5. The arithmetic mean of regression coefficients bxy and byx is more than or equal to the correlation coefficient r, that is, (by x + bx y)/2 ≥ r. For example, if byx = – 0.664 and bxy = – 0.234, then the arithmetic mean of these two values is (– 0.664 – 0.234)/2 = – 0.449, and this value is more than the value of r = – 0.394.
6. Regression coefficients are independent of origin but not of scale.

14.7 Least Squared Methods

Example 1: Use least squares regression line to estimate the increase in sales revenue expected from an increase of 7.5 per cent in advertising expenditure.

Table 1

Firm	Annual Percentage Increase in advertising expenditure	Annual Percentage Increase in Sales Revenue
A	1	1
B	3	2
C	4	2
D	6	4
E	8	6
F	9	8
G	11	8
H	14	9

Solution: Assume sales revenue (y) is dependent on advertising expenditure (x). Calculations for regression line using following normal equations are shown in Table 3

Σ y = na + bΣx and

Σ xy = a Σ x + bΣ x^2

Table 2: Calculations of Normal Equations

Sales revenue	Advertising expenditure	Annual Percentage Increase in Sales Revenue	
1	1	1	1
2	3	9	6
2	4	16	8
4	6	36	24
6	8	64	48
8	9	81	72
8	11	121	88
9	14	196	126
40	56	524	373

Σ y = na + bΣ x or 40 = 8a + 56b

Σ xy = aΣx + bΣ x^2 or 373 = 56a + 524b

Solving these equations, we get

a = 0.072 and b = 0.704

Substituting these values in the regression equation y = a + bx = 0.072 + 0.704x

For x = 7.5% or 0.075 increase in advertising expenditure, the estimated increase in sales revenue will be

y = 0.072 + 0.704 (0.075) = 0.1248 or 12.48%

Example 2: The owner of a small garment shop is hopeful that his sales are rising significantly week by week. Treating the sales for the previous six weeks as a typical example of this rising trend, he recorded them in Rs 1000's and analysed the results in table 3

TABLE 3

Week	Sales
1	2.69
2	2.62
3	2.80
4	2.70
5	2.75

6	2.81

Fit a linear regression equation to suggest to him the weekly rate at which his sales are rising and use this equation to estimate expected sales for the 7th week.

Solution: Assume sales (y) is dependent on weeks (x). Then the normal equations for regression equation: y = a + bx are written as:

Σ y = na + bΣx and

Σ xy = a Σ x + bΣ x^2

Calculations for sales during various weeks are shown in Table 4

Table 4: Calculations of Normal Equations

Week (x)	Sales(y)	x^2	xy
1	2.69	1	2.69
2	2.62	4	5.24
3	2.80	9	8.40
4	2.70	16	10.80
5	2.75	25	13.75
6	2.81	36	16.86
21	16.37	91	57.74

The gradient 'b' is calculated as:

$\hat{b} = \frac{SS_{XY}}{SS_{XX}}$= 0.445/17.5= 0.25

$SS_{XY} = \Sigma y - \frac{\Sigma x \Sigma y}{n}$= (57.74-21*16.37)/6= 0.445

$SS_{XX} = \Sigma x^2 - \frac{\Sigma x^2}{n}$= 91-(21)^2/6= 17.5

The intercept 'a' on the y-axis is calculated as

$$a = \bar{y} - b\bar{x} = \frac{16.37}{6} - 0.025 * \frac{21}{6}$$

= 2.728 – 0.025×3.5 = 2.64

Substituting the values, a = 2.64 and b = 0.025 in the regression equation, we have

y = a + bx = 2.64 + 0.025x

For x = 7, we have y = 2.64 + 0.025 (7) = 2.815

Hence the expected sales during the 7th week is likely to be Rs 2.815 (in Rs 1000's).

14.8 Deviations Method

Calculations to least squares normal equations become lengthy and tedious when values of x and y are large. Thus, the following two methods may be used to reduce the computational time.

a) Deviations Taken from Actual Mean Values of x and y: If deviations of actual values of variables x and y are taken from their mean values x and y , then the regression equations can be written as:

a.1) Regression equation of y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Where, b_{yx}= regression coefficient of y on x

The value of b_{yx} can be calculated using the formula

$$by_x = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

a.2) Regression equation of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

Where, b_{xy}= regression coefficient of y on x

The value of b_{xy}can be calculated formula

$$bxy = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

b) Deviations Taken from Assumed Mean Values for x and y

If mean value of either x or y or both are in fractions, then we must prefer to take deviations of actual values of variables x and y from their assumed means.

b.1) Regression equation of y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

where b_{yx}=

$$\frac{n_1\sum dx\, dy - (\sum dx)(\sum dy)}{n\Sigma d_x^2 - (\sum dx)^2}$$

n = number of observations

dx = x – A; A is assumed mean of x

dy = y – B; B is assumed mean of y

b.2) Regression equation of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

where b_{xy}=

$$\frac{n_1\sum dx\, dy - (\sum dx)(\sum dy)}{n\Sigma d_x^2 - (\sum dy)^2}$$

n = number of observations

dx = x – A; A is assumed mean of x

dy = y – B; B is assumed mean of y

c) Regression Coefficients in Terms of Correlation Coefficient

If deviations are taken from actual mean values, then the values of regression coefficients can be alternatively calculated as follows:

$$by_x = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$= \frac{Covariance(x,y)}{\sigma_x^2} = \text{r}.\ \sigma_y/\sigma_x$$

And

$$bxy = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

$$= \frac{Covariance(x,y)}{\sigma_x^2} = \text{r}.\ \sigma_x/\sigma_y$$

Example 3: The following data relate to the scores obtained by 9 salesmen of a company in an intelligence test and their weekly sales (in Rs 1000's)

Table 5:

Salesman	A	B	C	D	E	F	G	H	I
Test Scores	50	60	50	60	80	50	80	40	70
Weekly Sales	30	60	40	50	60	30	70	50	60

a) Obtain the regression equation of sales on intelligence test scores of the salesmen.

(b) If the intelligence test score of a salesman in 65, what would be his expected weekly sales.

Solution: Assume weekly sales (y) as dependent variable and test scores (x) as independent variable. Calculations for the following regression equation are shown in Table 5

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Weekly Sales,x	dx=x-60	d^2_x	Test score, y	dy=y-50	d^2_y	dxd_y
50	-10	100	30	-20	400	200
60	0	0	60	10	100	0
50	-10	100	40	-10	100	100
60	0	0	50	0	0	0
80	20	400	60	10	100	200
50	-10	100	30	-20	400	400
80	20	400	70	20	400	400
40	-20	400	50	0	0	0
70	10	100	60	10	100	100
540	0	1600	450	0	1600	1200

a) $\bar{x} = \Sigma x/n = 540/9 = 60$;

$\bar{y} = \Sigma y/n = 450/9 = 50$

$b_{yx} = \frac{\sum dx\,dy - (\sum dx)(\sum dy)}{\Sigma d_x^2 - (\sum dx)^2} = 1200/1600 = 0.75$

Substituting values in the regression equation, we have

y – 50 = 0.75 (x – 60) or y = 5 + 0.75x

For test score x = 65 of salesman, we have

y = 5 + 0.75 (65) = 53.75

Hence we conclude that the weekly sales is expected to be Rs 53.75 (in Rs 1000's) for a test score of 65.

Example 4: The following data give the ages and blood pressure of 10 women

Age	56	42	36	47	49	42	60	72	63	55
Blood	147	125	118	128	145	140	155	160	149	150

Pressure										

(a) Find the correlation coefficient between age and blood pressure.

(b) Determine the least squares regression equation of blood pressure on age.

(c) Estimate the blood pressure of a woman whose age is 45 years.

Solution: Assume blood pressure (y) as the dependent variable and age (x) as the independent variable. Calculations for regression equation of blood pressure on age are shown in Table 5

Table 5: Calculations for Regression Equation

Age, x	d_x = x-49	d^2_x	Blood, y	d_y=y-145	d^2_y	d_xd_y
56	7	49	147	2	4	14
42	-7	49	125	-20	400	140
36	-13	169	118	-27	729	351
47	-2	4	128	-17	289	34
49	0	0	145	0	0	0
42	-7	49	140	-5	25	35
60	11	121	155	10	100	110
72	23	529	160	15	225	345
63	14	196	149	4	16	56
55	6	36	150	5	25	30
522	32	1202	1414	-33	1813	1115

a) Coefficient of correlation between age and blood pressure is given by

$$\frac{n\sum dx\,dy - \sum dx\Sigma\, dg}{\sqrt{n\sum d_x^2 - (\sum dx)^2}\sqrt{n\Sigma d_y^2 - (\sum dy)^2}}$$

= 10(1115)-32(-33)/ $\sqrt{10(1202) - (32)^2}$* $\sqrt{10(1813) - (-33)^2}$

= 12206/13689=0.892

We may conclude that there is a high degree of positive correlation between age and blood pressure.

b) The regression equation of blood pressure on age is given by

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

a) $\bar{x}$= Σ x/n = 522/10 = 52.2;

$\bar{y}$ = Σ y/n = 1417/10 = 141.7

b_{yx}= $\frac{\sum dx\, dy - (\sum dx)(\sum dy)}{\Sigma d_x^2 - (\sum dx)^2}$= 10(1115)-32(-33)/(10(1202)-(32)²)= 12206/10996= 1.11

Substituting these values in the above equation, we have y – 141.7 = 1.11 (x – 52.2) or y = 83.758+1.11.

This is the required regression equation of y on x.

b) For a women, whose age is 45, the estimated average blood pressure will be

y = 83.758+1.11(45) = 83.758+49.95 = 133.708

Hence, the likely blood pressure of a woman of 45 years is 134.

Relationshipsbetween Correlation and Regression coefficients

The points given below, explains the difference between correlation and regression in detail:

1. A statistical measure which determines the co-relationship or association of two quantities is known as Correlation. Regression describes how an independent variable is numerically related to the dependent variable.
2. Correlation is used to represent the linear relationship between two variables. On the contrary, regression is used to fit the best line and estimate one variable on the basis of another variable.
3. In correlation, there is no difference between dependent and independent variables i.e. correlation between x and y is similar to y and x. Conversely, the regression of y on x is different from x on y.
4. Correlation indicates the strength of association between variables. As opposed to, regression reflects the impact of the unit change in the independent variable on the dependent variable.
5. Correlation aims at finding a numerical value that expresses the relationship between variables. Unlike regression whose goal is to predict values of the random variable on the basis of the values of fixed variable.

Comparison between Linear Correlation and Regression

	Correlation	*Regression*
• Measurement level	Interval or ratio scale	Interval or ratio scale
• Nature of variables	Both continuous, and linearly related	Both continuous, and linearly related
• $x - y$ relationship	x and y are symmetric	y is dependent, x is independent; regression of x on y differs from y on x
• Correlation	$b_{xy} = b_{yx}$	Correlation between x and y is the same as the correlation between y and x
• Coefficient of determination	Explains common variance of x and y	Proportion of variability of x explained by its least-squares regression on y

Summary

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

Keywords

Review Questions

1. (a) Explain the concept of regression and point out its usefulness in dealing with business problems.

(b) Distinguish between correlation and regression. Also point out the properties of regression coefficients.

2. Why should a residual analysis always be done as part of the development of a regression model?

3. What is the difference between a prediction interval and a confidence interval in regression analysis?

4. Explain what is required to establish evidence of a causeand-effect relationship between y and x with regression analysis.

5. The following calculations have been made for prices of twelve stocks (x) at the Calcutta Stock Exchange on a certain day along with the volume of sales in thousands of shares (y). From these calculations find the regression equation of price of stocks on the volume of sales of shares. Σ x = 580, Σy = 370, Σxy = 11494, Σ x2 = 41658, Σy 2 = 17206.

6. The following table gives the aptitude test scores and productivity indices of 10 workers selected at random:

Aptitude scores (x): 60 62 65 70 72 48 53 73 65 82

Productivity index (y): 68 60 62 80 85 40 52 62 60 81

Calculate the two regression equations and estimate (a) the productivity index of a worker whose test score is 92, (b) the test score of a worker whose productivity index is 75

Self Assessment

1. The line of 'best fit' to measure the variation of observed values of dependent variable in the sample data is

A. regression line
B. correlation coefficient
C. standard error
D. none of these

2. Regression coefficient is independent of..........................

A. Origin
B. Scale
C. Both a and b
D. Neither origin nor scale

3. Two regression lines are perpendicular to each other when

A. r = 0
B. r = 1/3
C. r = – 1/2
D. r = ± 1 18.

4. The change in the dependent variable y corresponding to a unit change in the independent variable x is measured by

A. bxy
B. byx
C. r
D. none of these

5. The regression lines are coincident provided

A. r = 0

B. r = 1/3

C. r = – 1/2

D. r = ± 1

6. If byx is greater than one, then bxy is

A. less than one

B. more than one

C. equal to one

D. none of these

7. If bxy is negative, then byx is

A. negative

B. positive

C. zero

D. none of these

8. The residual sum of square is

A. minimized

B. increased

C. maximized

D. decreased

9. The standard error of estimate Sy·x is the measure of

A. closeness

B. variability

C. linearity

D. none of these

10. If two coefficients of regression are 0.8 and 0.2, then the value of coefficient of correlation is

A. 0.16

B. – 0.16

C. 0.40

D. – 0.40

11. If two regression lines are: y = 4 + kx and x = 5 + 4y, then the range of k is

A. k ≤ 0

B. k ≥ 0

C. 0 ≤ k ≤ 1

D. 0 ≤ 4k ≤ 1

12. Correlation coefficient is the geometric mean of regression coefficients.

A. True

B. False

13. If the sign of two regression coefficients is negative, then sign of the correlation coefficient is positive.
A. True
B. False

14. Correlation coefficient and regression coefficient are independent.
A. True
B. False

15. The point of intersection of two regression lines represents average value of two variables
A. True
B. False

Answers for Self Assessment

1.	A	2.	A	3.	A	4.	B	5.	D
6.	A	7.	A	8.	A	9.	B	10.	A
11.	D	12.	A	13.	B	14.	B	15.	A

Further Readings

- Nagar, A.L. and R.K Das, 1989 : Basic Statistics, Oxford University Press, Delhi.
- Goon, A.M., M.K. Gupta and B.'Dasgupta, 19.87 : Basic Statistics, The World Press Pvt. Ltd., Calcutta.

Web Links

- Edwards, B. 1980. The Readable Maths and Statistics Book, George Allen and Unwin: London.
- Makridakis, S. and S. Wheelwright, 1978. Interactive Forecasting: Univariate and Multivariate Methods, Holden-Day: San Francisco.

LOVELY PROFESSIONAL UNIVERSITY
Jalandhar-Delhi G.T. Road (NH-1)
Phagwara, Punjab (India)-144411
For Enquiry: +91-1824-521360
Fax.: +91-1824-506111
Email: odl@lpu.co.in