Engineering Mathematics

Important Formulas

Formulas & Short Notes Handbook

IMPORTANT FORMULAS TO REMEMBER

CHAPTER 1: LINEAR ALGEBRA

1. Minor and Cofactors:

Minor:

The minor of the element a_{ij} is denoted M_{ij} and is the determinant of the matrix that remains after deleting row i and column j of A.

Co – factor:

The cofactor of a_{ij} is denoted C_{ij} and is given by:

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

Properties:

The value of a determinant does not change when rows and columns are interchanged i.e.
$|A^T| = |A|$
If any row (or column) of a matrix A is completely zero, then:
$|A| = 0$, Such a row (or column) is called a zero row (or column).
Also, if any two rows (or columns) of a matrix A are identical, then $|A| = 0$.
If any two rows or two columns of a determinant are interchanged the value of determinant is multiplied by −1.
If all elements of the one row (or one column) or a determinant are multiplied by same number k the value of determinant is k times the value of given determinant.
If A be n-rowed square matrix, and k be any scalar, then $|kA| = k^n|A|$.
(i) In a determinant the sum of the products of the element of any row (or column) with the cofactors of corresponding elements of any row or column is equal to the determinant value. (ii) In determinant the sum of the products of the elements of any row (or column) with the cofactors of some other row or column is zero.

Example:

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

Then, $\Delta = a_{11}A_{11} + b_{12}B_{12} + c_{13}C_{13}$ and,

$\Delta = a_{31}A_{21} + b_{32}B_{22} + c_{33}C_{23} = 0$

If to the elements of a row (or column) of a determinant are added k times the corresponding elements of another row (or column) the value of determinant thus obtained is equal to the value of original determinant.

i.e. $A \xrightarrow{R_i + kR_j} B$ then $|A| = |B|$

and $A \xrightarrow{C_i + kC_j} B$ then $|A| = |B|$

(i). $|AB| = |A| \times |B|$ and based on this we can prove the following:

(i) $|A^n| = (|A|)^n$

Proof:

$|A^n| = |A \times A \times A \times$ ...... n times.

$|A^n| = |A| \times |A| \times |A|$ ... n times

$|A^n| = (|A|)^n$

(ii) $|A\,A^{-1}| = |I|$

Proof:

$|A\,A^{-1}| = |I| = 1$

Now, $|A\,A^{-1}| = |A|\,|A^{-1}|$

$\therefore |A|\,|A^{-1}| = 1$

$\Rightarrow |A^{-1}| = \frac{1}{|A|}$

(j). Using the fact that $A \cdot \text{Adj } A = |A|.\, I$, the following can be proved for $A_{n\times n}$.

(i). $|\text{Adj } A| = |A|^{n-1}$

(ii). $|\text{Adj } (\text{Adj } (A))| = |A|^{(n-1)^2}$

2. Transpose of a Matrix:

The matrix obtained from any given matrix A, by interchanging rows and columns is called the transpose of A and is denoted by A^T or A'.

Thus, the transposed matrix of $A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix}$ is $A' = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix}$

Clearly, the transpose of an $m \times n$ matrix is an $n \times m$ matrix.

Also, the transpose of the transpose of a matrix coincides with itself i.e. $(A')' = A$.

Properties of Transpose of a Matrix:

If A^T and B^T be transpose of A and B respectively then,

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(kA)^T = kA^T$, k being any real number.
4. $(AB)^T = B^TA^T$
5. $(ABC)^T = C^TB^TA^T$

Special Matrices and Properties:

3. Row and Column Matrix:

- A matrix having a single row is called a row matrix, e.g., $\begin{bmatrix} 1 & 3 & 4 & 5 \end{bmatrix}$.
- A matrix having a single column is called a column matrix, e.g., $\begin{bmatrix} 2 \\ 7 \\ 9 \end{bmatrix}$

- Row and column matrices are sometimes called row vector and column vectors.

4. Square matrix:

- An m × n matrix for which the number of rows is equal to number of columns i.e. m = n, is called square matrix.
- It is also called an n-rowed square matrix.
- The element a_{ij} such that i = j, i.e. a_{11}, a_{22}… are called DIAGONAL ELEMENTS and the line along which they line is called Principle Diagonal of matrix. Elements other than principal diagonal elements are called off-diagonal elements i.e. a_{ij} such that $i \neq j$.

5. Diagonal Matrix:

A square matrix in which all off-diagonal elements are zero is called a diagonal matrix. The diagonal elements may or may not be zero.

Example: $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}$ is a diagonal matrix.

5.1 Properties of diagonal Matrix:

(a) diag [x, y, z] + diag [p, q, r] = diag [x + p, y + q, z + r]

(b) diag [x, y, z] × diag [p, q, r] = diag [xp, yq, zr]

(c) $(\text{diag } [x, y, z])^{-1} = \text{diag } [1/x, 1/y/ 1/z]$

(d) $(\text{diag}[x, y, z])^T = \text{diag}[x, y, z]$

(e) $\text{diag } [x, y, z]^n = \text{diag}[x^n, y^n, z^n]$

(f) Eigen values of diag [x, y, z] = x, y and z.

(g) Determinant of diag [x, y, z] = | diag[x, y, z]| = xyz

5.1.1 Scalar Matrix:

A scalar matrix is a diagonal matrix with all diagonal elements belong equal.

Example: $A = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}$ is a scalar matrix where a is any non-zero value.

5.1.2 Unit Matrix or Identity Matrix:

- A square matrix each of whose diagonal elements is 1 and each of whose non-diagonal elements are zero is called unit matrix or an identity matrix which is denoted by I.
- Identity matrix is always square.
- Thus, a square matrix $A = [a_{ij}]$ is a unit matrix if $a_{ij} = 1$ when i = j and $a_{ij} = 0$ when $i \neq j$.

Example: $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is unit matrix, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

5.1.3 Properties of Identity Matrix:

(a) I is identity element for multiplication, so it is called multiplicative identity

(b) AI = IA = A

(c) $I^n = I$

(d) $I^{-1} = I$

(e) |I| = 1

5.1.4 Null matrix:

- The m × n matrix whose elements are all zero is called null matrix. Null matrix is denoted by O.
- Null matrix need not be square.

Example : $O_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $O_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $O_{2\times1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

5.1.5 Properties of Null Matrix:

(a) A + O = O + A = A. So, O is additive identity.

(b) A + (–A) = O

5.1.6 Upper triangular Matrix:

- An upper triangular matrix is a square matrix whose lower off-diagonal elements are zero i.e. a_{ij} = 0 whenever i > j.
- It is denoted by U.
- The diagonal and upper off diagonal elements may or may not be zero.

Example : $U = \begin{bmatrix} 3 & 5 & -1 \\ 0 & 5 & 6 \\ 0 & 0 & 2 \end{bmatrix}$

5.1.7 Lower Triangular matrix:

- A lower triangular matrix is a square matrix whose upper off-diagonal triangular elements are zero, i.e., a_{ij} = 0 whenever i < j.
- The diagonal and lower off-diagonal elements may or may not be zero. It is denoted by L.

Example : $L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 5 & 0 \\ 2 & 3 & 6 \end{bmatrix}$

5.1.8 Idempotent Matrix:

A matrix A is called idempotent if $A^2 = A$.

Example: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$ are examples of idempotent matrices.

5.1.9 Involutory Matrix:

A matrix A is called involutory if $A^2 = I$.

Example: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is involutory.

Also $\begin{bmatrix} 4 & 3 & 3 \\ -1 & 0 & -1 \\ -4 & -4 & -3 \end{bmatrix}$ is involutory since $A^2 = I$.

5.1.10 Nilpotent Matrix:

A matrix A is said to be nilpotent of class m or index m iff $A^m = O$ and $A^{m-1} \neq O$ i.e., m is the smallest index which makes $A^m = O$

Example: The matrix $A = \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix}$ is nilpotent class 3, since A ≠ 0 and $A^2 \neq 0$, but $A^3 = 0$.

5.1.11 Singular Matrix:

A matrix will be singular matrix if its determinant is equal to zero.

$$\left[a_{ij}\right]_{n\times n} = \begin{bmatrix} a_{11} & a_{12} & \ldots\ldots & a_{1n} \\ . & & & . \\ . & & & . \\ . & & & . \\ . & & & . \\ a_{n1} & a21 & \ldots\ldots & a_{nn} \end{bmatrix}_{n\times n}$$

If $|a_{ij}| = 0 \Rightarrow$ Matrix will be singular.

If a given matrix is not singular, then it will be the Non – singular matrix.

6. Periodic Matrix:

A square matrix A is called periodic if $A^{k+1} = A$ where k is least positive integer and is called the period of A.

6.1 Classification of Real and Complex Matrices:

6.1.1. Real Matrices:

Real matrices can be classified into the following three types of the relationship between A^T and A.

6.1.2 Symmetric Matrix:

• A square matrix $A = [a_{ij}]$ is said to be symmetric if its $(i, j)^{th}$ elements is same as its $(j, i)^{th}$ element i.e. $a_{ij} = a_{ij}$ for all i and j.

• In a symmetric matrix: $A^T = A$

6.2 Properties of symmetric matrices: For any Square matrix A,

(a) AA^t is always a symmetric matrix.

(b) $\frac{A + A^t}{2}$ is always symmetric matrix.

(c) $A - A^T$ and $A^T - A$ are skew symmetric.

1. If A and B and symmetric, then:

(a) A + B and A – B are also symmetric

(b) AB, BA may or may not be symmetric.

(c) A^k is symmetric when k is set of any natural number.

(d) AB + BA is symmetric.

(e) AB – BA is skew symmetric.

(f) A^2, B^2, $A^2 \pm B^2$ are symmetric.

(g) KA is symmetric where k is any scalar quantity.

2. Every square matrix can be uniquely expressed as a sum of a symmetric and a skew-symmetric matrix. Let A be the given square matrix, then:

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

6.3 Skew – Symmetric Matrix:

- A square matrix A = $[a_{ij}]$ is said to be skew symmetric if $(i, j)^{th}$ elements of A is the negative of the $(j, i)^{th}$ elements of A if $a_{ij} = -a_{ij}\ \forall\ i, j$.
- In a skew symmetric matrix $A^T = -A$.
- A skew symmetric matrix must have all 0's in the diagonal.

Example: $A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$ is a skew-symmetric matrix.

Note :

(a) For any matrix A, the matrix $\frac{A - A^t}{2}$ is always skew symmetric.

(b) A ± B are skew symmetric.

(c) AB and BA are not skew symmetric.

(d) A^2, B^2, $A^2 \pm B^2$ are symmetric.

(e) A^2, A^4, A^6 are symmetric.

(f) A^3, A^5, A^7 are skew symmetric.

(g) kA is skew symmetric where k is any scalar number.

6.5 Orthogonal Matrices:

A square matrix A is said be orthogonal if: $A^T = A^{-1} \Rightarrow AA^T = AA^{-1} = 1$. Thus, A will be an orthogonal matrix if:

$AA^T = I = A^TA$.

Example: The identity matrix is orthogonal since $I^T = I^{-1} = I$

Note: Since for an orthogonal matrix A:

$\Rightarrow AA^T = I$

$\Rightarrow |AA^T| = |I| = 1$

$\Rightarrow |A|\ |A^T| = 1$

$\Rightarrow (|A|)^2 = 1$

$\Rightarrow |A| = \pm 1$

So, the determinant of an orthogonal matrix always has a modulus of 1.

6.6 Complex Matrices:

Complex matrices can be classified into the following three types based on relationship between A^θ and A.

6.6.1 Hermitian Matrix:

A necessary and sufficient condition for a matrix A to be Hermitian is that $A^\theta = A$.

Example: $A = \begin{bmatrix} a & b + ic \\ b - ic & d \end{bmatrix}$ is a Hermitian matrix.

6.6.2 Skew – Hermitian Matrix:

A necessary and sufficient condition for a matrix to be skew-Hermitian if $A^\theta = -A$.

Example: $A = \begin{bmatrix} 0 & -2-i \\ 2-i & 0 \end{bmatrix}$ is skew-Hermitian matrix.

6.6.3 Unitary Matrix:

A square matrix A is said to be unitary iff: $A^\theta = A^{-1}$.

Multiplying both sides by A, we get an alternate definition of unitary matrix as given below:

A square matrix A is said to be unitary iff:

$AA^\theta = I = A^\theta A$

6.6.4 Properties of addition and subtraction:

(a). Only matrices of the same order can be added or subtracted

(b). Addition of matrices is commutative i.e. A + B = B + A.

(c). Addition and subtraction of matrices is associative i.e. (A + B) – C = A + (B – C) = B + (A – C).

6.6.5 Multiplication of a Matrix by a Scalar:

The product of a matrix A by a scalar k is a matrix of which each element is k times the corresponding elements of A.

$$\text{Thus, } k\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} = \begin{bmatrix} ka_1 & kb_1 & kc_1 \\ ka_2 & kb_2 & kc_2 \end{bmatrix}$$

The distributive law holds for such products, i.e., k (A + B) = kA + kB.

Note:

All the laws of ordinary algebra hold for the addition or subtraction of matrices and their multiplication by scalars.

6.6.6 Multiplication of Matrices:

Two matrices can be multiplied only when the number of columns in the first is equal to the number of rows in the second. Such matrices are said to be conformable.

In general, if $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{bmatrix}$ be two m × n and n × p conformable matrices, then their product is defined as the m × p matrix:

$$AB = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

Where $c_{ij} = a_{i1} b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \ldots + a_{in}b_{nj}$ i.e. the element in the i^{th} row and the j^{th} column of the matrix AB is obtained by multiplying the i^{th} row of A with j^{th} column of B. The expression for c_{ij} is known as the inner product of the i^{th} row with the j^{th} column.

Properties of Matrix Multiplication:

1. Multiplication of matrices is not commutative. In fact, if the product of AB exists, then it is not necessary that the product of BA will also exist.

Example: $A_{3\times2} \times B_{2\times4} = C_{3\times4}$ but $B_{2\times4}\times A_{3\times2}$ does not exist since these are not compatible for multiplication.

2. Matrix multiplication is associative, if conformability is assured. i.e. A(BC) = (AB)C where A, B, C are m × n, n × p, p × q matrices respectively.

3. Multiplication of matrices is distributive with respect to addition matrices i.e. A (B + C) = AB + AC.

4. The equation AB = O does not necessarily imply that at least one of matrices A and B must be a zero matrix. For example, $\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\-1 & -1\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$.

5. In the case of matrix multiplication if AB = O then it is not necessarily imply that BA = O. In fact, BA may not even exist.

6. Both left and right cancellation laws hold for matrix multiplication as shown below:

AB = AC ⇒ B = C (iff A is non-singular matrix) and

BA = CA ⇒ B = C (iff is non-singular matrix).

7. Trace of Matrix:

Let A be a square matrix of order n. The Sum of elements lying along the principal diagonal is called the trace of A denoted by Tr(A).

Thus, if $A = [a_{ij}]_{n\times n}$ then:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ij} = a_{11} + a_{22} + a_{33} + \ldots\ldots + a_{nn}$$

7.1 Properties of trace of matrix:

(a). tr (λA) = λ tr(A)

(b). tr (A +B) = tr (A) + tr (B)

(c). tr (AB) = tr (BA)

8. Conjugate of the Matrix:

The matrix obtained from given matrix A on replacing its elements by the corresponding conjugate complex numbers is called the conjugate of A and is denoted by $\bar{A}$.

$$\text{Example}: A = \begin{bmatrix}2+3i & 4-7i & 8\\ -i & 6 & 9+i\end{bmatrix}$$

$$\text{Then, } \bar{A} = \begin{bmatrix}2-3i & 4+7i & 8\\ +i & 6 & 9-i\end{bmatrix}$$

9. Properties of Conjugate of a Matrix: If $\overline{A}$ & $\overline{B}$ be the conjugates of A and B respectively. Then,

(a). $\overline{(\overline{A})} = A$

(b). $\overline{(A+B)} = \overline{A} + \overline{B}$

(c). $\overline{(kA)} = \overline{k}\,\overline{A}$, k being any complex number

(d). $\overline{(AB)} = \overline{A}\,\overline{B}$, A and B being conformable to multiplication

(e). $\overline{A}$ = A iff A is real matrix

(f). $\overline{A}$ = –A iff A is purely imaginary matrix.

10. Transposed Conjugate of the Matrix:

The transpose of the conjugate of a matrix A is called transposed conjugate of A and is denoted by A^{θ} or A^* or $\left(\overline{A}\right)^T$. It is also called conjugate transpose of A.

10.1 properties: If A^{θ} and B^{θ} be the transposed conjugates of A and B respectively then,

(a). $(A^{\theta})^{\theta} = A$

(b). $(A + B)^{\theta} = A^{\theta} + B^{\theta}$

(c). $(kA)\theta = \overline{k}A^{\theta}$ where k → complex number

(d). $(AB)^{\theta} = B^{\theta}A^{\theta}$

11. Adjoint and Inverse of the Matrix:

11.1 Adjoint of a square matrix:

Let a square matrix $A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$. Then the transpose of matrix formed by the cofactors of the elements is called the transpose of the matrix and it is written as Adj(A).

$$\text{Cofactor} - \text{matrix}(C_{ij}) = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix}.\text{Then:}$$

$$\text{Adj}(A) = (C_{ij})^T = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{bmatrix}$$

Thus, adjoint of A matrix is the transpose of matrix formed by the cofactors of A.

12. Inverse of a matrix:

If A be any matrix, then a matrix B if it exists, such that:

AB = BA = I

Then, B is called the Inverse of A which is denoted by A^{-1} so that $AA^{-1} = I$.

Also $A^{-1} = \dfrac{\text{Adj}.A}{|A|}$, if A is non-singular matrix.

13. Properties of Inverse

(a). $AA^{-1} = A^{-1} A = I$

(b). A and B are are inverse of each other iff AB = BA = I

(c). $(AB)^{-1} = B^{-1} A^{-1}$

(d). $(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$

(e). If A be a n × n non-singular matrix, then $(A')^{-1} = (A^{-1})'$.

(f). If A be a n × n non-singular matrix then $(A^{-1})^{\theta} = (A^{\theta})^{-1}$.

(g). For a 2 × 2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ there is a short-cut formula for inverse as given below:

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{(ad - bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

14. Rank of Matrix:

The rank of a matrix is defined as the order of highest non-zero minor of matrix A. It is denoted by the notation ρ(A). A matrix is said to be of rank r when:

(i) it has at least one non-zero minor of order r, and

(ii) every minor of order higher than r vanishes.

14.1 Properties:

(a). Rank of A and its transpose is the same i.e. $\rho(A) = \rho(A')$.

(b). Rank of a null matrix is zero.

(c). Rank of a non-singular square matrix of order r is r.

(d). If a matrix has a non-zero minor of order r, its rank is ≥ r and if all minors of a matrix of order r + 1 are zero, its rank is ≤ r.

(e). Rank of a matrix is same as the number of linearly independent row vectors vectors in the matrix as well as the number of linearly independent column vectors in the matrix.

(f). For any matrix A, rank (A) ≤ min(m,n) i.e. maximum rank of $A_{m\times n}$ = min (m, n).

(g). If Rank (AB) ≤ Rank A and Rank (AB) ≤ Rank B:

so, Rank (AB) ≤ min (Rank A, Rank B)

(h). Rank (A^T) = Rank (A)

(i). Rank of a matrix is the number of non-zero rows in its echelon form.

(j). Elementary transformations do not alter the rank of a matrix.

(k). Only null matrix can have a rank of zero. All other matrices have rank of at least one.

(l). Similar matrices have the same rank.

15. Vectors:

An ordered n-tuple $X = (x_1, x_2, \ldots x_n)$ is called an n-vector and $x_1, x_2, \ldots x_n$ are called components of X.

Row Vector:

A vector may be written as either a row matrix $X = [x_1\ x_2 \ldots x_n]$ which is called row vector.

15.1 Column Vector:

A column matrix $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$ which is called column vector.

Thus, for a matrix A of order m×n, each row of A is an n-vector and each column of A is an m-vector.

In particular, if m=1 then A is a row vector & if n=1 then A is a column vector.

15.2 Multiplication of a vector by a scalar:

Let 'k' be any number and $X = (x_1, x_2, \dots x_n)$ then $kX = (kx_1, kx_2, \dots kx_n)$.

Example:

X = (1, 3, 2)

Then, 4X = (4, 12, 8)

15.3 Linear combination of vectors:

If $X_1, X_2, \dots X_r$ are r vectors of order n and $k_1, k_2, \dots k_r$ are r scalars then the expression of the form $k_1X_1+k_2X_2+ \dots +k_rX_r$ is also a vector and it is called linear combination of the vectors $X_1, X_2, \dots X_r$.

15.4 Linearly dependent vectors:

The vectors $X_1, X_2, \dots. X_r$ of same order n are said to be linearly dependent if there exist scalars (or numbers) $k_1, k_2, \dots k_r$ not all zero such that $k_1X_1+k_2X_2+\dots\dots+k_rX_r = O$ where O denotes the zero vector of order n.

15.5 Linearly independent vectors:

The vectors $X_1, X_2, \dots., X_r$ of same order n are said to be linearly independent vectors if every relation of the type:

$K_1X_1+k_2X_2+\dots+k_rX_r = O$

Such that all $k_1 = k_2 = \dots\dots = k_r = 0$

Note.7:

(i). If $X_1, X_2, \dots\dots, X_r$ are linearly dependent vectors then at least one of the vectors can be expressed as a linear combination of other vectors.

(ii). If A is a square matrix of order n and $|A| = 0$ then the rows and columns are linearly dependent.

(iii). If A is a square matrix of order n and $|A| \neq 0$ then the rows and columns are linearly independent.

(iii). Any subset of a linearly independent set is itself linearly independent set.

(iv). If a set of vectors includes a zero vector, then the set of vectors is linearly dependent set.

15.6 Inner product:

The inner product of two vectors $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ is denoted by $X \cdot Y$ and defined as

$$X \cdot Y = X^T Y = [x_1 x_2 \ldots\ldots x_n] \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2\, y_2 + \ldots\ldots + x_n y_n$$ which is a scalar quantity.

Note.8:

1. $X^TY=Y^TX$ i.e. Inner Product is symmetric
2. X.Y = 0 $\Rightarrow$ the vectors X and Y are perpendicular.
3. X.Y. = $\pm 1 \Rightarrow$ the vectors X and Y are parallel.

16. Length or norm of a vector:

If $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a vector of order n then the positive square root of inner product of X and X^T i.e. X^TX is called length of X and it is denoted by $\|X\|$.

$$\therefore \|X\| = \sqrt{X \cdot X} = \sqrt{x_1^2 + x_2^2 + \ldots\ldots + x_n^2}$$

16.1 Orthonormal vectors/Orthonormal set:

A set S of column vectors $X_1, X_2, \ldots. X_n$ of same order is said to be an orthonormal set if

$$X_i^T X_j = \delta_{ij} = \begin{cases} 0, & \forall i \neq j \\ 1, & \forall i = j \end{cases}.$$

Ex: $X_1 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$$\Rightarrow X_1^T X_2 = 0,\ X_1^T X_3 = 0,\ X_2^T X_3 = 0 \text{ and } X_1^T X_1 = 1,\ X_2^T X_2 = 1,\ X_3^T X_3 = 1$$

$\therefore$ X_1, X_2, X_3 are orthonormal vectors.

17. System of Linear Equations:

17.1 Homogenous System of Linear Equations:

If the system of m homogeneous linear equations in n variables $x_1, x_2, \ldots. x_n$ is given by:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots. + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots. + a_{2n}x_n &= 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots. + a_{mn}x_n &= 0 \end{aligned}\right\}$$

Then, the set of these equations can be written in matrix form as AX = O

where A is matrix of the coefficients and X is the column matrix of the variables.

Note.9:

For the system AX = O where A is the square matrix then:

(i). The system AX = O is always consistent.

(ii). If $|A| \neq 0$ and $\rho(A) = n$ (number of variables). Then, the system has unique solution (zero solution or trivial solution)

(iii). If $|A| = 0$ and $\rho(A) < n$ then the system has infinitely many non-zero (or non-trivial) solutions.

(iv). If $\rho(A) = r < n$ (number of variables) then the number of linearly independent solutions of $AX = O$ is (n – r).

(v). In a system of homogeneous linear equations, if the number of unknowns (or variables) exceeds the number of equations then the system necessarily possesses a non-zero solution.

17.2 Non – homogenous system of linear equations:

If the system of 'm' non-homogeneous linear equation in 'n' variables $x_1, x_2, \ldots x_n$ is given by

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}\right\} \quad \ldots\ldots\ldots (1)$$

Then, the set of these equations can be written in matrix form as:

$AX = B$ ……… (2)

Where A is coefficient matrix, X is column matrix of the variables and B is the column matrix of constants $b_1, b_2, \ldots. b_n$.

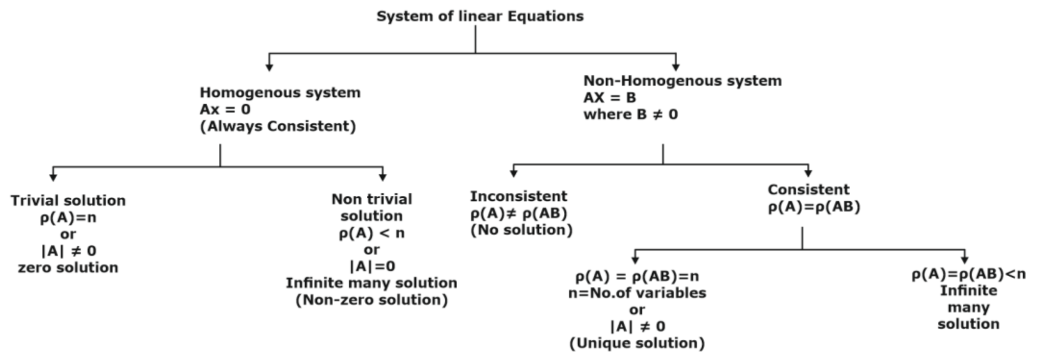
Note

(i). The system has a solution (consistent) if and only if Rank of A = Rank of [A|B].

(ii). The system AX = B has a unique solution if and only if Rank (A) = Rank (A|B) = n number of variables.

(ii) The system has infinitely many solutions if $\rho(A) = \rho(A\,|\,B) < n$ (number of variables).

(iii) The system has no solution (or is inconsistent) if $\rho(A) \neq \rho(A\,|\,B)$ i.e. $\rho(A) < \rho(A\,|\,B)$.



18. Eigen Values, Eigen vectors and Cayley Hamilton Theorem:

18.1 Eigen Values:

Let A = $[a_{ij}]_{n\times n}$ be any n-rowed square matrix and λ is a scalar. Then the matrix $A-\lambda I$ is called characteristic matrix of A, where I is the unit matrix of order n.

Then, the determinant $|A-\lambda I| = \begin{bmatrix} a_{11-\lambda} & a_{12} & a_{1n} \\ a_{21} & a_{22-\lambda} & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{nn-\lambda} \end{bmatrix}$ which is ordinary polynomial in λ of degree n is called "characteristic polynomial of A". The equation $|A-\lambda I| = 0$ is called "characteristic equation of A".

The λ values of this characteristic equation are called eigen values of A and the set of eigenvalues of A is called the "spectrum of A".

The corresponding non-zero solutions to X such that $AX = \lambda X$, for different eigen values are called as the eigen vectors of A.

18.2 Properties of Eigen Values:

(a). If $\lambda_1, \lambda_2 \ldots\ldots \lambda_n$ are the eigenvalues of A, then $k\lambda_1, k\lambda_2 \ldots\ldots k\lambda_n$ are eigenvalues of kA.

(b). The eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A. i.e. if $\lambda_1, \lambda_2 \ldots\ldots \lambda_n$ are the eigen value of A, then $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots \frac{1}{\lambda_n}$ are the eigen value of A^{-1}.

(c). If $\lambda_1, \lambda_2, \ldots \lambda_n$ are the eigen values of A, then $\lambda_1^m, \lambda_2^m, \ldots\ldots \lambda_n^m$ are the eigen values of A^m.

(d). If $\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_n$ are the eigen values of a non-singular matric A, then $\frac{|A|}{\lambda_1}, \frac{|A|}{\lambda_2} \ldots \frac{|A|}{\lambda_n}$ are the eigen values of Adj A.

(e). Eigen values of A = Eigen values of A^T.

(f). Maximum no. of distinct eigen values = size of A.

(g). If $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ …………..., λ_k are eigen values of matrix A of order n, then sum of eigen values = trace of A = sum of diagonal elements

i.e. $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \ldots, \lambda_k$ = trace of A

(h). Product of eigen values = $|A|$ (i.e. At least one eigen value is zero iff A is singular).

$\lambda_1 . \lambda_2 . \lambda_3 \ldots\ldots \lambda_k = |A|$

(i). In a triangular and diagonal matrix, eigen values are diagonal elements themselves.

(j). Similar matrices have same eigen values. Two matrices A and B are said to be similar if there exists a non-singular matrix P such that B = P^{-1} AP.

(k). If a + $\sqrt{b}$ is the one eigen value of a real matrix A then a - $\sqrt{b}$ other eigen value of matrix A.

(l). If a + ib is an eigen value of a real matrix A then a – ib is also other eigen value of A.

(m). If A and B are two matrices of same order, then the matrix AB and BA will have same characteristic roots.

18.3 Eigen Vectors:

The corresponding non-zero solutions to X such that $AX = \lambda X$, for different eigen values are called as the eigen vectors of A.

18.3.1 Properties of Eigen vectors:

(a). For each eigen value of a matrix there are infinitely many eigen vectors. If X is an eigen vector of a matrix A corresponding to the Eigen Value λ then KX is also an eigen vector of A for every non – zero value of K.

(b). Same Eigen vector cannot be obtained for two different eigen values of a matrix.

(c). Eigen vectors corresponding to the distinct eigen values are linearly independent.

(d). For the repeated eigen values, eigen vectors may or may not be linearly independent.

(e). The Eigen vectors of A and A^k are same.

(f). The eigen vectors of A and A^{-1} are same.

(g). The Eigen vectors of A and A^T are NOT same.

(h). Eigen vectors of a symmetric matrix are Orthogonal.

18.4 Cayley Hamilton Theorem:

Every square matrix A satisfies its own characteristic equation $A - \lambda I = 0$.

Example:

If $\lambda^2 - 5\lambda + 6 = 0$ is the Characteristic equation of the matrix A, then according to Cayley Hamilton theorem:

$A^2 - 5A + 6I = 0$

18.4.1 Applications of Cayley Hamilton theorem:

(a). It is used to find the higher powers of A such that A^2, A^3, A^4 etc.

(b). It can also be used to obtain the inverse of the Matrix.

18.5 Number of Linearly independent eigen vectors:

18.5.1 Algebraic Multiplicity:

The eigenvalues are the roots of the characteristic polynomials and a polynomial can have repeated roots.

i.e. $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 =$ -------------------$= \lambda_K$

If this happens then the eigenvalue is said to be of algebraic multiplicity k.

18.5.2 Geometric Multiplicity:

The number of linearly independent eigen vectors associated with that eigenvalue is called the Geometric multiplicity of that value.

Geometric Multiplicity (GM) corresponding to any eigen value λ_i is given by:

GM= n – Rank of $(A - \lambda_i I)$

Where n is the order of the matrix.

Thus, for a matrix A, the number of linearly independent eigen vectors is the sum of geometric multiplicities obtained corresponding to different eigen values.

19. Diagonalizable matrix:

If for a given square matrix A of order n, there exists a non – singular matrix P such that $P^{-1}AP = D$ or $AP = PD$ where D is the diagonal matrix then A is said to be diagonalizable matrix.

Note:

1. If X_1, X_2, X_3, ………., X_3 are linearly independent eigen vectors of $A_{3\times3}$ corresponding to eigen values λ_1, λ_2, λ_3 then P can be found such that $P^{-1}AP = D$ or $AP = PD$.

Where $D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$ and $P = [X_1, X_2, X_3]$

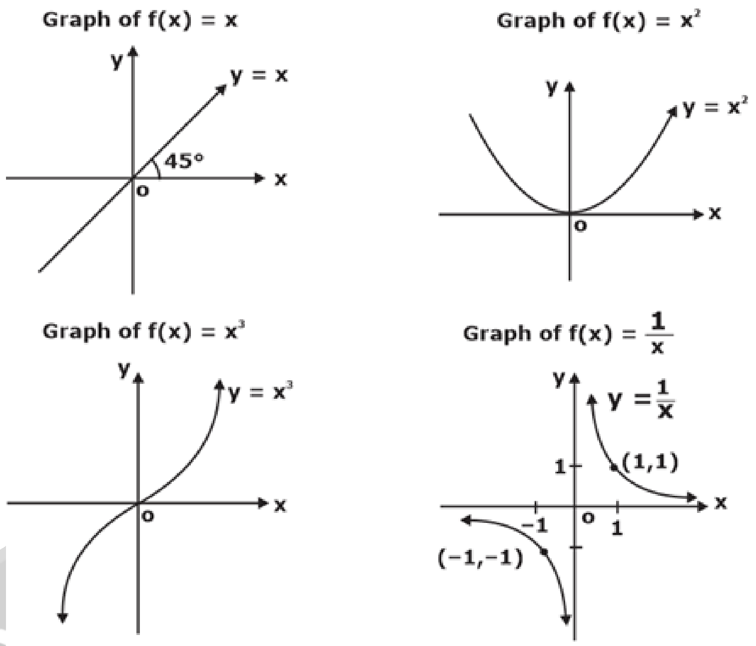
CHAPTER 2: CALCULUS

1. FUNCTIONS

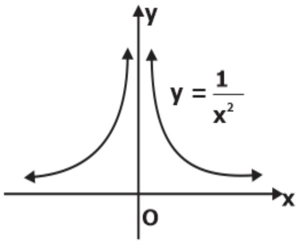
Definition:

We can define a function as a special relation which maps each element of set A with one and only one element of set B. Both the sets A and B must be nonempty. A function defines a particular output for a particular input.

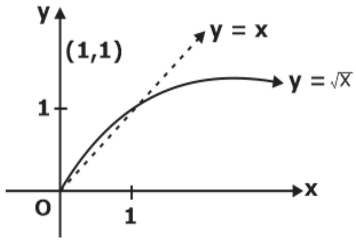
Basic graphs:



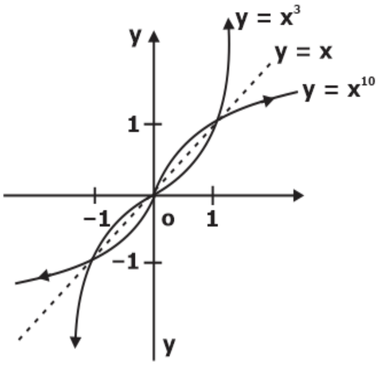
Graph of $f(x) = \frac{1}{x^2}$



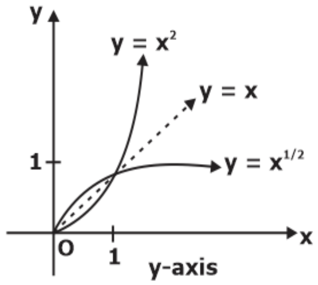
Graph of $f(x) = x^{1/2}$



Graph of $f(x) = x^{1/3}$



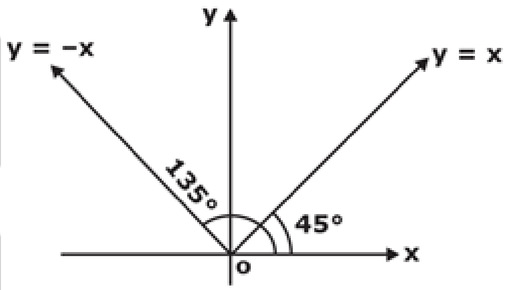
Graph of $f(x) = x^{1/2}$



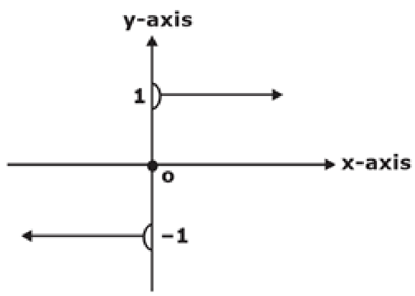
Modulus function

Signum function

Note.1: Graph of $f(x) = |x| = \begin{cases} -x & : \; x < 0 \\ 0 & : \; x = 0 \\ x & : \; x > 0 \end{cases}$

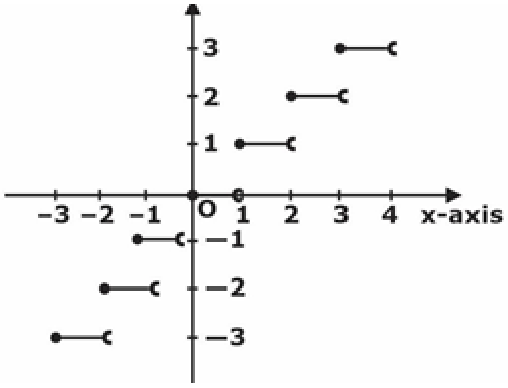


Note.2: Graph of $f(x) = \begin{cases} \dfrac{|x|}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases} = \begin{cases} -1 & : \; x < 0 \\ 1 & : \; x > 0 \\ 0 & : \; x = 0 \end{cases}$



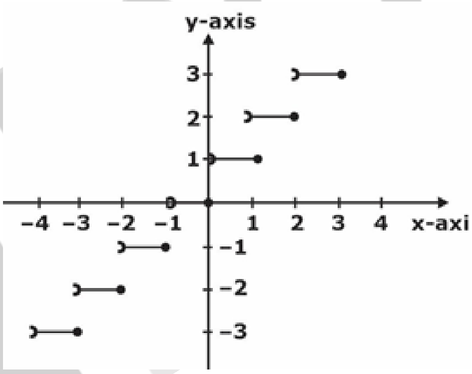
Note.3: Greatest Integer Function

Graph of f(x) = [x] = $\begin{cases} -2, & -2 \le x < -1 \\ -1, & -1 \le x < 0 \\ 0, & 0 \le x < 1 \\ 1, & 1 \le x < 2 \end{cases}$

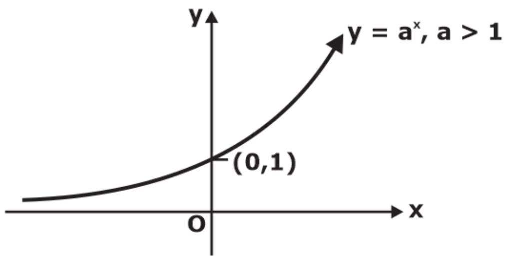


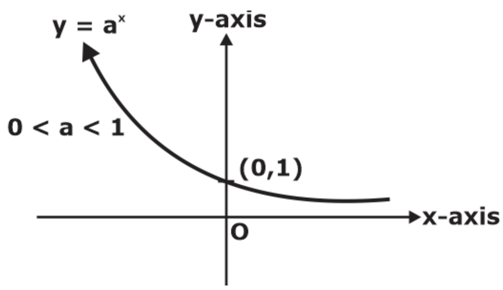
Note.4: Least Integer Function

Graph of f(x) = $\lceil x \rceil = \begin{cases} -1, -2 < x \le -1 \\ 0, -1 < x \le 0 \\ 1, 0 < x \le 1 \\ 2, 1 < x \le 2 \end{cases}$

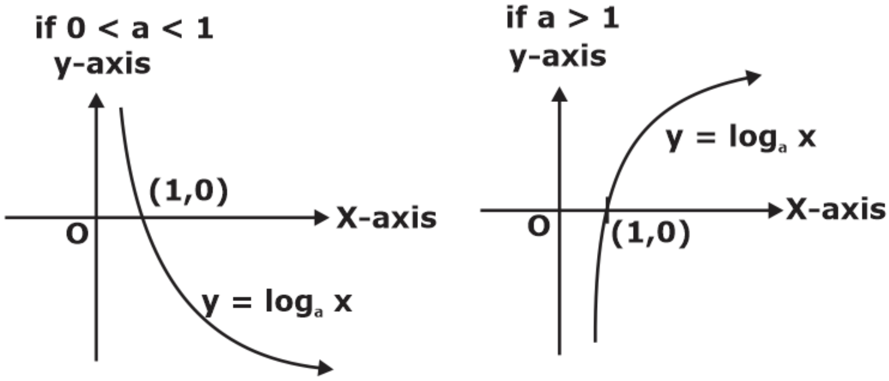


Exponential Function Graph of f(x) = a^x





Logarithmic function Graph of f(x) = log $_a$ x



Fundamental Theorem:

Rolle's Theorem:

If

(i) f(x) is continuous is the closed interval [a, b],

(ii) f′(x) exists for every value of x in the open interval (a, b) and

(iii) f (a) = f(b), then there is at least one value c of x in (a, b) such that f′ (c) = 0.

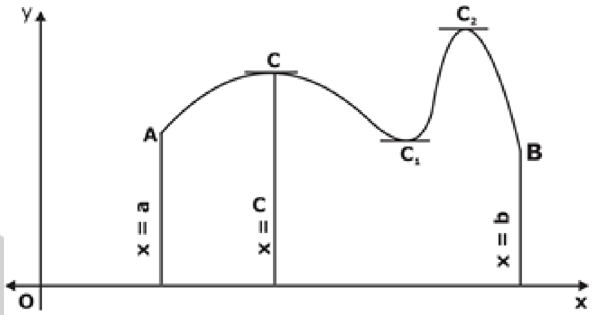


Fig.1

Consider the portion AB of the curve y = f(x), lying between x = a and x = b, such that

(i) It goes continuously from A to B,

(ii) It has a tangent at every point between A and B, and

(iii) Ordinate of A = ordinate of B.

From the fig. it is self-evident that there is at least one point C (may be more) of the curve at which the tangent parallel, to the x-axis.

i.e., slope of the tangent at C (x = c) = 0

But the slope of the tangent at C is the value of the differential coefficient of f(x) w.r.t x thereat, therefore f′(c) = 0. Hence the theorem is proved.

Lagrange's Mean-Value Theorem (LMVT):

If

(i) f(x) is continuous in the closed interval [a, b], and

(ii) f′(x) exists in the open interval (a, b),

then there is at least there is at one value c of x (a, b), such that $\frac{f(b)-f(a)}{b-a}=f'(c)$

Cauchy's Mean-value theorem:

If

(i) f(x) and g(x) be continuous in [a, b]

(ii) f′(x) and g′(x) exist in (a, b) and

(iii) g′(x) ≠ 0 for any value of x in (a, b)

Then there is at least one value c of x in (a, b), such that $\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(c)}{g'(c)}$

Taylor's series:

If f (x + h) can be expanded as an infinite series, then

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2!}f''(x)+\frac{h^3}{3!}f'''(x)+\ldots\infty$$

Replacing x by a and h by (x − a) in above, we get

$$f(x)=f(a)+(x-a)f'(a)+\frac{(x-a)^2}{2!}f''(a)+\frac{(x-a)^3}{3!}f'''(a)+\ldots\infty$$

Taking a = 0, we get Maclaurin's series.

Maclaurin's series:

If f(x) can be expanded as an infinite series, then

$$f(x)=f(0)+xf'(0)+\frac{x^2}{2!}f''(0)+\frac{x^3}{3!}f'''(0)+\ldots\infty$$

Expansion by use of known series:

When the expansion of a function is required only upto first few terms, it is often convenient to employ the following well-known series:

(i) $\sin\theta=\theta-\frac{\theta^3}{3!}+\frac{\theta^5}{5!}-\frac{\theta^7}{7!}+\ldots$

(ii) $\sinh\theta=\theta+\frac{\theta^3}{3!}+\frac{\theta^5}{5!}+\frac{\theta^7}{7!}+\ldots$

(iii) $\cos\theta=1-\frac{\theta^2}{2!}+\frac{\theta^4}{4!}-\frac{\theta^6}{6!}+\ldots$

(iv) $\cosh\theta=1+\frac{\theta^2}{2!}+\frac{\theta^4}{4!}+\frac{\theta^6}{6!}+\ldots$

(v) $\tan\theta=\theta+\frac{\theta^3}{3}+\frac{2}{15}\theta^5+\ldots$

(vi) $\tan^{-1}x=x-\frac{x^3}{3}+\frac{x^5}{5}-\ldots$

(vii) $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\ldots$

(viii) $\log(1-x) = -\left(x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots\right)$

(ix) $(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots..$

(x) $\log(1+x) = \left(x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots\right)$

2. LIMIT OF A FUNCTION

Let us consider a function f(x) defined in an interval I. If we see the behaviour of f(x) become closer and closer to a number l as $x \to a$ then l is said to be limit of f(x) at x=a.

Left Hand Limit –

Let function f(x) is said to approach l as $x \to a$ from left if for an arbitrary positive small number ε, a small positive number δ (depends on ε) such that

$|f(x) - l| < \varepsilon$ whenever $a - \delta < x < a$

It can also be written as

f (a -0) = $\lim\limits_{x \to a^-} f(x) = l$

Right Hand Limit

Let function f(x) is said to approach l as $x \to a$ from right if for an arbitrary positive small number ε, a small positive number δ (depends on ε) such that

$|f(x) - l| < \varepsilon$ whenever $a < x < a + \delta$

It can also be written as

f (a + 0) = $\lim\limits_{x \to a^+} f(x) = l$

if f (a+0) = f (a - 0) = l as $x \to a$, then the finite definite value l is said to be limit of f(x) at x = a

Important Results on Limits:

(i). $\lim\limits_{x\to 0} \frac{1-\cos mx}{1-\cos nx} = \frac{m^2}{n^2}$

(ii). $\lim\limits_{x\to 0} \frac{\cos ax - \cos bx}{\cos cx - \cos dx} = \frac{a^2 - b^2}{c^2 - d^2}$

(iii). $\lim\limits_{x\to 0} \frac{\cos mx - \cos nx}{x^2} = \frac{n^2 - m^2}{2}$

(iv). $\lim\limits_{x\to 0} \frac{\sin^p mx}{\sin^p nx} = \left(\frac{m}{n}\right)^p$

(v). $\lim\limits_{x\to 0} \frac{\tan^p mx}{\tan^p nx} = \left(\frac{m}{n}\right)^p$

(vi). $\lim_{x\to a}\frac{x^a - a^x}{x^x - a^a} = \frac{1-\log a}{1+\log a}$

(vii). $\lim_{x\to 0}\frac{(1+x)^m - 1}{(1+x)^n - 1} = \frac{m}{n}$

(viii). $\lim_{x\to 0}\frac{(1+bx)^m - 1}{(1+ax)^n - 1} = \frac{mb}{na}$

(ix). $\lim_{x\to 0}(1+ax)^{b/x} = \lim_{x\to\infty}\left(1+\frac{a}{x}\right)^{bx} = e^{ab}$

(x). $\lim_{n\to\infty}(x^n + y^n)^{\frac{1}{n}} = y, (0 < x < y)$

(xi). $\lim_{x\to\infty}\left(\frac{x\pm a}{x\pm b}\right)^{x+c} = e^{(a\mp b)}$

(xii). $\lim_{x\to 0}(\cos x + a\sin bx)^{1/x} = e^{ab}$

(xiii). $\lim_{x\to\infty}\frac{x^n}{e^x} = 0, \forall n$

(xiv). $\lim_{m\to\infty}\left(\cos\frac{x}{m}\right)^m = 1$

(xv). $\lim_{x\to\infty} a^x = \begin{bmatrix} 0, & 0 < a < 1 \\ 1, & a = 1 \\ \infty, & a > 1 \end{bmatrix}$

(xvi). $\lim_{n\to\infty}\cos\frac{x}{2}\cos\frac{x}{4}\cos\frac{x}{8}\ldots\cos\frac{x}{2^n} = \frac{\sin x}{x}$

(xvii). $\lim_{x\to 0}\frac{\sin x}{x} = \lim_{x\to 0}\frac{x}{\sin x} = 1$

(xviii). $\lim_{x\to 0}\frac{\tan x}{x} = \lim_{x\to 0}\frac{x}{\tan x} = 1$

(xix). $\lim_{x\to 0}\frac{\sin^{-1}x}{x} = \lim_{x\to 0}\frac{x}{\sin^{-1}x} = 1$

(xx). $\lim_{x\to 0}\frac{\tan^{-1}x}{x} = \lim_{x\to 0}\frac{x}{\tan^{-1}x} = 1$

(xxi). $\lim_{x\to 0}\frac{\sin x^\circ}{x} = \frac{\pi}{180^\circ}$

(xxii). $\lim_{x\to 0}\cos x = 1$

(xxiii). $\lim_{x\to a}\frac{\sin(x-a)}{x-a} = 1$

(xxiv). $\lim_{x\to a}\frac{\tan(x-a)}{x-a} = 1$

(xxv). $\lim\limits_{x \to a} \sin^{-1} x = \sin^{-1} a, \quad |a| \le 1$

(xxvi). $\lim\limits_{x \to a} \cos^{-1} x = \cos^{-1} a, \quad |a| \le 1$

(xxvii). $\lim\limits_{x \to a} \tan^{-1} x = \tan^{-1} a, \quad -\infty < a < \infty$

(xxviii). $\lim\limits_{x \to \infty} \dfrac{\sin x}{x} = \lim\limits_{x \to \infty} \dfrac{\cos x}{x} = 0$

(xxix). $\lim\limits_{x \to \infty} \dfrac{\sin \frac{1}{x}}{\frac{1}{x}} = 1$

(xxx). $\lim\limits_{x \to 0} \dfrac{(1+x)^n - 1}{x} = n$

INDETERMINATE FORMS:

Let us consider a function

$$F(x) = \frac{f(x)}{g(x)} \text{ then } \lim_{x \to a} F(x) = \lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim\limits_{x \to a} f(x)}{\lim\limits_{x \to a} g(x)}$$

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \text{ or } \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty$$

then the function

$$F(x) = \frac{f(x)}{g(x)}$$

is said to have indeterminate form of $\frac{0}{0}$ or $\frac{\infty}{\infty}$ respectively.

The other important indeterminate forms are $0 \times \infty$, $\infty - \infty$, 0^0, 1^∞ and ∞^0.

The limiting value of indeterminate forms is known as true value. The most standard form among all the indeterminate forms is $\frac{0}{0}$ or $\frac{\infty}{\infty}$. We can find the value of these two forms by using L-Hospital Rule.

L- Hospital Rule:

When $\lim\limits_{x \to a} f(x) = \lim\limits_{x \to a} g(x) = 0$ then $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = \lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)} = \ldots\ldots = \lim\limits_{x \to a} \dfrac{f^{(n)}(x)}{g^{(n)}(x)}$ provided $g'(x), \ldots g'^{(n)}(x)$ must not be zero, where $f^{(n)}$ and $g^{(n)}$ are n^{th} derivative of $f(x)$ and $g(x)$.

L- Hospital Rule for the form ($\infty - \infty$, $0 \times \infty$):

For the evaluation of $\lim\limits_{x \to \infty} [f(x) - g(x)]$, if it is in the form $(\infty - \infty)$, we will convert it into the form $\left(\frac{0}{0}\right)$ by simplification. The same process is also used in the form $(0 \times \infty)$. Then we use the L-Hospital Rule.

L-Hospital Rule for the form (0^0, 1^∞, ∞^0):

In the evaluation of $\lim_{x\to a}[f(x)]^{g(x)}$, we have to simply by taking the log and convert it into the form $\left(\frac{0}{0}\right)$. After that we can use the L-Hospital Rule

Note.5:

(i) $\log 1 = 0$ (ii) $\log 0 = -\infty$ (iii) $\log \infty = \infty$ (iv) $\log_1 x = \infty$

3. CONTINUITY

A function y = f(x) is said to be continuous if the graph of the function is a continuous curve. On the other hand, if a curve is broken at some point say x = a, we say that the function is not continuous or discontinuous.

Definition:

A function f(x) is said to be continuous at x = a if and only if the following three conditions are satisfied:

(i) f(x) exists; that is f(x) is defined at x = a

(ii) $\lim_{x\to a} f(x)$ exists

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

If the function is continuous at every point of a given interval [α, β], then it is said to be continuous in that interval.

Properties of continuous functions:

(i) A function which is continuous in a closed interval is also bounded in that interval.

(ii) A continuous function which has opposite signs at two points vanishes at least once between these points and vanishing point is called root of the function.

(iii) A continuous function f(x) in the closed interval [a, b] assumes at least once every value between f(a) and f(b), it being assumed that

$f(a) \neq f(b)$.

4. DIFFERENTIABILITY

Chain Rule of differentiability:

If $\phi(x) = \psi[f(x)]$, Then $\phi'(x) = \psi'[f(x)]f'(x)$

Note.6:

Let f and g be functions defined on an interval I and f, g are differentiable at

$x = a \in I$ then

(i) F $\pm$ G is differentiable and (F $\pm$ G)′ (a) = F′(a) $\pm$ G′(a).

(ii) cF is differentiable and (cF)′ (a) = c F′ (a): c $\in$ R.

(iii) F.G is differentiable and (FG)′(a) = F′(a)G(a) + F(a) G′(a)

(iv) $\frac{1}{F}$ is differentiable at x = a and $\left(\frac{1}{F}\right)(a) = -\frac{F'(a)}{[F(a)]^2}$: provided $F(a) \neq 0$.

(v) $\frac{F}{G}$ is differentiable at x = a and

$$\left(\frac{F}{G}\right)(a) = \frac{F'(a)G(a) - F(a)G'(a)}{[G(a)]^2} : \text{provided } G(a) \neq 0$$

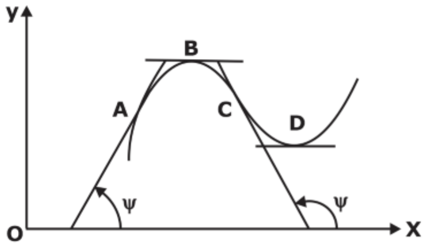
Note.7:

A Necessary condition for the Existence of a Finite Derivative

Continuity is a necessary but not the Sufficient for the existence of a finite derivatives.

5. INCREASING AND DECREASING FUNCTIONS

In the function y = f(x), if y increases as x increases (as at A), it is called an increasing function of x. On the contrary, if y decreases as x increases (as at c), it is called a decreasing function of x.



Let the tangent at any point on the graph of the function make an $\angle\psi$ with the x-axis so that

$$\frac{dy}{dx} = \tan\psi$$

At any point such as A, where the function is increasing $\angle\psi$ is acute i.e.,

$\frac{dy}{dx}$ is positive. At a point such as C, where the function is decreasing $\angle\psi$ is

Obtuse i.e. $\frac{dy}{dx}$ is negative. Hence the derivative of an increasing function is positive, and the derivative of a decreasing function is negative.

Note.8:

If the derivative is zero (as at B or D), then y is neither increasing nor decreasing. In such cases, we say that the function is **stationary**.

5.1. Concavity, Convexity and Point of Inflexion

(i) If a portion of the curve on both sides of a point, however small it may be, lies above the tangent (as at D), Then the curve is said to be **Concave upwards** at D where d^2y/dx^2 is positive.

(ii) If a portion of the curve on both sides of a point lies below the tangent (as at B), then the curve is said to be **Convex upwards** at B where $\frac{d^2y}{d^2x}$ is negative.

(iii) If the two portions of the curve lie on different sides of the tangent thereat (i.e., the curve crosses the tangent (as at C), then the point C is said to be a **Point of inflexion** of the curve.

At a point of inflexion $\frac{d^2y}{dx^2} = 0$ and $\frac{d^3y}{dx^3} \neq 0$.

6. MAXIMA AND MINIMA

Consider the graph of the continuous function y =f(x) in the interval (x_1, x_2) (Fig.). Clearly the point P_1 is the highest in its own immediate neighbourhood. So also is P_3. At each of these points P_1, P_3 the function is said to have a maximum value. On the other hand, the point P_2 is the lowest in its own immediate neighbourhood. So also is P_4. At each of these points P_2, P_4 the function is said to have a minimum value.

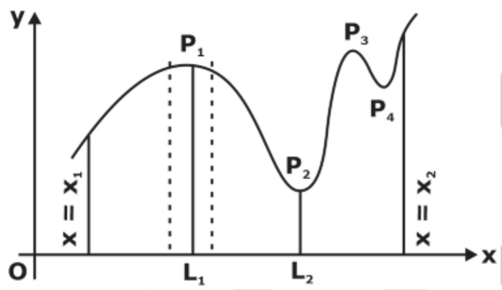


Fig.

Thus, we have

Definition:

A function f(x) is said to have a **maximum** value at x = a, if there exists a small number it, however small, such that f(a)> both f (a – h) and f (a + h).

A function f(x) is said to have a **minimum** value at x = a, if there exists a small number h, however small, such that f(a) < both f (a – h) and f (a + h).

Note.9:

The maximum and minimum values of a function taken together are called its extreme values and the points at which the function attains the extreme values are called the turning points of the function.

Note.10:

A maximum or minimum value of a function is not necessarily the greatest or least value of the function in any finite interval. The maximum value is simply the greatest value in the immediate neighbourhood of the maxima point or the minimum value is the least value in the immediate neighbourhood of the minima point. In fact. there may be several maximum and minimum values of a function in an interval and a minimum value may be even greater than a maximum value.

Note.11:

It is seen from the Fig. that maxima and minima values occur alternately.

(i) f(x) is maximum at x = a if f'(a) = 0 and f'' (a) is negative.

[i.e., f'(a) changes sign from positive to negative]

(ii) f(x) is minimum at x = a, if f' (a) = 0 and f''(a) is positive.

[i.e., f'(a) changes sign from negative to positive)

Note.12:

A maximum or a minimum value is a stationary value, but a stationary value may neither be a maximum non a minimum value.

Procedure for finding maxima and minima

(i) Put the given function = f(x)

(ii) Find f'(x) and equate it to zero.

Solve this equation and let its roots be a, b, c, …

(iii) Find f''(x) and substitute in it by turns x = a, b, c, …

If f'' (a) is negative, f(x) is maximum at x = a.

If f''(a) is positive, f(x) is minima at x = a.

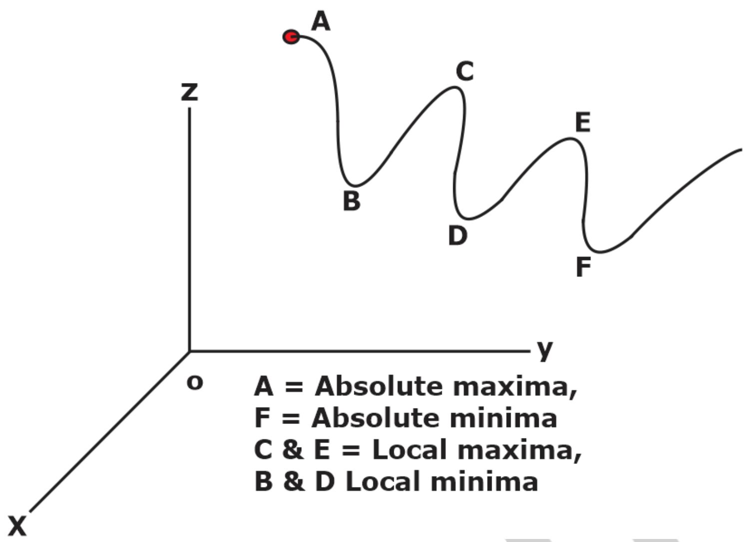
(iv) Sometimes f''(x) may be difficult to find out or f''(x) may be zero at x = a. In such cases, see if f'(x) changes sign from positive to negative as x passes through a, then f(x) is maximum at x = a.

If f'(x) changes sign from negative to positive as x passes through a, f(x) is minimum at x = a.

If f(x) does not change sign while passing through x = a, f(x) is neither maximum nor minimum at x = a.

7. MAXIMA - MINIMA OF FUNCTIONS OF TWO VARIABLES

Let Z = f (x, y) be a given surface shown in figure:



Maxima:

Let Z = f(x, y) be any surface and let P(a, b) be any point on it then f(x, y) is called maximum at P(a, b) if $f(a, b) > f(a + h, b + k)$ $\forall$ Positive and Negative values of h and k.

Minima:

Let Z = f(x, y) be any surface and let P(a, b) be any point on it then f(x, y) is called minimum at P(a, b) if $f(a, b) < f(a + h, b + k)$ $\forall$ Positive and Negative values of h and k.

Extremum:

The maximum or minimum value of the function f(x, y) at any point x = a and y = b is called the **extremum** value and the point is called **extremum point**.

Saddle Point:

It is a point where function is neither maximum nor minimum. At this point f is maximum in one direction while minimum in another direction. e.g. Consider Hyperbolic Paraboloid z = xy; since at origin (0, 0) function has neither maxima nor minima. So, origin is the saddle for Hyperbolic Paraboloid.

The Lagrange's conditions for maximum or minimum are:

Consider a function z = f(x,y) and let P(a, b) be any point on it, and let

$$r = \left(\frac{\partial^2 z}{\partial x^2}\right); s = \left(\frac{\partial^2 z}{\partial x \partial y}\right); t = \left(\frac{\partial^2 z}{\partial y^2}\right)$$

(i) If $rt - s^2 > 0$ and $r < 0$, then f (x,y) has maximum value at (a,b).

(ii) If $rt - s^2 > 0$ and $r > 0$, then f (x, y) has minimum value at (a, b).

(iii) If $rt - s^2 < 0$, then f (x, y) has neither a minimum nor minimum i.e. (a, b) is saddle point.

(iv) If $rt - s^2 = 0$, then case fail, and we need further investigations to calculate maxima or minima.

Flowchart to find Maxima and Minima:

(i). Consider a given function Z = f (x, y)

(ii). Calculate the values of x & y by using $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$.

(iii). Let we get x = a and y = b from step (2) then critical point is P (a. b)

(iv). Check Lagrange's conditions for maxima/minima.

(v). Now, maximum or minimum value is given by f(a, b).

8. PARTIAL DERIVATIVES

Let z = f(y) be a function of two variables x and y.

If wo keep y as constant and vary x alone, then z is a function of x only. The derivative of z with respect to x, treating y as constant, is called the partial derivative of z with respect to x and is denoted by one of the symbols.

Partial differentiation and its applications:

$$\frac{\partial z}{\partial x}, \frac{\partial f}{\partial x}, f_x(x,y), D_x f \qquad \text{Thus } \frac{\partial z}{\partial x} = \underset{\delta x \to 0}{\text{Lt}} \frac{f(x+\delta x, y) - f(x,y)}{\delta x}$$

Similarly, the derivative of z with respect to y, keeping x as constant, is called the partial derivative of z with respect to y and is denoted by one of the symbols.

$$\frac{\partial z}{\partial y}, \frac{\partial f}{\partial y}, f_y(x,y), D_y f. \qquad \text{Thus } \frac{\partial z}{\partial y} = \underset{\partial y \to 0}{\text{Lt}} \frac{f(x, y+\delta y) - f(x,y)}{\delta y}$$

Sometimes we use the following notation

$$\frac{\partial z}{\partial x} = p, \frac{\partial z}{\partial y} = q, \frac{\partial^2 z}{\partial x^2} = r, \frac{\partial^2 z}{\partial x \partial y} = s, \frac{\partial^2 z}{\partial y^2} = t$$

Total Derivative:

If u = f (x, y), where x = ϕ (t) and y = Ψ (t), then we can express u as a function of alone by substituting the values of x and y in f (x, y). Thus we can find the ordinary derivative du/dt which is called the total derivative of u to distinguish it from the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$.

(i). If u = f (x, y, z), where x,y, z are all functions of a variable t, then Chain rule is

$$\frac{du}{dt} = \frac{\partial u}{\partial x}.\frac{dx}{dt} + \frac{\partial u}{\partial y}.\frac{dy}{dt} + \frac{\partial u}{\partial z}.\frac{dz}{dt}$$

Chain rule:

$$\frac{du}{dt} = \frac{\partial u}{\partial x}.\frac{dx}{dt} + \frac{\partial u}{\partial y}.\frac{dy}{dt}$$

(ii). Differentiation of implicit functions.

If f (x, y) = c be an implicit relation between x and y which defines as a differentiable function of x, then

$$\frac{dy}{dx} = -\frac{\partial f}{\partial x} / \frac{\partial f}{\partial y} \qquad , \left[\frac{\partial f}{\partial y} \neq 0\right]$$

Change of Variables:

If u = f (x, y), Where x = ϕ (s, t) and y = ψ (s, t)

The necessary formulae for the change of variables are easily obtained.

If t is regarded as a constant, then x, y, u will be functions of s alone. Therefore, by, we have

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

Similarly, regarding s as constant, we obtain as

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

Homogeneous Functions:

An expression of the form $a_0x^n + a_1x^{n-1}y + a_2x^{n-2}y^2 + \ldots + a_n y^n$ in which every term is of the nth degree, is called a homogeneous function of degree n. This can be rewritten as $x^n [a_o + a_1(y/x) + a_2(y/x)^2 + \ldots + a_n (y/x)^n]$.

A function f(x, y) is said to be homogeneous function of degree n if f(kx, ky) = $k^n f(x, y)$

Note.13: f (x, y) = $x^3 \sin^{-1}\left(\frac{x}{y}\right)$ is homogeneous of degree 3.

Note.14: f(x,y) = $\frac{x^3 + y^3}{x - y} + x^{-8}\cos^{-1}\left(\frac{x}{y}\right)$ is not homogeneous

Note.15: f (x, y) = $\sin^{-1}(x^6 + y^6)$ is not homogeneous.

Euler's Theorem:

If u = f (x, y) is homogeneous function of degree n.

Then

(i) $$x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = nu$$

(ii) $$x^2\frac{\partial^2 u}{\partial x^2} + 2xy\frac{\partial^2 u}{\partial x \partial y} + y^2\frac{\partial^2 u}{\partial y^2} = n(n-1)u$$

Note.16:

If u = f (x, y) + g (x, y) where f and g are homogeneous functions of degree m, n respectively.

Then

(i) $$\frac{x\partial u}{\partial x} + y\frac{\partial u}{\partial y} = mf + ng$$

(ii) $$\frac{x^2\partial^2 u}{\partial x^2} + 2xy\frac{\partial^2 u}{\partial x \partial y} + y^2\frac{\partial^2 u}{\partial y^2} = m(m-1)f + n(n-1)g$$

If u = f(x, y) is not homogeneous but F(u) is homogeneous of degree n then

(i) $$x\frac{\partial u}{\partial x}+y\frac{\partial u}{\partial y}=n\frac{F(u)}{F'(u)}=g(u)\text{ say}$$

(ii) $$x^2\frac{\partial^2 u}{\partial x^2}+2xy\frac{\partial^2 u}{\partial x\partial y}+y^2\frac{\partial^2 u}{\partial y^2}=g(u)\left[g'(u)-1\right]$$

Note.17:

• If $u=\tan^{-1}\left(\frac{x^3+y^3}{x-y}\right)$ is not homogeneous then, $\tan u=\frac{x^3+y^3}{x-y}$ is homogeneous of degree 2.

$$x\frac{\partial u}{\partial x}+y\frac{\partial u}{\partial y}=n\frac{F(u)}{F'(u)}-\frac{2\tan u}{\sec^2 u}=\frac{2\sin u}{\cos u}\frac{\cos^2 u}{1}=2\sin u\cos u=\sin 2u=g(u)\text{ say}$$

$$x^2\frac{\partial^2 u}{\partial x^2}+y^2\frac{\partial^2 u}{\partial y^2}+2xy\frac{\partial^2 u}{\partial x\partial y}=g(u)\left[g'(u)-1\right]=\sin 2u.(2\cos 2u-1)$$

9. INTEGRATION

This is the inverse process of differentiation, if the differentiation of F(x) with respect to x be f(x) then the integration of f(x) with respect to x is F(x) i.e.,

$$\frac{d}{dx}F(x)=f(x)\Rightarrow\int f(x)dx=F(x)$$

But the derivative of a constant term is zero then

$\frac{d}{dx}\left[F(x)+C\right]=f(x)$, so we have

$$\int f(x)dx=F(x)+C$$

The process of finding the integral of a function is said to be integration and the function which is to be integrated is known as integrand.

Standard Formulae:

(i). $\int(ax+b)^n dx=\frac{(ax+b)^{n+1}}{a(n+1)}\, n\neq -1$

(ii). $\int\frac{1}{ax+b}dx=\frac{1}{a}\log(ax+b)$

(iii). $\int e^{ax+b}dx=\frac{1}{a}e^{ax+b}$

(iv). $\int a^{bx+c}dx=\frac{1}{b}a^{bx+c}\log_a e$

(v). $\int\sin(ax+b)dx=-\frac{1}{a}\cos(ax+b)$

(vi). $\int\cos(ax+b)dx=\frac{1}{a}\sin(ax+b)$

(vii). $\int\tan(ax+b)dx=\frac{1}{a}\log\sec(ax+b)$

(viii). $\int \cot(ax+b)dx = \frac{1}{a}\log \sin(ax+b)$

(ix). $\int \sec^2(ax+b)dx = \frac{1}{a}\tan(ax+b)$

(x). $\int co\sec^2(ax+b)dx = -\frac{1}{a}\cot(ax+b)$

(xi). $\int \sec(ax+b)\tan(ax+b)dx = \frac{1}{a}\sec(ax+b)$

(xii). $\int co\sec(ax+b)\cot(ax+b)dx = -\frac{1}{a}co\sec(ax+b)$

(xiii). $\int \sec(ax+b)dx = \frac{1}{a}\log\left(\sec(ax+b)+\tan(ax+b)\right)$

(xiv). $\int \text{cosec}(ax+b)dx = \frac{1}{a}\log\left(\text{cosec}(ax+b)-\cot(ax+b)\right)$

(xv). $\int \sinh(ax+b)dx = \frac{1}{a}\cosh(ax+b)$

(xvi). $\int \cosh(ax+b)dx = \frac{1}{a}\sinh(ax+b)$

(xvii). $\int \tanh(ax+b)dx = \frac{1}{a}\log\cosh(ax+b)$

(xviii). $\int \coth(ax+b)dx = \frac{1}{a}\log\sinh(ax+b)$

(xix). $\int \text{sech}^2(ax+b)dx = \frac{1}{a}\tanh(ax+b)$

(xx). $\int co\sec^2(ax+b)dx = -\frac{1}{a}\coth(ax+b)$

(xxi). $\int \text{sech}(ax+b)\tanh(ax+b)dx = -\frac{1}{a}\text{sech}(ax+b)$

(xxii). $\int \text{cosech}(ax+b)\coth(ax+b)dx = -\frac{1}{a}\text{cosech}(ax+b)$

(xxiii). $\int \frac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\frac{x}{a}$

(xxiv). $\int \frac{dx}{a^2+x^2} = \frac{1}{a}\tan^{-1}\frac{x}{a}$

(xxv). $\int \frac{dx}{\sqrt{a^2+x^2}} = \sinh^{-1}\frac{x}{a} = \log\left(x+\sqrt{x^2+a^2}\right)$

(xxvi). $\int \frac{dx}{\sqrt{x^2-a^2}} = \cosh^{-1}\frac{x}{a} = \log\left(x+\sqrt{x^2+a^2}\right)$

(xxvii). $\int \frac{dx}{x\sqrt{x^2-a^2}}dx = \sec^{-1}\frac{x}{a}$

(xxviii). $\int \sqrt{a^2-x^2}dx = \frac{x}{2}\sqrt{a^2-x^2}+\frac{a^2}{2}\sin^{-1}\frac{x}{a}$

(xxix). $\int \sqrt{a^2+x^2}\,dx = \frac{x}{2}\sqrt{a^2+x^2}+\frac{a^2}{2}\sinh^{-1}\frac{x}{a} = \frac{x}{2}\sqrt{a^2+x^2}+\frac{a^2}{2}\log\left(x+\sqrt{x^2+a^2}\right)$

(xxx). $\int \sqrt{x^2-a^2}\,dx = \frac{x}{2}\sqrt{x^2-a^2}-\frac{a^2}{2}\cosh^{-1}\frac{x}{a} = \frac{x}{2}\sqrt{x^2-a^2}-\frac{a^2}{2}\log\left(x+\sqrt{x^2-a^2}\right)$

(xxxi). $\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\log\frac{x-a}{x+a}, x>a = \frac{1}{2a}\log\frac{a-x}{a+x}, x<a$

(xxxii). $\int e^{ax}\sin(bx+c)\,dx = e^{ax}\frac{a\sin(bx+c)-b\cos(bx+c)}{a^2+b^2}$

(xxxiii). $\int e^{ax}\cos(bx+c)\,dx = e^{ax}\frac{a\cos(bx+c)+b\sin(bx+c)}{a^2+b^2}$

(xxxiv). $\int f.g\,dx = f\int g\,dx - \int\left[f'.\int g\,dx\right]dx$, where f, g are functions of x

(xxxv). $\int_0^{\pi/2} \sin^m x\cos^n x\,dx = \frac{\Gamma\frac{m+1}{2}\Gamma\frac{n+1}{2}}{2\Gamma\frac{m+n+2}{2}}$ (Gamma Function)

Important integration and Their Hints:

Integration	**Hints**
(i). $\sqrt{a^2+x^2}$	Put x = a tan θ ot a cot θ
(ii). $\sqrt{x^2-a^2}$	Put x = a sec θ or a cosec θ
(iii). $\sqrt{a^2-x^2}$	put x = a sin θ or acos θ
(iv). $\sqrt{\frac{a+x}{a-x}}$ or $\sqrt{\frac{a-x}{a+x}}$	put x = a cos 2 θ
(v). $\sqrt{ax^2+bx+c}$	by making perfect square
(vi). $\frac{1}{X\sqrt{Y}}$	Put $\sqrt{Y}=y$

Where X, Y are both linear.

Definite Integrals:

The definite integral is denoted by $\int_a^b f(x)\,dx$

and is read as "the integral of the function f(x) w.r.t. 'x' from x = a to x = b",

Let $\frac{d}{dx}F(x)=f(x)$ then $\int_a^b f(x)\,dx=F(b)-F(a)$;

Where F(b) and F(a) are the values of the functions F(x) at x = b and x = a respectively

Property I. $\int_a^b f(x)\,dx = \int_a^b f(t)\,dt$

Property II. $\therefore \int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$

Property III. $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$

Property IV. $\int_0^a f(x)dx = \int_0^a f(a-x)dx$

Property V. $\int_{-a}^a f(x)dx = 2\int_0^a f(x)dx$, if f(x) is an even function,

= 0 if f(x) is an odd function.

Property VI. $\int_0^{2a} f(x)dx = 2\int_0^a f(x)dx$, if f (2a – x) = f(x)

= 0 if f(2a – x) = -f(x)

Wallis formula:

$$\int_0^{\pi/2} \sin x^n dx = \int_0^{\pi/2} \cos x^n dx$$

$$= \frac{(n-1)(n-3)(n-5).....}{n(n-2)(n-4).....} \times \left(\frac{\pi}{2}, \text{Only if n is even}\right)$$

$$I_n = \int_0^{\frac{\pi}{2}} \sin x^n dx = -\left|\frac{\sin^{n-1} x \cos x}{n}\right|_0^{\frac{\pi}{2}} + \left(\frac{n-1}{n}\right)\int_0^{\frac{\pi}{2}} \sin^{(n-2)} x dx$$

$$I_n = \frac{(n-1)}{n} I_{n-2}$$

Case-I. When n is odd,

$$I_{n-2} = \left(\frac{n-3}{n-2}\right) I_{n-4}, \; I_{n-4} = \left(\frac{n-5}{n-4}\right) I_{n-6}$$

From these we get

$$I_n = \frac{(n-1)(n-3)(n-5)......2}{n(n-2)(n-4).....3.1}$$

Case-II. When n is even,

$$I_{n-2} = \frac{(n-3)}{(n-2)} I_{n-4},$$

$$I_{n-4} = \frac{(n-5)}{(n-4)} I_{n-6}$$

From these, we obtain

Note.18:

Reduction formula for $\int \sin^m x dx.\cos^n x dx$

Here a generalized formula

$$\int_0^{\frac{\pi}{2}} \sin^m x.\cos^n x.dx = \frac{(m-1)(m-3)...(n-1)(n-3)}{(m+n)(m+n-2)(m+n-4)...} \times K$$

When m and n both are even $K = \frac{\pi}{2}$

Otherwise, K =1,

Note.19: Leibnitz rule of Differentiation:

Let f (x, t) is integrand which is function of two variable x and t then

$$\frac{d}{dx}\left[\int_{\varphi(x)}^{\psi(x)} f(x,t)\,dt\right] = \int_{\varphi(x)}^{\psi(x)} \frac{\partial}{\partial x} f(x,t)\,dt + \frac{d\psi}{dx}.f(x,\psi) - \frac{d\varphi}{dx}.f(x,\varphi)$$

(i). Take care, here $\psi(x)$ and $\varphi(x)$ are replaced in place of t in 2nd & 3nd term.

(ii). If integrand is function of 't' alone then

$$\frac{d}{dx}\left[\int_{\varphi(x)}^{\psi(x)} f(t)\,dt\right] = \frac{d\psi}{dx}.f(\psi) - \frac{d\varphi}{dx}.f(\varphi)$$

Gamma Functions:

$\Gamma n = \int_0^\infty e^{-x}x^{n-1}dx$, n > 0 and n may not be an integral value.

Use Formula $\Gamma n\Gamma(1-n) = \frac{\pi}{\sin n\pi}$

Beta function:

$\beta(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}\,dx,$ m, n > 0 not necessarily an integer.

Property:

(i). Beta function is symmetrical about m and n i.e. β (m, n) = β (n, m)

(ii). Another useful transformation of beta functions $\beta(m,n) = \int_0^\infty \frac{x^{n-1}}{(1+x)^{m+n}}dx$

(iii). Relation between beta and gamma function $\beta(m,n) = \frac{\Gamma m\Gamma n}{\Gamma(m+n)}$

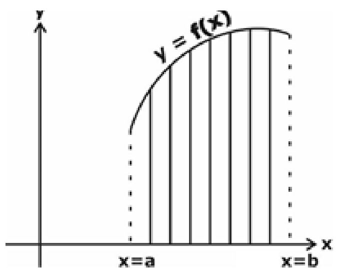
(iv). $\int_0^{\pi/2} \sin^m\theta\cos^n\theta\,d\theta = \frac{\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}{2\Gamma\left(\frac{m+n+2}{2}\right)}$ where m > -1 and n > -1

Areas of Cartesian curves:

Theorem: -

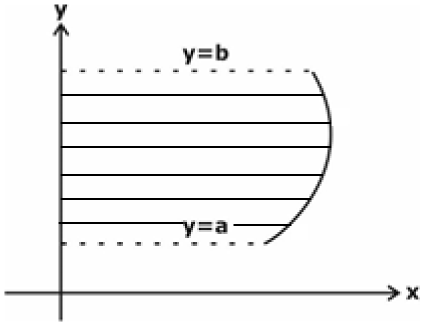
(i). Area bounded by the curve y = f(x) the x-axis and the ordinates

x = a, x = b is $\int_a^b y\,dx = \int_a^b f(x)\,dx$

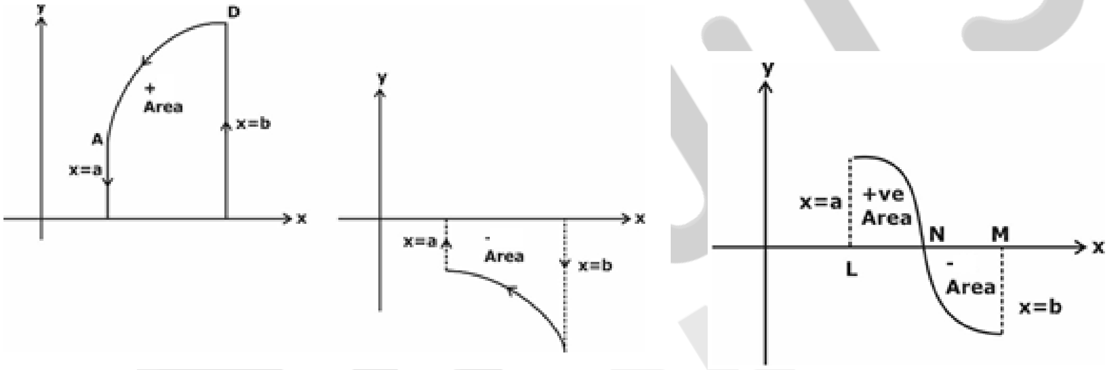


(ii). The area bounded by the curve x = f(y), the x-axis and the abscissa y = a, y = b is

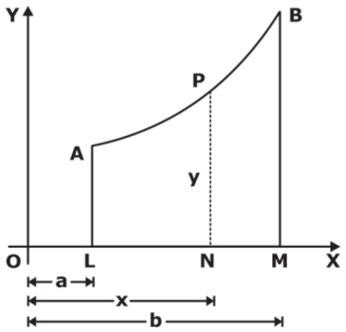
$$\int_a^b x\,dy = \int_a^b f(y)\,dy$$



Sign of an area:



Length of Curves:



(i). The length of the arc of the curve y = f(x) between the points where x = a and x = b is

$$\int_a^b \sqrt{\left[1+\left(\frac{dy}{dx}\right)^2\right]}\,dx$$

(ii). The length of the arc of the curve x = f(y) between the point where y =a and y = b, is

$$\int_a^b \sqrt{\left[1+\left(\frac{dx}{dy}\right)^2\right]}\,dy$$

(iii). The length of the arc of the curve $x = f(t), y = \phi(t)$ between the points where t =a and t =b, is

$$\int_a^b \sqrt{\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right]}dt$$

(iv). The length of the arc of the curve $r = f(\theta)$ between the point where $\theta = \alpha$ and $\theta = \beta$, is

$$\int_\alpha^\beta \sqrt{\left[r^2 + \left(\frac{dr}{d\theta}\right)^2\right]}d\theta$$

(v). The length of the arc of the curve $\theta = f(r)$ between the point where r = a and r = b, is

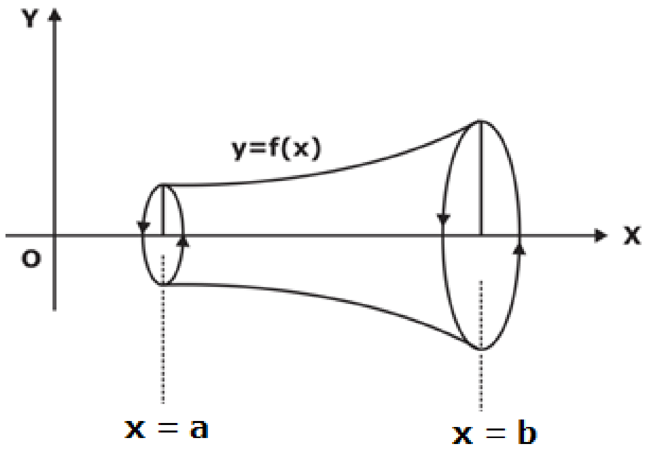
$$\int_a^b \sqrt{\left[1 + \left(r\frac{d\theta}{dr}\right)^2\right]}dr$$

Volumes of Revolution:

Revolution about x-axis:

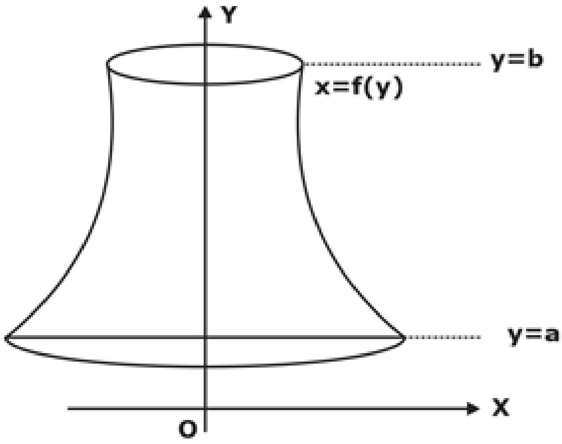
The volume of the solid generated by the revolution about the x-axis, of the area bounded by the curve y =f(x), the x-axis and the ordinates $x = a, x = b$ is $\int_a^b \pi y^2 dx$.

Let AB to the curve y =f(x) between the ordinates A(x = a) and B (x=b).



Revolution about the y-axis:

The volume of the solid generated by the revolution, about y-axis, of the area, bounded by the curve x = f(y), the y-axis and the abscissa y = a, y = b is $\int_a^b \pi x^2 dy$



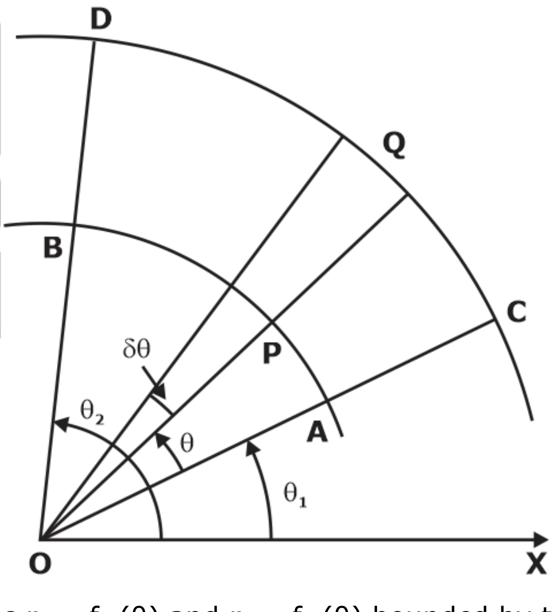
Change of order of integration:

In a double integral with variable limits, the change of order of integration changes the limit of integration. While doing so, sometimes it is required to split up the region of integration and the given integral is expressed as the sum of a number of double integrals with changed limits.

The change of order of integration quite often facilities the evaluation of a double integral.

Double Integrals in Polar Coordinates:

To evaluate $\int_{\theta_1}^{\theta_2}\int_{r_1}^{r_2} f(r,\theta)drd\theta$, we first integrate w.r.t. r between limits $r = r_1$ and $r = r_2$ keeping θ fixed and the resulting expression in integrated w.r.t. θ from θ_1 to θ_2. In this integral r_1, r_2 are functions of θ and θ_1, θ_2 are constants.



Here AB and CD are the curves $r_1 = f_1(\theta)$ and $r_2 = f_2(\theta)$ bounded by the lines $\theta = \theta_1$ and $\theta = \theta_2$. PQ is a wedge of angular thickness δθ.

Then $\int_{r_1}^{r_2} f(r,\theta)dr$ indicates that the integration is along PQ from P to Q while the integration w.r.t. θ corresponds to the turning of PQ from AC to BD.

Thus, the whole region of integration is the area ACDB. The order of integration may be changed with appropriate changes in the limits.

Triple Integrals:

Consider a function f (x, y, z) defined at every point of the 3-dimensional finite region V. Divide V into n elementary volumes δV_1, δV_2, ……, δV_n. Let (x_r, y_r, z_r) be any point within the r^{th} sub-division δV_r. Consider the sum $\sum_{r=1}^{\infty} f(x_r, y_r, z_r)\delta V_r$

The limit of this sum, if it exists, as $n \to \infty$ and $\delta V_r \to 0$ is called the triple integral of

f(x, y, z) over the region V and is denoted by $\iiint f(x,y,z)\,dV$

For purpose of evaluation it can also be expressed as the repeated integral

$\int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} f(x,y,z)\,dxdydz$

If x_1, x_2 are constants; y_1, y_2 are either constants or functions of x and z_1, z_2 are either constants or functions of x and y, then this integral is evaluated as follows.

First f(x, y, z) is integrated w.r.t. z between the limits z_1 and z_2 keeping x and y fixed. The resulting expression is integrated w.r.t. y between the limits y_1 and y_2 keeping x constant. The result just obtained is finally integrated w.r.t. x from x_1 to x_2.

Thus

$$I = \boxed{\int_{x_1}^{x_2} \boxed{\int_{y_1(x)}^{y_2(x)} \boxed{\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)\,dz}\,dy}\,dx}$$

Where the integration is carried out from the innermost rectangle to the outermost rectangle. The order of integration may be different for different types of limits.

CHAPTER 3: PROBABILITY & STATISTICS

1. PROBABILITY

DEFITITION

A. Random Experiments-

For any invention, number of experiments are done. Consider an experiment whose results is not predictable under almost similar working condition then these experiments are known as Random Experiments**.**

These are some cases of random experiments-

Case 1: If we toss a coin, then the result of the experiment whether it is going to come head or tail is not predictable under very similar conditions.

Case 2: If we throw a dice, then the outcome of this cannot be predicted with certainty that which number is going to turn.

B. Sample Space, S –

Each random experiments of some possible outcomes, if we make a set of all the possible outcomes of random experiments then Set 'S' is known as the Sample Space & each possible outcome is Sample Point.

Case 1: If we roll a die, then set of all possible outcomes, is given by {1, 2, 3, 4, 5, 6} then this will be the sample space of given experiment and 1, 2, 3, 4, 5 & 6 are sample points.

Similarly, if our objective is getting odd number on rolling same die then the Sample space will be {1, 3, 5} & for even number Sample space will be {2, 4, 6}.

Case 2: If the outcome of our experiment is to determine whether a male is married or not then our Sample space will be {Married, Unmarried}.

C. Event, E

An event is a subset A of the sample space S, i.e., it is a set of possible outcomes.

An Event is a set of consisting some of the possible outcomes from the sample space of the experiment.

Case 1: On tossing a coin twice, all possible outcomes (Sample space) is {HH, HT, TH, TT} whereas {HH}, {HH, TT}, {HT, HH}, {HH, HT, TT} are the events.

If the event consists only single outcome, then it is known as **Simple Events**.

If the events consist of more than one outcome, then it is known as **Compound Events**.

Types of Events-

(i) Complementary Event – Any Event E^C is called complementary event of event E if it consists of all possible outcomes of sample space which is not present in E.

Exp. - If we roll a die, then set of all possible outcomes, is given by {1, 2, 3, 4, 5, 6}.

An event of getting outcome in multiple of 3 is

E (multiples of 3) = {3,6}

Then, E^C = {1,2,4,5}

(ii) Equally Likely Event – if any two event of sample space are in such a way that the chance of both the events are equal, then this type of events is known as Equally likely events.

Exp. – Chances of a new-born baby to be a boy or girl is 50% means either it can be a girl or boy.

(iii) Mutually Exclusive Events – Two events are called as mutually exclusive when occurring of both the simultaneously is not possible.

If E_1 & E_2 are mutually exclusive, then $E_1 \cap E_2 = \varphi$

Exp. – if we toss a coin then either head or tail can occur, occurrence of both simultaneously is not possible.

(iv) Collectively Exhaustive Events - Two events are called as Collectively exclusive when sample points of both the events incudes all the possible outcomes.

If E_1 & E_2 are mutually exclusive, then $E_1 \cup E_2 = S$

Exp. – if we toss a coin & E_1 is the occurrence of head and E_2 is the occurrence of a tail. Then both the events are collectively exhaustive because both of them collectively include all possible outcomes.

(v) Independent Events – Two events are called as independent when occurring of 1st event does not affect the occurrence of 2nd.

Exp. – On rolling two dice simultaneously, occurrence of 5 in 1st die does not affect the occurrence of 4 in second die. Their occurrence is independent to each other.

Definition of Probability – If an experiment is conducted under essentially given condition up to 'n' times and let 'm' cases are favourable to an event 'E', then probability of 'E' is denoted by P(E) & defined as

$$P(E) = \frac{\text{Number of favourable cases to E}}{\text{Total number of Events}} = \frac{m}{n}$$

$$P(\bar{E}) = \frac{\text{Number of non favourable cases to E}}{\text{Total number of Events}} = 1 - \frac{m}{n}$$

$$P(\bar{E}) == 1 - P(E)$$

$$P(E) + P(\bar{E}) = 1$$

The Axioms of Probability

Consider an Experiment whose sample space is S. For each event E of the sample space, we associate a real number P(E). Then P is called a probability function, and P(E) the probability of the event E, then P(E) will satisfy the following axioms.

Axiom 1:

For every event E, $P(E) \geq 0$

Probability of an event can never be negative.

Axiom 2:

In case of sure or certain event E, $P(E) = 1$

Probability of an event with 100% surety is 1.

Axiom 3:

For any number of **mutually exclusive** events E_1, E_2, ….,

$P(E_1 \cup E_2 \cup E_3 \ldots) = P(E_1) + P(E_2) + P(E_3) \ldots\ldots$

In particular, for two mutually exclusive events E_1, E_2

$P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Some Important Theorems on Probability

From the above axioms we can now prove various theorems on probability

Theorem 1: For every event E,

$0 \le P(E) \le 1$,

i.e., probability lies between 0 and 1.

Theorem 2: $P(\Phi) = 0$

i.e., the impossible event has probability zero.

Theorem 3: If E^C is the complementary of E i.e. that event E will not happen, then

$P(E^C) = 1 - P(E)$

DeMorgan's Law

1. $\left(\bigcup_{i=1}^{i=n} E_i\right)^C = \bigcap_{i=1}^{i=n} E_i^C$

2. $\left(\bigcap_{i=1}^{i=n} E_i\right)^C = \bigcup_{i=1}^{i=n} E_i^C$

Exp.

Let E_1, E_2 are two events,

then

$(E_1 \cup E_2)^C = E_1^C \cap E_2^C$

$E_1 \cup E_2$ is the event either E_1 or E_2 (or both).

$E_1^C \cap E_2^C$ is the event neither E_1 nor E_2.

De-Morgan's law is often used to find the probability of neither E_1 nor E_2.

Corollary:1

From Theorem 3

If E^C is the complement of E, then

$P(E^C) = 1 - P(E)$

And from De-Morgen's theorem

$(E_1 \cap E_2)^C = E_1^C \cup E_2^C$

Combining both the results

$P\left(E_1^C \cap E_2^C\right) = P\left((E_1 \cup E_2)^C\right) = 1 - P(E_1 \cup E_2)$

$P(\text{neither } E_1 \text{ nor } E_2) = 1 - P(\text{Either } E_1 \text{ or } E_2)$

Theorem 4: If $E = E_1 \cup E_2 \cup E_3 .... \cup E_n$, where $E_1, E_2, ... E_n$ are mutually exclusive events, then

$P(E) = P(E_1) + P(E_2) + ... + P(E_n) = 1$

If E = S, the sample space, then

$P(E_1) + P(E_2) + ... + P(E_n) = 1$

Theorem 5: If A and B are any two events, then

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

If both the events are Mutually Exclusive,

Then,

$P(E_1 \cap E_2) = 0$

Thus,

$P(E_1 \cup E_2) = P(E_1) + P(E_2)$

More generally,

if E_1, E_2, E_3 are any three events, then

$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_3 \cap E_1) + P(E_1 \cap E_2 \cap E_3)$

Theorem 6: If E_1 & E_2 are two independent events, then

$P(E_1 \cap E_2) = P(E_1) \times P(E_2)$

Then, $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Will converts into $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1) \times P(E_2)$ (for independent events)

Theorem 7: If an event E must result in the occurrence of one of the mutually exclusive events E_1, $E_2, ... E_n$, then

$P(E) = P(E \cap E_1) + P(E \cap E_2) + ... + P(E \cap E_n)$

This is also known as Rule of total probability.

Theorem 8: Conditional Probability

Let E_1 and E_2 be two events such that $P(E_1) > 0$.

The probability of E_2, given that E_1 has occurred denoted by $P(E_2/E_1)$ and given by,

$$P(E_2 \mid E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \qquad P(E_1) \neq 0$$

or $P(E_1 \cap E_2) = P(E_1)\, P(E_2 \mid E_1)$

Similarly,

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \qquad P(E_2) \neq 0$$

This rule is also known as multiplication rule of probability.

- if E_1 & E_2 are independent events

Then, $P(E_1 \cap E_2) = P(E_1) \times P(E_2)$

$$P(E_2 \mid E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{P(E_1) \times P(E_2)}{P(E_1)}$$

$$P(E_2 \mid E_1) = P(E_2)$$

Similarly,

$$P(E_1 \mid E_2) = P(E_1)$$

For any three events E_1, E_2, E_3,

we have

$P(E_1 \cap E_2 \cap E_3) = P(E_1)\,(E_2 | E_1)\,P(E_3 | E_1 \cap E_2)$

In words, the probability that E_1 and E_2 and E_3 all occur is equal to the probability that E_1 occurs times the probability that E_2 occurs given that E_1 has occurred times the probability that E_3 occurs given that both E_1 and E_2 have occurred.

Theorem 9: Bayes' Theorem

It is an extended form of Conditional probability.

Suppose that E_1, E_2, E_3 ……. E_m are the mutually exclusive events whose union is the sample space and E is an event

Then, as per the Bayes' theorem

$$P(E_n \mid E) = \frac{P(E_n) \times P\left(E/E_n\right)}{\sum_{i=1}^{n} P(E_i) \times P\left(E/E_i\right)}$$

In general form,

If A and B are two mutually exclusive event

$$P(A \mid E) = \frac{P(A \cap E)}{P(E)} = \frac{P(A \cap E)}{P(A \cap E) + P(B \cap E)}$$

$$P(A \mid E) = \frac{P(A) \times P\left(E/A\right)}{P(A) \times P\left(E/A\right) + P(B) \times P\left(E/B\right)}$$ (Using theorem 8 & 9)

2. PROBABILITY DISTRIBUTION

(A) Random Variables –

Suppose that to each point of a sample space we assign a number. We then have a function defined on the sample space. This function is called a random variable or more precisely a random function. It is usually denoted by a capital letter such as X or Y. Random variable X associated with the outcome of an experiment which is not certain, and its value depend upon the chance.

If a random variable takes a finite set of values then it is called as **Discrete random variable**, whereas when a random variable takes an infinite set of values (or any value from a continuous range or graph) then it is called as **Continuous Random variable**.

Based on this, we can divide distributions also in two category-

(i) Discrete probability Distribution

(ii) Continuous probability distribution

(B) Discrete Probability Distributions

Let X be a discrete random variable and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \ldots$, arranged in some order.

These values are assumed with probabilities given by

$P(X = x_k) = f(x_k)$ where $k = 1, 2, \ldots \ldots$ (1)

It is convenient to introduce the probability function, also referred to as probability distribution, given by

$P(X = x) = f(x)$

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

The properties of discrete probability distribution are

(i) $P(x_i) \geq 0$ for all values of i

(ii) $\sum P(x_i) = 1$

(iii) Mean of Random variable, μ (or E)

$E(x) = \mu = \sum x_i P(x_i)$

It is also called expected value (Expectation) or average value of random variable.

(iv) Variance of Random variable, V (σ^2)

$\sigma^2 = V(x) = \sum (x_i - \mu)^2 P(x_i)$

As we know

$\sum P(x_i) = 1$, $\mu = \sum x_i P(x_i)$

$\sigma^2 = \sum x_i^2 P(x_i) - \mu^2$

(v) Standard deviation, σ (SD) – it is square root of the variance. It is the measure of variation amongst data.

Types of Discrete distributions are

(i) Binomial Distribution

(ii) Poisson distribution

(iii) Geometric distribution

(C) Continuous Random Variables

A non-discrete random variable X is said to be continuous, or simply continuous, if its distribution function may be represented as

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx \quad (-\infty < x < \infty)$

where the function f(x) has the properties

1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

3. $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

4. $V(X) = E(x^2) - (E(x))^2$

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \left(\int_{-\infty}^{\infty} x f(x)\, dx\right)^2$$

It follows from the above that if X is a continuous random variable, then the probability that X takes on any one value is zero.

Whereas the interval probability that X lies between two different values, say, a and b, is given by

$$P(a < X < b) = \int_a^b f(x)\, dx$$

$$P(a < X < b) = P(a < X \le b) = P(a \le X < b) = P(a \le X \le b) = \int_a^b f(x)\, dx$$

Some examples of continuous distribution area as follows

(i). Normal Distribution

(ii). Exponential Distribution

(iii). Uniform Distribution

D) Properties of Expectation and Variance:

If x_1 and x_2 are two random variance and a and b are constants,

$E(ax_1 + b) = a\,E(x_1) + b$

$V(ax_1 + b) = a^2\,V(x_1)$

$E(ax_1 + bx_2) = a\,E(x_1) + b\,E(x_2)$

$V(ax_1 + bx_2) = a^2V(x_1) + b^2V(x_2) + 2ab\,Cov(x_1, x_2)$

Where Cov (x_1, x_2) represents the covariance between x_1 and x_2, which is the ratio of standard deviation and mean.

If x_1 and x_2 are independent, then $Cov(x_1, x_2) = 0$

Hence, above formula reduces to

$V(ax_1 + bx_2) = a^2V(x_1) + b^2V(x_2)$

If x_1 and x_2 are independent, then

$E(x_1 \times x_2) = E(x_1) \times E(x_2)$

Binomial Distribution –

Suppose that we have an experiment such as tossing a coin or rolling a die repeatedly or choosing a marble from an urn repeatedly. Each toss or selection is called a trial. In any single trial there will be a probability associated with a particular event such as head on the coin, 4 on the die, or selection of a particular colour of marble.

In some case this probability will not change from one trial to the next (as in tossing a coin or die). Such trials are then said to be independent and are often called Bernoulli trials.

Let p be the probability that an event will happen in any single Bernoulli trial (called the probability of success). Then q = 1 – p is the probability that the event will fail to happen in any single trial (called the probability of failure). The probability that the event will happen exactly x times in n trials (i.e., x times successes and (n – x) times failures will occur) is given by the probability function

$$F(x) = P(X = x) = n_{C_x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

where,

the random variable X denotes the number of successes in n trials and x = 0, 1, ... n.

Case – 1

When p = q,

$F(x) = P(X = x) = {}^{n}C_{x} p^x q^{n-x} = {}^{n}C_{x} p^x p^{n-x} = {}^{n}C_{x} p^n$

some assumptions are made by Bernoulli before reaching the conclusion

1. There is only 2 outcomes are possible, success or failure.
2. Probability of success (p) and probability of failure q remains same from trial to trial.
3. The trials event are independent. i.e., The outcome of one trial does not affect the subsequent trials.

Some Properties of the Binomial Distribution

Mean/ Expected value	$\mu = np$
Variance	$\sigma^2 = npq$
Standard deviation	$\sigma = \sqrt{npq}$

Poisson's Distribution –

Let X be a discrete random variable that can take on the values 0, 1, 2, . . . such that the probability function of X is given by

$F(x) = P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ where, x = 0, 1, 2…….

where $\lambda(>0)$ is a given positive constant. This distribution is called the Poisson distribution and a random variable having this distribution is said to be Poisson distributed.

Some Properties of the Poisson Distribution

Mean/ Expected value	$\mu = \lambda$
Variance	$\sigma^2 = \lambda$
Standard deviation	$\sigma = \sqrt{\lambda}$

From the table, we can see that expected value and variance is same for Poisson's distribution.

Geometric distribution –

Consider repeated trial of Bernoulli experiment with probability of success p, and failure q=(1-p). If the experiment is repeated until success is not achieved, then the distribution of variable is given by geometric distribution.

If experiment is performed "k" times, then experiment must be failed in 'k-1' times.

Then probability of success is given by

$P(X = k) = pq^{k-1}$

Some Properties of the Geometric Distribution

Mean/ Expected value	$\mu = \dfrac{1}{p}$
Variance	$\sigma^2 = \dfrac{q}{p^2}$
Standard deviation	$\sigma = \sqrt{\dfrac{q}{p^2}}$

Normal Distribution:

One of the most important examples of a continuous probability distribution is the normal distribution, some-times called the Gaussian distribution.

The density function for this distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{where, } -\infty < x < \infty$$

where μ and σ are the mean and standard deviation, respectively.

Standard normal distribution –

If we replace $\mu = 0$ & $\sigma = 1$ then normal distribution will reduce to standard normal distribution.

In such cases the density function for Z will be reduced to

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

This is often referred to as the standard normal density function.

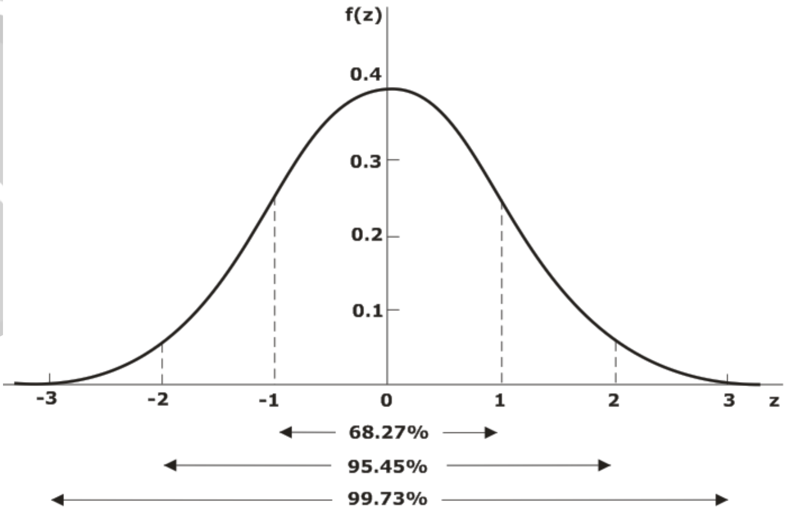
The corresponding distribution function is given by

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-u^2/2} du$$

Z be the standardized variable corresponding to X, i.e.

$$Z = \frac{X-\mu}{\sigma}$$

A graph of the density function sometimes called the standard normal curve, is shown in figure. It is a bell-shaped curve which is symmetric about mean and area under the curve is equal to 1 unit.



In this graph we have indicated the areas within 1, 2, and 3 standard deviations of the mean (i.e., between z = – 1 and + 1, z = –2 and +2, z = –3 and +3) as equal, respectively, to 68.27%, 95.45% and 99.73% of the total area, which is 1.

This means,

$P(-1 \le Z \le 1) = 0.6827 = 68.27\%$

$P(-2 \le Z \le 2) = 0.9545 = 95.45\%$

P (− 3 ≤ Z ≤ 3) = 0.9973 = 99.73%

Exponential Distribution:

It is a continuous random variable whose density function is given by

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geq 0 \\ 0 & x \text{ less than zero} \end{cases}$$

Its probability distribution function will be given as,

$$F(x) = P(x \leq k) = \int_0^k \alpha e^{-\alpha x} dx \quad \text{where } k \geq 0$$

$$F(x) = P(x \leq k) = 1 - e^{-\alpha k}$$

$$\text{Mean, } \mu = \frac{1}{\alpha}$$

$$\text{Variance, } \sigma^2 = \frac{1}{\alpha^2}$$

$$\text{Standard deviation, } \sigma = \frac{1}{\alpha}$$

Continuous Uniform Distribution

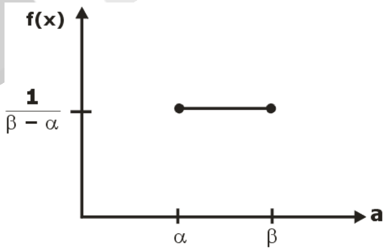
In general, we say that X is a uniform random variable on the interval (a, b). If its probability density function is given by:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

The distribution given by above density function is uniform distribution.

Since f(x) is a constant, all values of x between α and β are equally likely (uniform).

Graphical Representation:



For Discrete Uniform Distribution:

$$\text{Mean} = E[x] = \int_\alpha^\beta x.f(x)dx$$

$$\mu = \frac{1}{\beta - \alpha}\int_\alpha^\beta x dx$$

$$\mu = \frac{\beta + \alpha}{2}$$

$$E(x) = \mu = \frac{\beta + \alpha}{2}$$

$$\text{Variance} = V(X) = \int_{\alpha}^{\beta} x^2 f(x)dx$$

$$\text{Or } \sigma^2 = V(X) = \frac{(\beta - \alpha)^2}{12}$$

3. STATISTICS

(i) Introduction

Statistics deals with the method of collection, classification, and analysis of numerical data for drawing valid conclusion and making reasonable decision. It is a branch of mathematics which gives us the tools to deal with large quantities of data.

In this method of calculation, we find a representative value for the given data. This value is called the **measure of central tendency**.

(i) mean (arithmetic mean)

(ii) median

(iii) mode

These are the three measures of central tendency

Measure of central tendency indicates an average value of given data.

But the measures of central tendency are not sufficient to give complete information about a given data. Variability is another factor which is required to be studied under statistics.

Like 'measures of central tendency' a single number is assigned to describe variability of the data. This single number is called a '**measure of dispersion.**

(i) Standard deviation

(ii) Variance

(iii) Coefficient of Variation

(iv) Range

'Measures of Dispersion' denotes the scattering of the data from a fixed point and that fixed point is measure of central tendency. It tells about how data is closely packed around the central mean value

Arithmetic Mean

Arithmetic Mean for Raw Data

Arithmetic mean is simply the average of the given data that is ratio of sum of the data or observation divided by total number of observations.

If X_1, X_2, X_3............X_n are the observations

Then arithmetic mean will be given as

$$\text{Mean} = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n}$$

It is denoted by $\bar{X}$

Thus, it can also be written as,

$$\bar{x} = \frac{\sum x}{n}$$

$\bar{x}$ - arithmetic mean

x - refers to the value of an observation

n - number of observations.

The Arithmetic Mean for Grouped Data (Frequency Distribution)

if x_1, x_2, …… x_n are observations with respective frequencies f_1, f_2… ….,f_n then this means observation x_1 occurs f_1 times, x_2 occurs f_2 times, and so on, then mean of the data will be given as

$$\text{Mean}, \bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \ldots\ldots + f_nX_n}{f_1 + f_2 + f_3 + \ldots\ldots\ldots + f_n}$$

This formula can be rewritten as

$$\bar{x} = \frac{\sum (f.x)}{\sum f}$$

3.2. Median-

Median is the positional average of the given data, i.e. of we arrange the data in ascending or descending order than the middle term will be the median of the given set of data.

So, we can say that,

For median, it is the 'number of values' greater than the median which balances against the 'number of values' of less than the median

Median for Raw Data

In general, if we have n values of x, they can be arranged in ascending order as:

$x_1 < x_2 < \ldots. < X_n$

Suppose n is odd, then

$$\text{Median} = \frac{N+1}{2}^{th} \text{ value}$$

That is if we arrange data in ascending order, then middle term will be median of the given data.

However, if n is even, we have two middle points

$$\text{Median} = \frac{\left(\frac{n}{2}\right)^{th} \text{value} + \left(\frac{n}{2}+1\right)^{th} \text{value}}{2}$$

That is if we arrange data in ascending order, then mean of the two middle term will be median of the given data.

Median for Grouped Data

1. Identify the median class which contains the middle observation

$$\left(\left(\frac{N+1}{2}\right)^{th} \text{observation}\right)$$

This can be done by observing the first class in which the cumulation frequency is equal to or more than $\frac{N+1}{2}$. Here. N = $\sum f$ = total number of observations.

2. Calculate Median as follows:

$$\text{Median} = L + \left[\frac{\left(\frac{N+1}{2}\right) - (f+1)}{f_m}\right] \times h$$

Where,

L = Lower limit of median class

N = Total number of data items = $\sum f$

f = Cumulative frequency of the class immediately preceding the median class

f_m = Frequency of median class

h = difference between upper limit and lower limit of median class

3.3. Mode –

Mode is defined as the value of the variable which occurs most frequently i.e. the value of maximum frequency.

Mode for Raw Data

In a raw data, most frequently occurring data is mode of that data.

Suppose in a given set of data,

X_1 occurs n_1 times, X_2 occurs n_2 times, X_3 occurs n_3 times………, X_n occurs n_n

And $n_1 > n_2 > n_3 >$………$> n_n$

Then occurrence of X_1 is highest, thus mode of the given data will be X_1.

If there is more than one data which having same & highest frequency, then each of them is a mode.

Thus, we have Unimodal (single mode), Bimodal (two modes) and Trimodal (three modes) data sets.

Mode for Grouped Data

Mode is that value of x for which the frequency is maximum.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies. Here, we can only locate a class with the maximum frequency, called the **modal class**. The mode is a value inside the modal class, and is given by the formula:

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

Where,

L = Lower limit of the modal class

f_0 = Largest frequency (frequency of Modal Class)

f_1 = Largest Frequency in the class preceding the modal class

f_2 = Frequency of the class succeeding to the modal class

h = Width

Properties of Mean, Mode & Median -

In symmetrical distribution, mean, mode & median coincides, but for an unsymmetrical distribution all are different and related by an empirical formula

Empirical mode = 3 (median) –2 (mean)

Skewness - Skewness measure the degree of asymmetry.

There are three types of frequency distributions.

Depending upon the asymmetry, distribution curve can be of 3 types.

(i) Positively skewed distribution

(ii) Symmetric distribution

(iii) Negatively skewed distribution

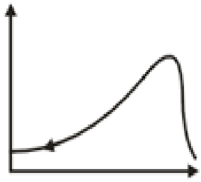
In positively skewed distribution, frequency curve has longer tail to the right i.e. mean is to the right of the mode.

Mode $\leq$ Median $\leq$ Mean

(a) Positively Skewed

In negatively skewed distribution, frequency curve has longer tail to the left i.e. mean is to the left of the mode.

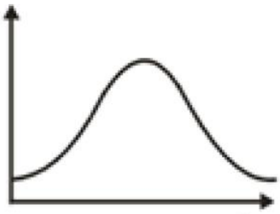
Mean $\leq$ Median $\leq$ Mode



Negatively Skewed

In symmetric distribution, mean, mode & median coincides.

Mean = Median = Mode



Symmetric

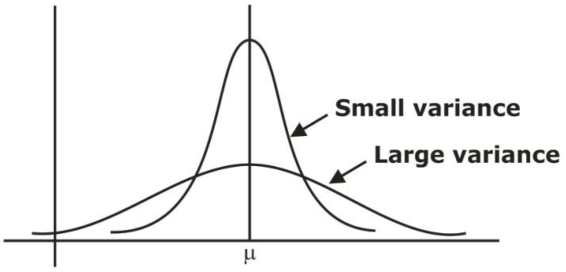
Standard Deviation and Variance

Standard Deviation is a measure of dispersion or variation amongst data. Deviation of an observation x from a fixed value 'a' is the difference (x – a) & the absolute values of these differences are the mean deviation.

But there is possibility that some dispersion comes out positive and some comes out negative, which may cancel each other and results in zero deviation (zero error).

So, to eliminate this, instead of calculating mean deviation, we may square each deviation and obtain the arithmetic mean of squared deviations. This gives us the 'variance' of the values. The positive square root of the variance is called the 'Standard Deviation' of the given values.

If the values tend to be concentrated near the mean, the variance is small; while if the values tend to be distributed far from the mean, the variance is large. The situation is indicated graphically in Figure. For the case of two continuous distributions having the same mean μ.



Comparison of standard deviation of two continuous graph

Standard Deviation for Raw Data

Suppose $x_1, x_2 \ldots . x_n$ are n values of the x,

Then, arithmetic mean will be given as

$$\bar{x} = \frac{\sum x_i}{n}$$

then, $x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x} \ldots . x_n - \bar{x}$ are the deviations of the values of x from $\bar{x}$.

Then Variance of these data will be given as

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

Which can also be written as

$$\sigma^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n^2}$$

The above expression represents the variance whereas square root of the variance will give the standard deviation.

Variance is represented by σ^2 whereas standard deviation is represented by σ.

$$\sigma = +\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2 - \bar{x}^2}{n}} = \sqrt{\frac{n\sum x_i^2 - (\sum x_i)^2}{n^2}}$$

Standard deviation of the combination of two groups –

If m_1, σ_1 are the mean & standard deviation of a sample size of n_1 and m_2, σ_2 are the mean & standard deviation of a sample size of n_2

Then, mean, m & standard deviation, σ of combined sample size $n_1 + n_2$ is given by

$$(n_1 + n_2)\sigma^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1D_1^2 + n_2D_2^2$$

$$\text{where, } D_1 = m_1 - m$$

$$D_2 = m_2 - m$$

m is mean of the combined data which can be calculated as

$$\text{mean, } m = \bar{x} = \frac{n_1x_1 + n_2x_2}{n_1 + n_2}$$

Coefficient of Variation

The ratio of standard deviation to mean is known as coefficient of variation.

The standard deviation is an absolute measure of dispersion and hence cannot be used for comparing variability of 2 data sets with different means. Thus, a new variable is introduced which can compare the variation between the two groups with different mean.

Therefore, such comparisons are done by using a relative measure of dispersion called coefficient of variation (CV).

$$\text{Coefficient of variation, } CV = \frac{\sigma}{\mu}$$

where σ is the standard deviation and μ is the mean of the data set

CV is often represented as a percentage,

$$CV\% = \frac{\sigma}{\mu} \times 100$$

When comparing data sets, the data set with larger value of CV% is more variable (less consistent) as compared to a data set with lesser value of CV%.

4. CORRELATION

Correlation is the method to examine relation between two variables.

When the changes in one variable are associated or followed by changes in the other, is called correlation. Such a data connecting two variables is called bivariate population.

If an increase (or decrease) in the values of one variable corresponds to an increase (or decrease) in the other, the correlation is said to be positive. i.e. Variables moves in same direction.

If the increase (or decrease) in one corresponds to the decrease (or increase) in the other, the correlation is said to be negative. i.e. Variables moves in opposite direction.

If there is no relationship indicated between the variables, they are said to be independent or uncorrelated.

If $x_1, x_2, x_3 \ldots\ldots\ldots\ldots x_n$ are the 'n' observations of 'x' & $y_1, y_2, y_3 \ldots\ldots\ldots\ldots y_n$ are the 'n' observations of y

Then, arithmetic mean is given as

$$\bar{x} = \frac{\sum x}{n}, \quad \bar{y} = \frac{\sum y}{n}$$

Their standard deviation is given as

$$\sigma_x = +\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2 - \bar{x}^2}{n}} = \sqrt{\frac{n\sum x_i^2 - (\sum x_i)^2}{n^2}}$$

$$\sigma_y = +\sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum y_i^2 - \bar{y}^2}{n}} = \sqrt{\frac{n\sum y_i^2 - (\sum y_i)^2}{n^2}}$$

Then,

Covariance of x, y is defined as

$$\mathrm{Cov}(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}$$

The sign of covariance between x and y determines the sign of the correlation coefficient. The standard deviations are always positive. If the covariance is zero, the correlation coefficient is always zero

And coefficient of correlation denoted by 'r' & defined as

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y}$$

By putting the 1st value of standard deviation

We can get,

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \times \sqrt{\sum (y - \bar{y})^2}}$$

Which can also be rewritten as

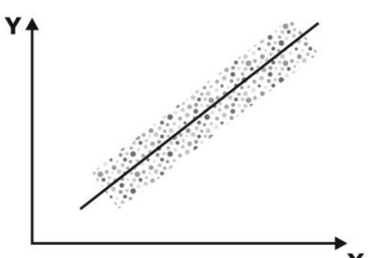
$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

Properties of Correlation Coefficient -

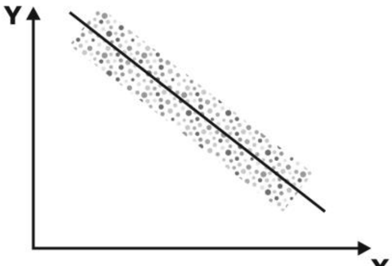
- A negative value of r indicates an inverse relation. A change in one variable is associated with change in the other variable in the opposite direction.
- If r is positive the two variables move in the same direction.
- The value of the correlation coefficient lies between minus one and plus one, $-1 \le r \le 1$. If, in any exercise, the value of r is outside this range it indicates error in calculation.
- If r = 0, the two variables are uncorrelated.

There is no linear relation between them. However other types of relation may be there.

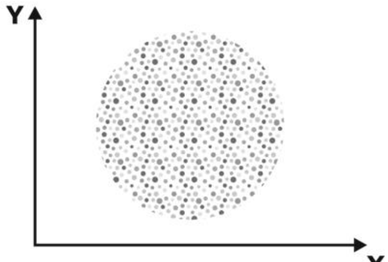
- If r = 1 or r = −1 the correlation is perfect and there is exact linear relation.
- A high value of r indicates strong linear relationship. Its value is said to be high when it is close to +1 or −1.
- A low value of r (close to zero) indicates a weak linear relation.



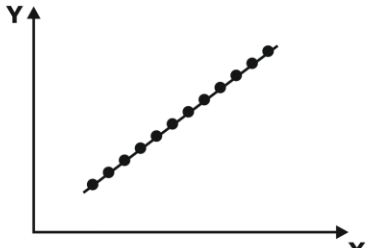
Positive Correlation



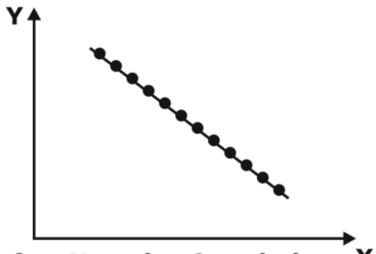
Negative Correlation



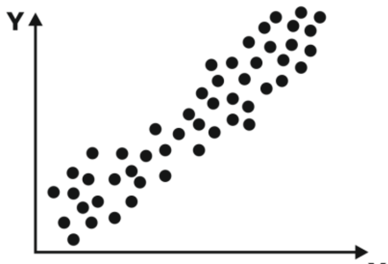
No Correlation



Perfect Positive Correlation



Perfect Negative Correlation



Perfect non-linear relation

5. LINES OF REGRESSION

When comparing two different variables, two questions come to mind: "Is there a relationship between two variables?" and "How strong is that relationship?" These questions can be answered using regression and correlation. Regression answers whether there is a relationship and correlation answers how strong the linear relationship is.

It frequently happens that the dots of the scatter diagram generally, tend to cluster along a well-defined direction which suggests a linear relationship between the variables x and y. Such a line of best fit for the gives distribution of dots is called the line of regression.

There are two such lines, one giving the best possible mean values of y for each 8pecified value of x and the other giving the best possible mean values of x for given values of y. The former is known as the line of regression of y on x and the latter as the line of regression of x on y.

Consider first the line of regression of y on x.

Let the straight line satisfying the general trend of n dots in a scatter diagram be

$y = a + bx$

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

$\frac{1}{n}\Sigma y = a + b \cdot \frac{1}{n}\Sigma x$

$\bar{y} = a + b\bar{x}$ .... (1)

$y = a + bx$

$xy = ax + bx^2$

$\Sigma xy = a\Sigma x + b\Sigma x^2$ ...... (2)

This shows that $(\bar{x}, \bar{y})$, i.e., the means of x and y, lie on (1).

Shifting the origin to $(\bar{x}, \bar{y})$,

Thus replacing, x from $x - \bar{x}$, y from $y - \bar{y}$

Thus, equation will become,

$\Sigma(x-\bar{x})(y-\bar{y}) = a\Sigma(x-\bar{x}) + b\Sigma(x-\bar{x})^2,$

but $a\Sigma(x-\bar{x}) = a\sum x - a\sum \bar{x}$

$\bar{x} = \frac{\sum x}{n} \Rightarrow \sum x = n\bar{x},$

$\sum \bar{x} = \bar{x}\sum 1 = n\bar{x}$

$a\Sigma(x-\bar{x}) = an\bar{x} - an\bar{x} = 0$

$\therefore\ b = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x^2} = r\frac{\sigma_y}{\sigma_x}$ $\left[\because r = \frac{\Sigma XY}{n\sigma_x\sigma_y}\right]$

Thus, the line of best fit becomes $y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$

which is the equation of the line of regression of y on x.

Its slope is called the regression coefficient of y on x.

Interchanging x and y, we find that the line of regression of x on y is

$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$

Thus, the regression coefficient of y on $x = r\sigma_y/\sigma_x$

and the regression coefficient of x on $y = r\sigma_x/\sigma_y$

Note - The correlation coefficient r is the geometric mean between the two regression coefficients.

For $r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y} = r^2$.

6. SAMPLING THEORY

A small section selected from the population is called a sample and the process of drawing sample is called sampling.

It is essential that a sample must be a random selection so that each member of the population has the same chance of being included in the sample. Thus, the fundamental assumption underlying theory of sampling is Random sampling.

A special case of random sampling in which each event has the same probability, P of success and the chance of success of different events are independent whether previous trials have been made or not, is known as simple sampling.

Objectives of sampling –

Sampling aims at gathering the maximum information about the populations with the minimum effort, cost and time. The logic of the sampling theory is the logic of induction in which we pass from a particular (sample) to general (population).

Sampling distribution

Consider all possible samples of size n which can be drawn. from a given population at random. For each sample, we can compute the mean. The means of the samples will not be identical. If we group these different means according to their frequencies, the frequency distribution so formed is known as sampling distribution of the mean.

Similarly, we can have sampling distribution of the standard deviation etc.

While drawing each sample, we put back the previous sample so that the parent population remains the same. This is called sampling with replacement and all the subsequent formulae will pertain to sampling with replacement.

Standard error. The standard deviation of the sampling distribution is called the standard error (S.E.).

Similarly, the standard error of the sampling distribution of means is called standard error of means.

The standard error is used to assess the difference between the expected and observed values.

The reciprocal of the standard error is called precision.

If $n \geq 30$, a sample is called large otherwise small. The sampling distribution of large samples is assumed to be normal.

Testing a hypothesis -

To reach decisions about populations on the basis of sample information, we make certain assumptions about the populations involved. Such assumptions, which may or may not be true, are called statistical hypothesis.

By testing a hypothesis is meant a process for deciding whether to accept or reject the hypothesis or we can say it is the process of cross checking our assumption whether it is correct or not.

The method consists in assuming the hypothesis as correct and then computing the probability of getting the observed sample. If this probability is less than a certain preassigned value, the hypothesis is rejected.

Errors -

If a hypothesis is rejected while it should have been accepted, we say that a Type I error has been committed.

On the other hand, if a hypothesis is accepted while it should have been rejected, we say that Type II error has been made.

The statistical testing of hypothesis aims at limiting the Type I error to a press signed value (upto 5%) and to minimize the Type II error. The only way to reduce both types of errors is by increasing the sample size so that more accurate prediction can be made but increasing the sample size is always not possible.

Null hypothesis –

The hypothesis formulated for the sake of rejecting it, under the assumption that it is true. is called the null hypothesis and is denoted by H_o. To test whether one procedure is better than another, we assume that there is no difference between the procedures. Similarly, to test whether there is a relationship between two variates, we take Ho that there is no relationship. By accepting a null hypothesis, we mean that on the basis of the statistic calculated from the sample, we do not reject the hypothesis. It however, does not imply that the hypothesis is proved to be true. Nor its rejection implies that it is disproved.

Level of significance –

The probability level below which we reject the hypothesis is known as level of significance.

The region in which a sample value falling is rejected then this region is known as critical region.

Generally, it is taken as 5% (2.5% on each side) of the normal curve or 95% of which inside the acceptance region.

Simple sampling of attributes –

Sampling of attributes may be regarded as the selection of sample from a population whose members possesses the attribute K.

The presence of K may be called as success.

Suppose we draw a simple sample of n items.

Since this follows normal distribution

Thus, its mean will be

$$m = \mu = np$$

And standard deviation will be

$$\sigma = \sqrt{npq}$$

Where p & q are the probability of success & failure respectively & n is the sample size.

If we consider the proportion of successes,

Then,

(i) mean proportion of success, $\frac{np}{n} = p$

(ii) standard error of the proportion of success, $\sqrt{n \times \frac{p}{n} \times \frac{q}{n}} = \sqrt{\frac{pq}{n}}$

(iii) Precision of the proportions of success = reciprocal of standard error of the proportion of success, $\sqrt{\frac{n}{pq}}$

****