AI/ML

Maths

SUKALA
Page
Date

Index :

ch : 1 Linear Algebra

ch : 2 Statistics

ch : 3 Probability

ch : 4 Calculus

Ch: 1
Linear Algebra

SUKALA Page Date

Index :

1. Basics of Linear Algebra
   - 1.1 scalars, vectors, Matrices, Tensors
   - 1.2 Indexing and Slicing
2. Vector operations
   - 2.1 vector addition and subtraction
   - 2.2 scalar multiplication
   - 2.3 Dot product (Inner product)
   - 2.4 cross product
   - 2.5 Norms (L1, L2 Norms / magnitude / Length)
   - 2.6 unit vectors and direction
3. Matrix operations
   - 3.1 Matrix addition and subtraction
   - 3.2 scalar multiplication
   - 3.3 Matrix multiplication (Dot product)
   - 3.4 Transpose of matrix
   - 3.5 Element wise operations
4. Special matrices and properties.
   - 4.1 Identity matrix, zero matrix
   - 4.2 Diagonal matrix, orthogonal matrix
   - 4.3 Symmetric matrix, Inverse matrix
   - 4.4 Rank of matrix, Trace of matrix
5. System of Linear Equations.
   - 5.1 Representing Linear system as Matrix equati
   - 5.2 Row Echelon form & Gaussian Elimination
   - 5.3 Matrix Inverse method.
   - 5.4 Cramer's Rule.
6. Matrix Inverse and Determinant
   - 6.1 Determinant of 2X2 and NXN matric
   - 6.2 Properties of determinants.
   - 6.3 Invertibility (condit^n of matrix to be Invertibl

SUKALA
Page
Date

7. Linear Transformations
   7.1 Function and Vector Transformation.
   7.2 Linear Transformation
   7.3 Linear Transformation in Data Science.
   7.4 Projections.
8. Inverse Function or Transformation.
   8.1 Inverse of function
   8.2 Application of Inverse Function
9. Eigen value and Eigen vector
10. Equation of Line, plane and Hyperplane

SUKALA
Page
Date

(1) Basics of Linear Algebra :

(1.1) Scalars, Vectors, Matrices, Tensors

→ Scalars :
scalar is a just single number.
(single value with no direction, no dimension)

eg: Age = 21
Tempreture = 47°c
Price = 69 ₹
any number like 2/7, 6.1 etc.

→ Vectors :
vector is a list of numbers (1D Array).
(A line of numbers - 1D (one direction)). (row or col
eg: Juhil's marks in three subject = [80, 85, 82]
position in 2D space [x=5, y=7]
Daily temprature for week : [30, 29, 41, 20
35, 32, 31].

→ Matrix :
Matrix is combination of rows and columns (2D)
(2D grid like excel sheet)
eg: Marks of 3 students in 4 subjects
[[80, 85, 90, 75],
[70, 72, 92, 81],
[60, 78, 85, 80]] (3×4)

SUKALA
Page
Date

→ Tensor :

Tensor is a multi-dimensiona array (more than 2D).
(can be 3D, 4D ... nD).

eg :- colored image have 3D structure
(Height x width x 3 colors - RGB)

eg :- Shape : [100, 100, 3]
It's mean 100 x 100 pixels and 3 color channels

- video data : [Frames, height, width, channels]

eg with deep explain :

one Box, its length = 5
width = 4
Depth = 2

than it can be simply write like
[5, 4, 2] - but
in tensor it represent like
(2 depth, each with 5 rows & 4 cols)

```
tensor = [
  [ [1, 2, 3, 4],
    [5, 6, 7, 8],
    [9, 10, 11, 12],
    [13, 14, 15, 16],
    [17, 18, 19, 20] ](5,4)        } depth 1
  [ [21, 22, 23, 24],
    [25, 26, 27, 28],
    [29, 30, 31, 32],
    [33, 34, 35, 36],
    [37, 38, 39, 40] ](5,4)        } depth 2
](5,4,2)
```

SUKALA
Page
Date

(1.2) Indexing and Slicing

→ scalar :
scalar is a single value - so no indexing or slicing needed.
eg : x = 10

→ Vector (1D) :
Indexing :
v = [10, 20, 30, 40]
0 1 2 3
Print (v[0]) o/p : 10
Print (v[2]) o/p : 30
Slicing : Extract(a part (range) From vector
Print (v[1:3]) o/p : [20, 30]
Print(v[:2]) o/p : [10, 20]
print (v[2:]) o/p : [30, 40]

→ Matrix (2D) :
Matrix is a list of lists
Indexing :
m = [[1, 2, 3],
[4, 5, 6],
[7, 8, 9]
] (3x3)
Print (m[0]) o/p : [1, 2, 3]
Print [m[1][2]) o/p : 6
↑ row ↑ column
Slicing :
Print (m[:2]) o/p : First two rows
print (m[0] [:2]) o/p : [1, 2]

→ Tensor (3D or more):
It's like a cube (list of matrices)

```
tensor = [
    [ # depth 0
      [1,2],
      [3,4]
    ],
    [ # depth 1
      [5,6],
      [7,8]
    ]
]
```

Indexing:

```
print(t[0])    o/p:: First depth matrix
                     [1,2]
                     [3,4]
Print(t[1][0][1])
      ↑  ↑  ↑        o/p : 6
   Depth row col
```

Slicing:

```
Print(t[1][1][0:1]    o/p: 7
      Depth row col
```

SUKALA
Page
Date

(2) Vector Operations :

(2.1) vector addition and subtraction

Note: vectors must be the same size (same no. of dimensions)

- vector Addition :

$\vec{a} = [2, 4, 6]$, $\vec{b} = [1, 3, 5]$

$\vec{a} + \vec{b} = [2+1, 4+3, 6+5]$

$\vec{a} + \vec{b} = [3, 7, 11]$

- vector Subtraction :

$\vec{a} - \vec{b} = [2-1, 4-3, 6-5]$

$\vec{a} - \vec{b} = [1, 1, 1]$

(2.2) scalar multiplication.

- multiplication of vector by single number (scalar).

if: $\vec{V} = [v_1, v_2, v_3, \ldots, v_n]$

and: k = scalar

then: $k \cdot \vec{V} = [k \cdot v_1, k \cdot v_2, k \cdot v_3, \ldots k \cdot v_n]$

eg : $\vec{V} = [2, 4, 6]$

$k = 3$

$3 \cdot [2, 4, 6] = [6, 12, 18]$

Means that :

The length (magnitude) of the vector increases or decreases.

if $k > 1$: vector gets longer

if $0 < k < 1$: vector gets shorter

if $k < 0$: vector reverse direction.

SUKALA
Page
Date

(2.3) Dot product (Inner product)

- The dot product of two vectors is a single number (scalar) you get by multiplying their corresponding elements and adding the results.

if : $\vec{a} = [a_1, a_2, a_3]$, $\vec{b} = [b_1, b_2, b_3]$

$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$

$\therefore \vec{a} \cdot \vec{b} = \sum_{i=1}^{N} a_i b_i$

eg : $\vec{a} = [2, 3, 4]$, $\vec{b} = [5, 6, 7]$

$\vec{a} \cdot \vec{b} = 2 \times 5 + 3 \times 6 + 4 \times 7$

$\vec{a} \cdot \vec{b} = 10 + 18 + 28$

$\vec{a} \cdot \vec{b} = 56$

Usecase :

- Measure how much two vectors align with each other.
- If dot product = 0 than vectors are perpendicular.

(2.4) Cross product (only in 3D)

- The cross product of two 3D vectors is a new vector (not a scalar) that is perpendicular to both original vectors.

if : $\vec{a} = [a_1, a_2, a_3]$, $\vec{b} = [b_1, b_2, b_3]$

$\vec{a} \times \vec{b} = [(a_2 b_3 - a_3 b_2), (a_3 b_1 - a_1 b_3), (a_1 b_2 - a_2 b_1)]$

eg : $\vec{a} = [1, 2, 3]$, $\vec{b} = [4, 5, 6]$

$\vec{a} \times \vec{b} = [(2 \times 6 - 3 \times 5), (3 \times 4 - 1 \times 6), (1 \times 5 - 2 \times 4)]$

$\vec{a} \times \vec{b} = [-3, 6, -3]$

usecase :
- finding a vector that perpendicular to two vectors.
- used in physics & 3D graphics (not much in deep learning)

(2.5) (L1, L2 Norms / magnitude / length)

→ L1 Norms
This is the sum of absolute values of the vector elements
$\|\vec{v}\|_1 = |v_1| + |v_2| + \ldots + |v_n|$
eg: $\vec{v} = [3, -4]$
$\|\vec{v}\|_1 = |3| + |-4| = 7$

→ L2 Norms (Euclidean Norm)
This is the most common norm -
It tells you how long the vector is (like distance from origin)
$\|\vec{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$
eg: $\vec{v} = [3, 4]$
$\|\vec{v}\|_2 = \sqrt{3^2 + 4^2} = \sqrt{9+16} = 5$

(2.6) Unit vector :
A unit vector is a vector with a length of 1, pointing the same direction as the original.

formula : $\hat{v} = \frac{\vec{v}}{\|\vec{v}\|_2}$

eg : $\vec{v} = [3, 4]$

$\|\vec{v}\|_2 = 5$

$\therefore \hat{v} = \left[\frac{3}{5}, \frac{4}{5}\right]$

$\therefore \hat{v} = [0.6, 0.8]$

$\rightarrow \|\hat{v}\| = \sqrt{0.6^2 + 0.8^2} = 1$

(3) Matrix operations :

(3.1) Matrix addition & subtraction

- You can add or subtract two matrices only if they have same dimensions.

eg : $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$

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

$$A-B = \begin{bmatrix} 1-5 & 2-6 \\ 3-7 & 4-8 \end{bmatrix} = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

(3.2) Scalar multiplication

- Multiply every element of matrix by scalar value.

eg $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $K = 3$

$$\therefore K.A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot 3 = \begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix}$$

(3.3) Matrix Multiplication (Dot product):

Rule:

- you can multiply two matrices $A(m \times n)$ and $B(n \times p)$ if:
- The number of columns in A = The number of rows in B.
- Multiply row of A with column of B and then sum the products.

eg: $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}_{(2\times2)}$, $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}_{(2\times2)}$

$$A \cdot B = \begin{bmatrix} (1\times5)+(2\times7) & (1\times6)+(2\times8) \\ (3\times5)+(4\times7) & (3\times6)+(4\times8) \end{bmatrix}$$

$$A \cdot B = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

(3.4) Transpose of matrix:

- flip rows into columns (and vice versa)

eg:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{(2\times3)}$$

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}_{(3\times2)}$$

SUKALA
Page
Date

(3.5) Element wise operation in matrix:

(i) Element wise Addition

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

$$\therefore A+B = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

(ii) Element wise subtraction

$$\therefore A-B = \begin{bmatrix} 1-5 & 2-6 \\ 3-7 & 4-8 \end{bmatrix} = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

(iii) Element wise Multiplication
(Hadamard product)

$$\therefore E = A \circ B$$

$$\therefore A \circ B = \begin{bmatrix} 1\times5 & 2\times6 \\ 3\times7 & 4\times8 \end{bmatrix} = \begin{bmatrix} 5 & 12 \\ 21 & 32 \end{bmatrix}$$

(iv) Element wise division

$$\therefore A/B = \begin{bmatrix} 1/5 & 2/6 \\ 3/7 & 4/8 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.33 \\ 0.42 & 0.5 \end{bmatrix}$$

It's all operations useful in:

- Image processing : Adjust Brightness etc
- ML : Apply activation Funct^n, loss Funct
- Data Analysis
- Scientific computation.

SUKALA
Page
Date

(4) Special matix and properties :

(4.1) - Identity matrix and zero matrix
- main diagonal (From top left → bottom right) are 1, and other elements are 0.

eg : $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- Zero matrix have all elements are 0.

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

Properties : → $A + 0 = A$ & $0 + A = A$
→ $A0 = 0$ & $0A = 0$

(4.2) Diagonal matrix and Orthogonal matrix
- main diagonal have different values and other elements are 0.

eg : $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

- Orthogonal matrix :
properties : $A^T \cdot A = I$ OR $A^T = A^{-1}$
(Transpose X matrix = Identity)

eg : $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $A^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$\therefore A \cdot A^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$

(4.3) Symmetric matrix & Inverse Matrix

- symmetric matrix :

A square matrix where

$\therefore A = A^T$

eg : $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$, $A^T = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$

$\therefore$ This is symmetric matrix

- Inverse matrix :

$\therefore A \cdot A^{-1} = I$ & $A^{-1} \cdot A = I$

How to find Inverse matrix

$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$

Now Interchange 1 & 4 and change sign of 2 & 3.

$\therefore A^{-1} = \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$

Not divide by It's determinant

$\therefore A^{-1} = \frac{1}{-2} \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix}$

(4.4) Rank of matrix & Trace OF matrix

- Rank OF matrix :

The rank of matrix is the number of linearly independent rows or colu

- Independent = "not a copy or combination of other"
- Rank tell us how much actual information is in the matrix.

→ Methods to find Rank of matrix.

(i) Row Echelon Form (REF):

row operations:

(i) $R_2 \to R_2 - 2R_1$

(ii) $R_3 \to R_3 - R_1$

(iii) $R_3 \leftrightarrow R_2$

→ use row operation to convert the matrix into upper triangular 

eg: $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix}$

Now, performing row operation.

(i) $R_2 \to R_2 - 2R_1$

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

(ii) $R_3 \to R_3 - R_1$

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

(iii) $R_3 \leftrightarrow R_2$

$\therefore A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$ (upper triangular Form).

→ Total non-zero rows = 2, $\therefore$ Rank = 2

(ii) using determinants.
(For square sub-matrices)

→ Find largest square submatrix
→ with non-zero determinant.
→ The size of the submatrix
= rank.

eg: $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix}$

→ largest square submatrix is 3×3.
∴ d (3×3) = 0 ∴ Rank ≠ 3
→ Now 2×2 square submatrix
(at least one) with determinant
of non-zero means it's
rank is 2.

∴ $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

∴ d(A) = (1×4) − (2×2)
∴ d(A) = 4 − 4 = 0
→ Now change (2×2) matrix.
∴ $A = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$
∴ d(A) = (2×6) − (4×3) = 0
→ Again change (2×2) matrix
∴ $A = \begin{bmatrix} 4 & 6 \\ 1 & 1 \end{bmatrix}$
∴ d(A) = (4×1) − (6×1) = (−2).
(Non zero) ∴ Rank = 2

- Trace of matrix
It's sum of all diagonal elements of square matrix.

eg : $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$ ∴ Trace = 2+5 = 7.

(5) System of Linear equations :

(5.1) Representing Linear system as Matrix equations

- Let's say we have

$$\begin{cases} 2x + 3y = 8 \\ 4x - y = 2 \end{cases}$$

- coefficients : $\begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix}$
- Variables : $\begin{bmatrix} x \\ y \end{bmatrix}$
- Constants : $\begin{bmatrix} 8 \\ 2 \end{bmatrix}$
- Matrix equation form :

$$A \cdot X = B$$

∵ A = coefficient
∵ X = Variables
∵ B = Constants

So, $\begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$

SUKALA
Page
Date

eg: 3 equations, 3 variables

$$\begin{cases} x + 2y + 3z = 14 \\ 4x + 5y + 6z = 32 \\ 7x + 8y + 9z = 50 \end{cases}$$

Matrix form of 3 equations.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 14 \\ 32 \\ 50 \end{bmatrix}$$

(5.2) Row Echelon Form & Gaussian Elimination

→ What is Row Echelon Form?

- All zero rows (if any) are at the bottom.
- Leading entry from left side of row should be 1. (not for all matrix)

Example matrix in REF.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 4 \\ 0 & 1 & 3 & 7 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

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

↑ This is REF Form

SUKALA
Page
Date

⇒ Gaussian Elimination.

$$\begin{cases} x + 2y + z = 9 \\ 2x + 3y + 3z = 21 \\ 3x + y + 2z = 14 \end{cases}$$

Step 1: Augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 9 \\ 2 & 3 & 3 & 21 \\ 3 & 1 & 2 & 14 \end{array}\right]$$

Step 2: convert (2,1) & (3,1) positioned values into (0).

$\therefore R2 \rightarrow R2 - 2R1$

$\therefore R3 \rightarrow R3 - 3R1$

$$\therefore \left[\begin{array}{ccc|c} 1 & 2 & 1 & 9 \\ 0 & -1 & 1 & 3 \\ 0 & -5 & -1 & -13 \end{array}\right]$$

Step 3: convert (3,2) positioned value into (0).

$\therefore R3 \rightarrow R3 - 5R2$

$$\therefore \left[\begin{array}{ccc|c} 1 & 2 & 1 & 9 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & -6 & 2 \end{array}\right]$$

SUKALA
Page
Date

Step 4: Back substitution.

→ start from the last row.

$\therefore -6z = 2$

$\therefore z = -\frac{1}{3}$

→ now second last row

$\therefore -y + z = 3$

$\therefore -y - \frac{1}{3} = 3$

$\therefore y = -3.33$

→ Now first row

$\therefore x + 2y + z = 9$

$\therefore x + 2(-3.33) + \left(-\frac{1}{3}\right) = 9$

$\therefore x + (-6.66) - \frac{1}{3} = 9$

$\therefore x - 6.66 - \frac{1}{3} = 9$

$\therefore 3x - 19.98 - 1 = 27$

$\therefore 3x = 27 + 19.98 + 1$

$\therefore 3x = 48$

$\therefore x = 48/3$ $\therefore x = 16$

SUKALA
Page
Date

(5.3) Matrix Inverse Method.

suppose, we have equations like

$$\begin{cases} 2x + 3y = 8 \\ 4x + y = 10 \end{cases}$$

$$\therefore \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \end{bmatrix}$$

$$(A) \quad (x) \quad (B)$$

$$\therefore A \cdot X = B$$

formula $\therefore X = A^{-1} \cdot B$

find A^{-1}

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

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

$$\therefore A^{-1} = \frac{1}{-10} \begin{bmatrix} 1 & -3 \\ -4 & 2 \end{bmatrix}$$

$$\therefore A^{-1} = \begin{bmatrix} -0.1 & 0.3 \\ 0.4 & -0.2 \end{bmatrix}$$

$$\rightarrow X = A^{-1} \cdot B = \begin{bmatrix} -0.1 & 0.3 \\ 0.4 & -0.2 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 10 \end{bmatrix}$$

$\therefore x = (-0.1)(8) + (0.3)(10) = 2.2$

$\therefore y = (0.4)(8) + (-0.2)(10) = 1.2$

$\therefore \quad x = 2.2 \ , \ y = 1.2$

(5.4) Cramer's Rule

- Cramer's Rule is a method to solve a system of linear equation using determinants.
- It works only when (square matrix) (2x2) or (3x3)
- Determinant of the coefficient matrix is not zero.

eg:

$$2x + 3y = 8$$
$$4x + y = 10$$

$\therefore AX = B$

$$\therefore A = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \text{ (coefficient matrix)}$$

$$\therefore B = \begin{bmatrix} 8 \\ 10 \end{bmatrix}$$

Step 1: find Determinant of A.

$\therefore \det(A) = 2 \times 1 - 3 \times 4$

$\det(A) = 2 - 12$

$\det(A) = -10.$

Step 2: Replace each column of A with B. to get Ax & Ay

$\therefore Ax = \begin{bmatrix} 8 & 3 \\ 10 & 1 \end{bmatrix}$

$\therefore \det(Ax) = 8 \times 1 - 3 \times 10 = (-22)$

$\therefore Ay = \begin{bmatrix} 2 & 8 \\ 4 & 10 \end{bmatrix}$

$\therefore \det(Ay) = 2 \times 10 - 4 \times 8 = (-12)$

Step 3 : Apply cramer's rule

$\therefore x = \frac{\det(Ax)}{\det(A)} = \frac{-22}{-10} = 2.2$

$\therefore y = \frac{\det(Ay)}{\det(A)} = \frac{-12}{-10} = 1.2$

$\rightarrow \therefore x = 2.2 \ \& \ y = 1.2$

(6) Matrix Inverse and Determinant :

(6.1) Determinant of 2x2 and NxN matrices.

→ Determinant is a special number calculated From Square matrix.

(1) Determinant of 2x2 matrix

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow \det(A) = ad - bc$

eg : $A = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} \rightarrow \det(A) = (3)(5) - (2)(4)$

$\therefore \det(A) = 7$

(2) Determinant of 3×3 or NxN matrix

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

$$\det(A) = 1 \cdot \begin{vmatrix} 1 & 4 \\ 6 & 0 \end{vmatrix} - 2 \cdot \begin{vmatrix} 0 & 4 \\ 5 & 0 \end{vmatrix} + 3 \cdot \begin{vmatrix} 0 & 1 \\ 5 & 6 \end{vmatrix}$$

$\therefore \det(A) = 1 \times (-24) - 2 \times (-20) + 3 \times (-5)$

$\therefore \det(A) = -24 + 40 - 15$

$\therefore \det(A) = 1$

(6.2) Properties of Determinants.

(i) only for square matrices.
(ii) if $\det(A) = 0$, then A is non-invertible
(iii) Multiplying a rows multiplies (by k), determinant by k.
(iv) $\det(AB) = \det(A) \times \det(B)$
(v) $\det(A^T) = \det(A)$

(6.3) Invertibility.

→ A square matrix A is invertible (also called non-singular) if there exists another matrix A^{-1} such that:

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

$\because A^{-1}$ = Inverse of A
$\because I$ = Identity matrix

SUKALA
Page
Date

→ When is a matrix Invertible?
only if :
- It is a square matrix (2×2) (n×n)
- It's determinant is not zero.
∴ det (A) ≠ 0

When det (A) = 0 then
- The matrix is singular
- It has no Inverse
- System of equations may have no or infinite solutions.

→ Diff. between singular vs Non-singular Matrix.

Type	condition	Inverse exists?
singular	det (A) = 0	No
Non-singular	det (A) ≠ 0	Yes

→ Compute Inverse of matrix.

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

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

Eg : $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$

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

$$\therefore A^{-1} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$$

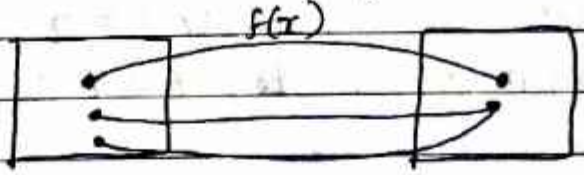
(7) Linear Transformations:

(7.1) Function and vector transformation

Function : A function is a mathematical relationship that uniquely associate elements of one set (Domain) with element of another set (codomain).
In simple, a function maps inputs to output in a specific way.

$$f : x \rightarrow y$$

If x is an element of X, then $f(x)$ is the corrusponding element in Y.



$$f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in R^3 \rightarrow \begin{bmatrix} x+y \\ 6z \end{bmatrix} \in R^2$$

→ function (f) convert 3 dimension to 2 dimension.

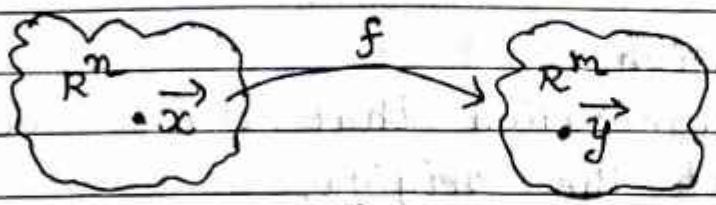
$$f : R^3 \rightarrow R^2$$

∴ $f(x)$ ⇒ Transformation.

SUKALA
Page
Date

Vector transformation :

$f : \vec{x} \rightarrow \vec{y}$



$\therefore f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

eg: $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $f(x, y, z) = (x + y, 2z)$

$\therefore f\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

→ This function $\vec{x} \in \mathbb{R}^3$ transformed to $\vec{y} \in \mathbb{R}^2$.

$\therefore f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$

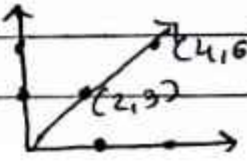
defination of vector transformation :

- vector transformation refer to operations that map vectors from one space (R^n) to another space (R^m), often changing their magnitude, direction or both.

eg: (1) scaling :

scaling is a transformation that change the magnitude of vector but direction remain same.

$v' = 2v = 2\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$

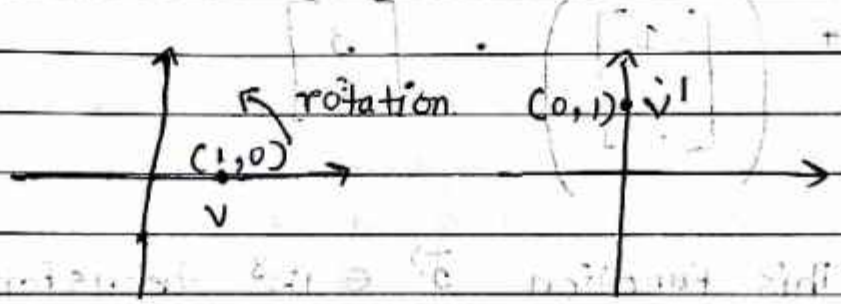


Application : (i) data normalization
(ii) Resize objects
(Image, paint etc).

(2) Rotation :
Transformation that turns vectors around the origine.

$$v = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in R^2$$

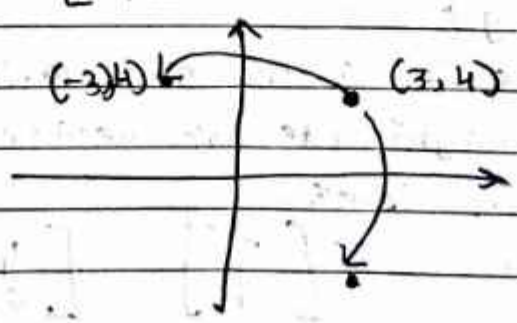
$$v' = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in R^2$$



Application : (i) Image processing ⇒ Rotate Image.
(ii) 3D graphics ⇒ Rotating object.

(3) Reflection :
Transformation that flip vectors over a specifice axis or plane.

$v = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ ⇒ across the y axis



Application :
· Mirroring Images.

SUKALA
Page
Date

(7.2) Linear Transformation

- A Linear transformation is a function between two vector space that preserve the operation of vector addition and scalar multiplication
- This mean that if T is a linear transformat" From a vector space V to a vector space W, then

$T: V \rightarrow W$ ⇒ Linear transformation.

properties.

(i) Additivity $T(u+v) = T(u) + T(v)$

(ii) Homogenity $T(cu) = c \cdot T(u)$

$\because u, v \in V$ and c is scalar value.

eg: Reflection

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2, \quad T(x) = \begin{bmatrix} -x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T(X) = A \cdot X$$

$$T(X) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}$$

(i) check Additivity :

Let $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ be two vector

SUKALA
Page
Date

$$T(u+v) = T(u) + T(v)$$

∴ For LHS :

$$u+v = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}$$

Now $T(u+v) = A \cdot (u+v)$

$$= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}$$

$$LHS = T(u+v) = \begin{bmatrix} -(u_1+v_1) \\ u_2+v_2 \end{bmatrix}$$

∴ For RHS :

$$T(u) = A \cdot u = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -u_1 \\ u_2 \end{bmatrix}$$

$$T(v) = A \cdot v = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -v_1 \\ v_2 \end{bmatrix}$$

$$RHS = T(u) + T(v)$$

$$= \begin{bmatrix} -u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} -v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -(u_1+v_1) \\ u_2+v_2 \end{bmatrix}$$

∴ LHS = RHS

First condition satisfied.

SUKALA
Page
Date

(ii) check Homogenity :

Let $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \in \mathbb{R}^2$ & c is a scalar

$$T(cu) = c \cdot T(u)$$

For LHS :

$$cu = c \cdot \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}$$

$$\therefore T(cu) = A \cdot (cu)$$

$$= \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix} = \begin{bmatrix} -cu_1 \\ cu_2 \end{bmatrix}$$

For RHS :

$$c \cdot T(u) = c \cdot (A \cdot u)$$

$$= c \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = c \begin{bmatrix} -u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -cu_1 \\ cu_2 \end{bmatrix}$$

$\therefore$ LHS = RHS

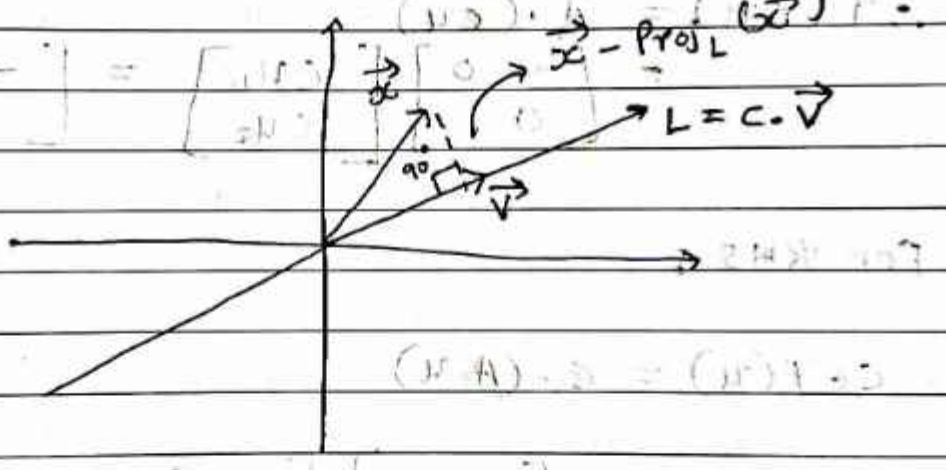
$\therefore$ second condition also satisfied.

SUKALA
Page
Date

(7.3) Linear Transformation in Data Science.

(i) Dimensionality Reduction (PCA)
(ii) Feature engineering
(iii) Data Preprocessing
Normalization and standardization
(iv) Neural Networks (Forward propogation, Activation Function)
(v) Image and signal processing.
(vi) Optimization.

(7.4) Projections :



$Proj_L(\vec{x})$ ⇒ Projection of $\vec{x}$ on the Line L.

→ Note: Projected vector is perpendicular to $\vec{x} - Proj_L(\vec{x})$.

Formula : Projection of $\vec{x}$ on ~~line~~ vector V.

$$Proj_V(\vec{x}) = c \cdot \vec{V}$$
$$= \left(\frac{\vec{x} \cdot \vec{V}}{\vec{V} \cdot \vec{V}}\right) \cdot \vec{V}$$

OR simple Formate :

Projection of $\vec{a}$ onto $\vec{b}$ is:

$$\text{Proj}_{\vec{b}}\,\vec{a} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{b}\|^2}\cdot\vec{b}$$

eg: $\vec{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Projection of $\vec{a}$ on $\vec{b}$.

$\therefore \vec{b} = \{ c \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$$\text{Proj}_{\vec{b}}\,\vec{a} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{b}\|^2}\cdot\vec{b}$$

$$\therefore = \frac{\begin{bmatrix} 2 \\ 3 \end{bmatrix}\cdot\begin{bmatrix} 2 \\ 1 \end{bmatrix}}{5}\cdot\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \end{bmatrix}\cdot\frac{7}{5}$$

$$= \begin{bmatrix} \frac{14}{5} \\ \frac{7}{5} \end{bmatrix}$$

$\therefore \text{Proj}_{\vec{b}}\,\vec{a} = \begin{bmatrix} 2.8 \\ 1.4 \end{bmatrix}$

(8) Inverse Function or Transformation

(8.1) Inverse of Function

- Inverse of Function is a function that "Reverse" the effect of original Function.

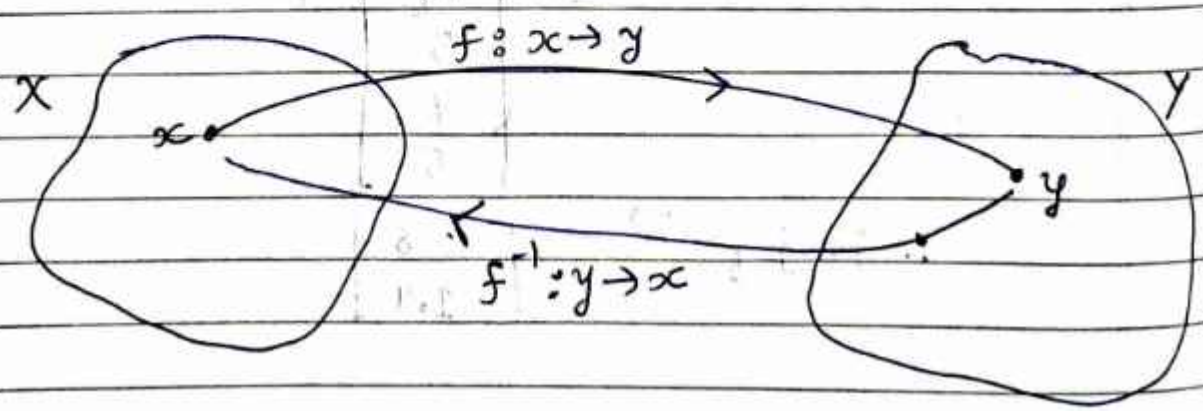
- IF you have a function f that maps an element x from set X to an element y in a set Y, the inverse function f^{-1} map y back to x.

Function $f : X \to Y$
Inverse function $f^{-1} : Y \to X$

For ~~any~~ every, $y \in Y$, there is a unique $x \in X$ such that $f(x) = y$

- The Inverse Function f^{-1} satisfies the following condition.

1) For all $x \in X : f(f^{-1}(y)) = y$
2) for all $y \in Y : f^{-1}(f(x)) = x$



SUKALA
Page
Date

⇒ Identity function :

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

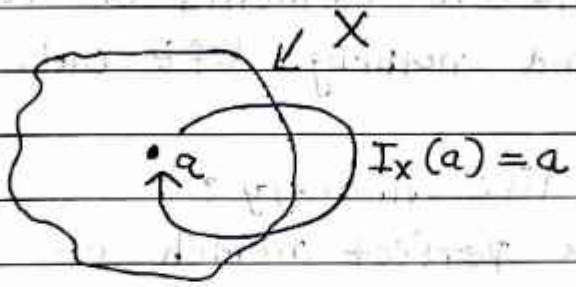
$$IV = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$I_X = X \to X$ and $I_X(a) = a$
where $a \in X$

→ For a set X, the Identity Function (I_X) is defined as :

$$I_X(a) = a \quad \text{For all } a \in X.$$

→ I_X is the identity function on the set X and it maps every element x in X to itself.



→ Properties of Identity Function :

(i) Preservation : Does not alter any element. If x is the domain, then the image of x under the identity fn is x.

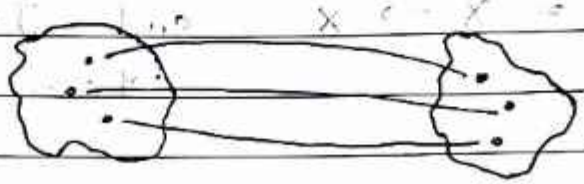
(ii) Identity function is a linear transformation
- $I(u+v) = I(u) + I(v)$
- $I(cu) = c \cdot I(u)$

SUKALA
Page
Date

→ Existence and Uniqueness:

- A function f has an inverse if and only if it is bijective.

(i) Injective (one to one): Different element in the domain map to different element in codomain.



(ii) Surjective (onto): Every element in codomain is the image of atleast one element in the domain.

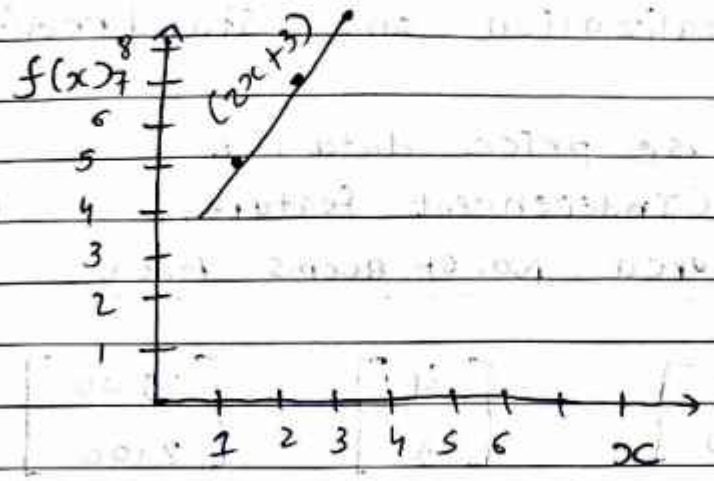
- Bijective function is a perfect pairing between elements of two sets - no duplicates, and nothing left out.

Real life Analogy:
Think perfect match of dance partners:
- Each person has exactly one partner (no two people share same partner → Injective)
- Everyone has a partner (nobody is left → surjective)

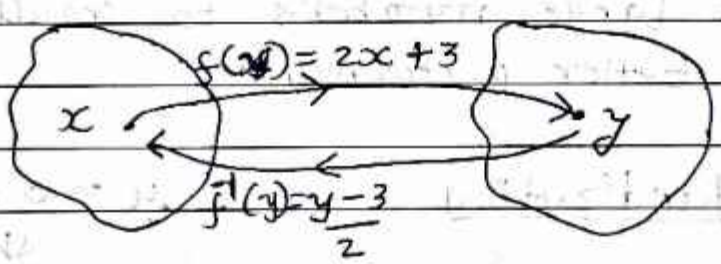
SUKALA
Page
Date

Eg: Linear Function $f(x) = 2x+3$

$x = [1, 2, 3]$



$\therefore y = [5, 7, 9]$



find the Inverse

$y = 2x+3$ for x :

$y = 2x+3$

$y-3 = 2x$

$\therefore x = \frac{y-3}{2}$

The Inverse Function $\boxed{f^{-1}(y) = \frac{y-3}{2}}$

Verification:

(i) $f(f^{-1}(y)) = f\left(\frac{y-3}{2}\right) = 2\left(\frac{y-3}{3}\right)+3 = y-3+3 = y$

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

(8.2) Application of Inverse Function:

(i) Normalization and Standardization:

eg: House price dataset

(Independent features)			(dependent)
carpet area	No. of Rooms	Area	Price
1500	4	1800	50 Lakhs
1800	3	2100	60 Lakhs

for the model training, we scaling down the large numbers to small numbers For better performance.

Standardization $(\mu = 0 \;\&\; \sigma = 1)$

$\Downarrow$

Standard Normal distrib.

No. of Rooms.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \in R \Rightarrow I = \frac{x_i - \mu}{\sigma}$$

No. of Rooms

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \in R \xrightarrow{f(x)} \begin{bmatrix} -1.5/2 \\ -0.5/2 \\ 0.5/2 \\ 1.5/2 \end{bmatrix} \quad \text{No. of Rooms}$$

$$\xleftarrow{f^{-1}(x)}$$

$\therefore (\mu = 2.5 \;\&\; \sigma = 2)$:

SUKALA
Page
Date

original Transformation : $z = \frac{x_i - \mu}{\sigma}$

Inverse Transformation : $x_i = z\sigma + \mu$.

→ use case:
After training a machine learning model on standardized data, the prediction after rescaled back to the original scale to interpret the result in meaningful way.

Normalization

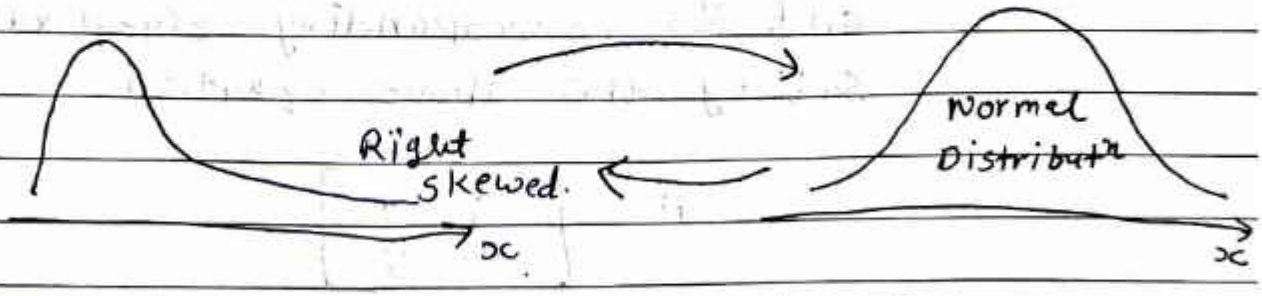
feature scaling with min max Normalization.

Original transformation : $z = \frac{x - \min(x)}{\max(x) - \min(x)}$

Inverse transformation : $x = z(\max(x) - \min(x)) + \min(x)$

(ii) Distribution of data :

- Logarithmic distribution.



Original Transformation $y = \log(x)$
Invers Transformation $x = e^y$

SUKALA
Page
Date

Visit ↓

(9) Eigen Value and Eigen Vector. (shad.io/MatVis/)

- Eigenvector of a matrix is a special vector that doesn't change direction when the matrix is applied to it. It only gets stretched or squished by some factor (called eigen value).

- Eigen value (λ) : A scalar that indicates how much an eigen vector is stretched or compressed during Linear transformation.

- Eigen Vector (v) : A non zero vector that only changes in scale (not direction), when a linear transformation is applied.

$$A \cdot V = \lambda \cdot V$$

$\because$ A = squar matrix
$\because$ λ = Eigen value
$\because$ V = Eigen Vector

→ For a square metrix A, an eigen vector and its corresponding eigen value λ satisfy the above equation

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

SUKALA
Page
Date

(i) Find Eigen values.

$$\det(A - \lambda I) = 0$$

Now, $A - \lambda I = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$

$$A - \lambda I = \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix}$$

$$\det(A - \lambda I) = (4-\lambda)(3-\lambda) - (2)(1)$$

$$= 12 - 4\lambda - 3\lambda + \lambda^2 - 2$$

$$\det(A - \lambda I) = \lambda^2 - 7\lambda + 10$$

Solve this equation.

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

$$\lambda = \frac{7 \pm \sqrt{49 - 40}}{2}$$

$$\lambda = \frac{7 \pm \sqrt{9}}{2}$$

$$\lambda_1 = \frac{7+3}{2}, \quad \lambda_2 = \frac{7-3}{2}$$

$$\therefore \lambda_1 = 5, \quad \lambda_2 = 2.$$

This is Eigen value, $\lambda_1 = 5$ or $\lambda_2 = 2$

(ii) Find Eigen vector.

$$(A - \lambda I)X = 0$$

For $\lambda_1 = 5$:

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

$$A - 5I = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

Now, $\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 0$

$$-x + y = 0$$

$$\therefore x = y$$

For $\lambda_2 = 2$

$$A - 2I = \begin{bmatrix} 4-2 & 1 \\ 2 & 3-2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

Now, $\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 0$

$$2x + y = 0$$

$$\therefore y = -2x$$

$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$, $\lambda_1 = 5$ & $\lambda_2 = 2$

For $\lambda_1 = 5$, $V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

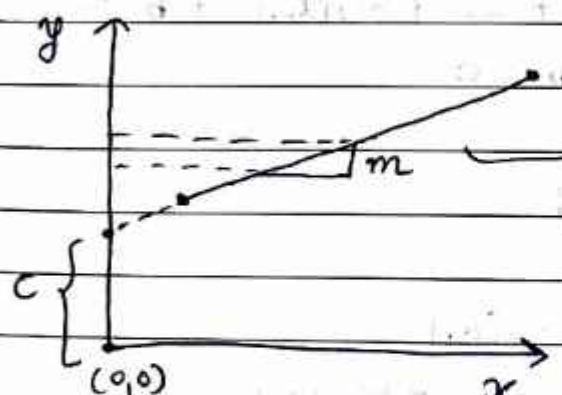
For $\lambda_2 = 2$, $V_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

Application :

(i) PCA (Dimentionality Reduction)
(ii) Latent Semantic Analysis (LSA) in NLP

(10) Equation of line ; plane and Hyperplane

- Line : (2D)



$y = mx + c$

$y = \beta_0 + \beta_1 x$

[m = slope, c = intercept]

or

$ax + by + c = 0$

$w_1 x_1 + w_2 x_2 + b = 0$

$\therefore w^T \cdot x + b = 0$

→ Linear Regression model prediction salary based on Experience.

Salary = 5000 · Experience + 20000

$y = m \cdot x + c$

SUKALA
Page
Date

- Plane (3D):

$$ax + by + cz = d$$
or $$w_1x_1 + w_2x_2 + w_3x_3 + b = 0$$

$$\therefore w^T \cdot x + b = 0$$

(a, b, c) is normal vector perpendicular to the plane
d is a scalar.

eg: multiple linear regression with 2 Features.

$$\text{Price} = 2000 \cdot \text{size} + 1500 \cdot \text{Rooms} + 10000$$
$$y = m_1 \cdot x_1 + m_2 \cdot x_2 + c$$

- Hyperplane (n-D):

$$w_1x_1 + w_2x_2 + \dots + w_nx_n + b = 0$$
or $$w^T \cdot x + b = 0$$

Application in DS:

(i) simple Regression
(ii) Multivariable Regression
(iii) Logistic Regression
(iv) SVM
(v) PCA
(vi) clustering (K-mean).

x x

Ch: 2

Statistics

SUKALA
Page
Date

Index :

1. Introduction to statistics
   - 1.1 Introduction and types of statistics
   - 1.2 Population and sample data
   - 1.3 Types of Sampling
   - 1.4 Types of data and scale of measurement
2. Descriptive statistics
   - 2.1 Measure of central tendency
   - 2.2 Measure of Dispersion
   - 2.3 Percentile and Quartile.
   - 2.4 5 Number summary
   - 2.5 Histogram and skewness
   - 2.6 Correlation and covariance
3. Inferential Statistics and hypothesis testing
   - 3.1 Hypothesis testing and it's mechanism
   - 3.2 P value and hypothesis testing
   - 3.3 z test and hypothesis testing
   - 3.4 t test and hypothesis testing
   - 3.5 Type 1 & ptype 2 error
   - 3.6 Baye's Theorem
   - 3.7 Chi square test
   - 3.8 Annova test

SUKALA
Page
Date

(1) Introduction to Statistics :

(1.1) Introduction and Types of Statistics

- Statistics is the science of collecting, organizing and analyzing data.

Application :

(i) Data Exploration and Summarize
(ii) Model Building and Validation
(iii) Statistical Analysis
(iv) Hypothesis testing
(v) Optimization and efficiency.
(vi) Reporting.

- Types of statistics

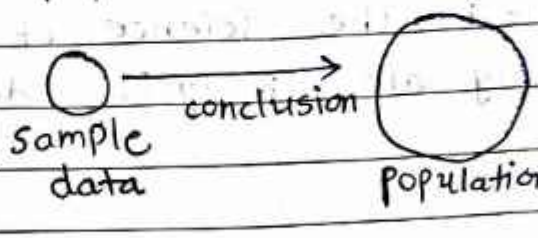
(i) Descriptive statistics :

It involves a methods for summarizing and organizing data to make it understandable.

- Measure of central tendency
- Measure of Dispresion
- Data distribution
  i) Histogram
  ii) Box plot
  iii) Pie chart
- Summary statistics
  (Five number summary)

(ii) Infrential Statistics

It involves methods for making prediction about population based on sample data.



- Hypothesis testing
- P. Value
- Confidence interval
- Analysis test
  - z test
  - T test
  - Anova / F test
  - chi square

(1.2) Population and Sample data

(i) Population data:

- A population is a entire set of individuals of interest in perticula study.

characteristics:

i) complete set: contain all the observatio of interest

ii) Parameter: A numerical value summari the entire population
  - Population mean (μ)
  - Population variance (σ^2)

SUKALA
Page
Date

Example :

(i) Population in school study
- all students enrolled in a school.
- determine Avg. height of students.

(ii) Population in market study.
- all consumers in a city
- To understand purchasing behavior of all costomers.

(ii) Sample data :
- A semple is a subset of the population that is used to represent the entire group.

characteristics :
i) subset : Represent a portion of the population.
ii) statistic : A numerical value summarizing the sample data.
[sample mean, sample variance]
iii) Random sampling : Sample should randomly selected to avoid bias

Example :

(i) Sample in school study.
- A group of 50 students from school
- Estimate the Avg. height of students in school.

SUKALA
Page
Date

(1.3) Types of Sampling :

(1) Probability sampling
(2) Non probability sampling.

→ (1) Probability sampling :

(a) simple random sampling :
- Every member has an equal chance of being selected.
eg: selecting people randomly

(b) Systematic sampling :
- Select every n^{th} member From population.
eg: selecting every even number student 2^{nd}, 4^{th}, 6^{th} etc.

(c) Stratified sampling :
- Divide the population into strata (group) based on specific characteristics and than randomly sampling From each strata.
eg: Divide employees by department than randomly select a specific number of employee from each department.

(d) cluster sampling :
- Divide the population in clusters than select random clusters, Its mean all the members of selected clusters are sample.

→ (2) Non probability sampling :

(a) Convenience sampling :
- selecting individual who are easier to reach.
eg : surveying people at mall.

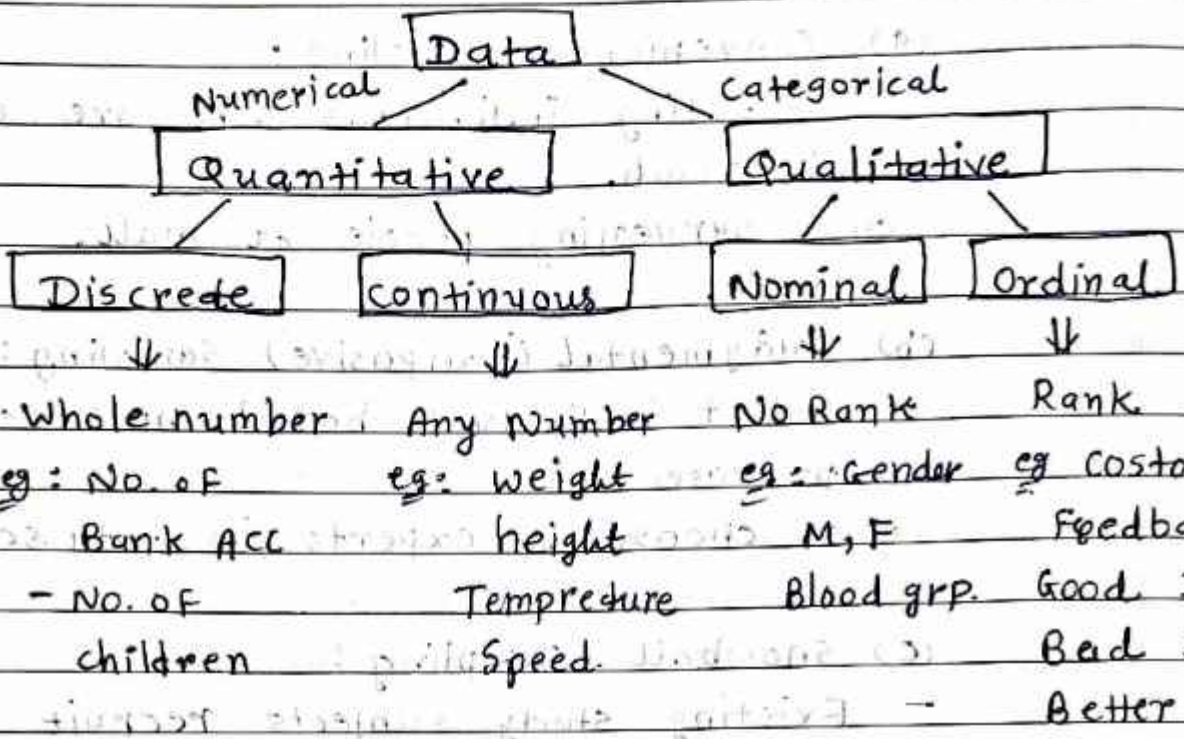
(b) Judgmental (purposive) sampling :
- Select individual based on researcher's purpose.
eg : choose a experts in data science.

(c) Snowball sampling :
- Existing study subjects recruit Future subject From their network.
eg : Internal employee select candidate From their network for job possition.

(d) Quota sampling :
- Divide population into categories like (age, gender, income etc) and then select a fixed number (quota) of subjects From each category.
eg : Total 100 seats For vacancy and 20 seats quota For disables.

SUKALA
Page
Date

(1.4) Types of Data and Scale of measurment



- Scale of measurement of data.
  - (i) Nominal scale
  - (ii) Ordinal scale
  - (iii) Interval
  - (iv) Ratio

- (i) Nominal scale
  - This scale classifies data into distinct categories that do not have an interinsic order
  - Qualitative / categorical data.

Characteristics

i) Data is categorized based on lables, names
ii) categories are mutually exclusive
iii) No logical order among catagories (Rank)

eg: Types of cuisines
- Italian
- Chinese
- Mexican

(ii) ordinal scale
- This scale classify data into categories that can be ranked or ordered.

characteristics
(i) Data is categorized and ranked in specific order
ii) The interval between ranks are not necesarily equal

eg: Education level

High school	0
Bachlor	1
master	2
Doctorate	3

, Costomer Feedback

Poor	0
good	1
Excellent	2

(iii) Interval scale
- Numerical data with meaningful differences.
- No true zero value.

eg: Tempreture in celsius
- 20°c is 10° more than 10°c.
- 0°c is not mean no tempreture

→ you can add/subtract, but ratios don't make sense.
(eg: 20°c is not "twice as hot" as 10°c)

(iv) Ratio scale :

- Like interval, but hase true zero value
- Allow all maths operation like add, subtract, multiply, divide.

eg : Age, height, weight, income, distance
→ you can say 20kg is twice heavy as 10 kg.

(2) Descriptive Statistics :

(2.1) Measure of central tendency.

- measure of central tendency are statistical metrics that describe the center point or typical value of a dataset. They provide a single value that summarize a set of data by identifying the central position within that dataset.

(i) mean :

- mean is the sum of all values divide by the number of values.

Population mean (μ) — population (N)

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N}$$

Sample mean ($\bar{x}$) — sample (n)

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

eg : $x = \{5, 8, 12, 15, 20\}$

$N = 5$

$\therefore \mu = \frac{5+8+12+15+20}{5} = \frac{60}{5} = 12$

- characteristics
  - Affected by exteme outliers.
  - used for interval and ratio data

(ii) median :
- The median is the middle value in a dataset when the values are arranged in ascending or descending order.

$X = \{1, 2, 3, 4, 5\}$
median = 3

$X = \{1, 2, 3, 4, 5, 100\}$
$\text{median} = \frac{3+4}{2} = 3.5$

- characteristics:
  - Not affected by exterme outliers.
  - used for ordinal, interval & Ratio data.

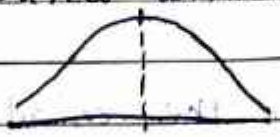
(iii) mode :
- The most Frequent value in a dataset.

eg :- Dataset : $\{2, 4, 4, 6, 7, 7, 7, 9\}$
$\therefore$ Mode = 7
- Dataset : $\{3, 5, 5, 6, 6, 8\}$
$\therefore$ Mode = 5, 6 (Bimodal)

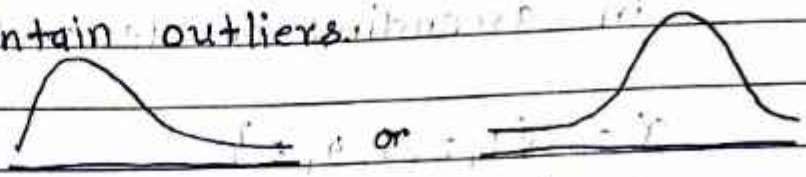
- characteristics:
  - Not affected by exterme outlier.
  - Used for Nominal, ordinal, interval & Ration

- choosing the Appropriate measure

i) mean : Best use when data is symmetrically distributed without outliers.



ii) median : Best use when data is skewed or contain outliers.



iii) mode : Best use for categorical data to identify the most comon category.

- Real world Application. (EDA & Feature engi...)

Age	weight	Salary	Gender	Degree
24	70	40 K	M	BE
25	80	70 K	F	-
27	95	45 K	F	-
24	-	50 K	M	PHD
32	-	60 K	-	BE
-	60	-	-	Master
-	65	55 K	-	BSc.
40	72	-	M	BE

→ There are some missing values in dataset, we use mean or median for age and weight, mode for Gender & degree.

(2.2) Measure of Dispersion

- Measure of dispersion describe the spread or variability of a dataset. They indicate how much the values in a dataset is differ from the central tendency.

- common measure of dispersion
  (i) Range
  (ii) Variance
  (iii) Standard deviation.
  (iv) Interquartile Range.

- (i) Range
  - Range is a difference between the maximum and minimum value in dataset.
  - Rang = max. value − min. value

eg: Ages = { 14, 13, 10, 20, 25, 75, 15 }
Range = 75 − 10 = 65

- characteristics :
  - simple to calculate
  - sensitive to outliers
  - Rough measure of dispersion

- (ii) Variance
  - variance measures the average squared deviation of each value from the mean.

SUKALA
Page
Date

Population variance

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

$x_i \Rightarrow$ Data points
$\mu \Rightarrow$ population mean
$N \Rightarrow$ population size

sample variance

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

$x_i \Rightarrow$ Data point
$\bar{x} \Rightarrow$ sample mean
$n \Rightarrow$ sample size

eg: size of flower petals

$\{5, 8, 12, 15, 20\}$

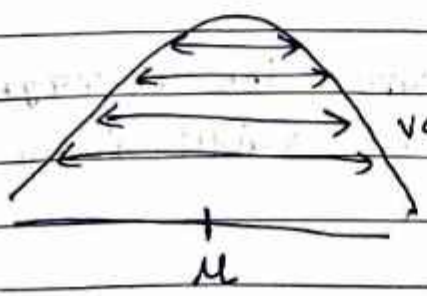
$N = 5$

$\mu = \frac{5+8+12+15+20}{5} = 12$

$\sigma^2 = \frac{(5-12)^2 + (8-12)^2 + (12-12)^2 + (15-12)^2 + (20-12)^2}{5}$

$\sigma^2 = 27.6$

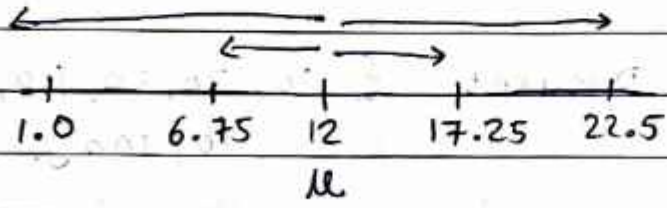
- characteristics.
  - provide a precise measure of variability
  - units are squared of the original data unit.
  - more sensitive to outliers than the Range



- (iii) standard Deviation :
  - The standard deviation is the square root of the variance.

$$\sigma = \sqrt{27.6}$$

$$\therefore \sigma = 5.25$$



- characteristics
  - provide a clear measure of spread in the same unit as the data.
  - Sensitive to outliers.

Big Question ?
- Why sample variance is divided by N-1.

(2.3) Percentile and Quartile.

(i) Percentile :

- A percentile tells you the relative standing of a value in a dataset.

eg : in a class of 100 students, if you scored at 90th posi percentile, it means you scored better than 90 students.

SUKALA
Page
Date

(ii) Quartile :

- Quartile is a specific percentile that divide your data into Four equal parts.

Q1 — 1st quartile (25-1-)
Q2 — 2nd quartile (50-1-).
Q3 — 3rd quartile (75-1-).

eg : Dataset = { 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 }.

$$\text{Rank} = \frac{P}{100} \times (n+1)$$

$$\therefore \text{Rank} = \frac{25}{100} \times (10+1) = 0.25 \times 11$$
$$= 2.75^{th} \text{ position}$$

$\therefore$ Q1 = 30

$$Q2 = \frac{50}{100} \times 11 = 0.5 \times 11 \Rightarrow 5.5$$

$\therefore$ Q2 = 55

$$Q3 = \frac{75}{100} \times 11 = 0.75 \times 11 = 8.25$$

$\therefore$ Q3 = 80

SUKALA
Page
Date

(2.4) Five number summary

- Five number summary is a quick overview of a dataset's distribution.

- components of 5 number summary:
  1. minimum - the smallest data value.
  2. Q1 (First Quartile) - 25th percentile
  3. Median (Q2) - 50th percentile
  4. Q3 (Third Quartile) - 75th Percentile
  5. maximum - the largest data value.

eg: {1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, 27}
N = 19.

Lower fence = Q1 − 1.5 (IQR).
Higher Fence = Q3 + 1.5 (IQR)

$\therefore Q1 = \frac{\text{Percentile}}{1000} \times (n+1) = \frac{25}{100} \times (20) = 5^{th}$ position
$= 3$ //.

$\therefore Q3 = \frac{75}{100} \times 20 = 15^{th}$ position $= 7$

IQR = Q3 − Q1 = 7 − 3 = 4

Lower Fence = Q1 − 1.5(IQR)
= 3 − 1.5(4)
= 3 − 6 $\boxed{= -3}$

Higher Fence = Q3 + 1.5(IQR)
= 7 + 1.5(4)
$\boxed{= 13}$

→ 27 will be removed because it is higer than Higer Fence.

minimum = 1
1st Quartile = 3
Median = 5
3rd Quartile = 7
Maximum = 9

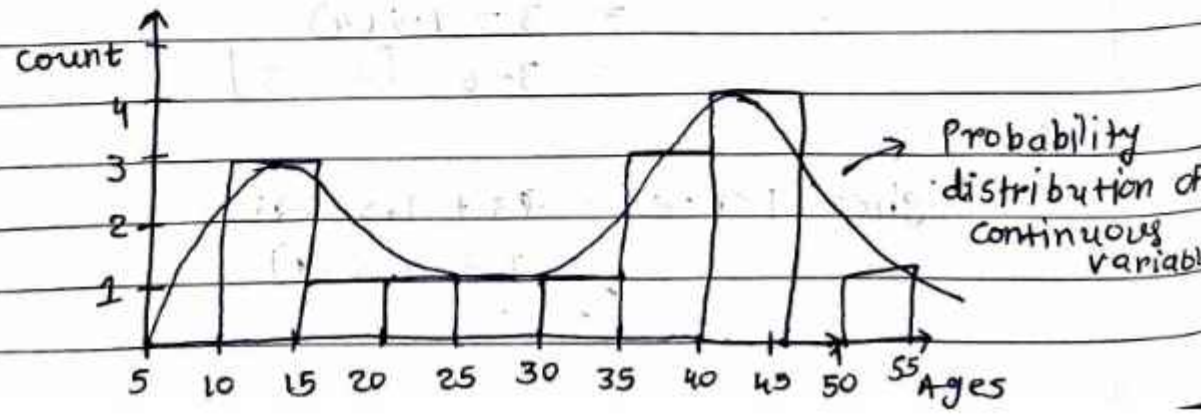
(2.5) Histogram and skewness

- Histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable and is used to visualize the shape, central tendency and variability of dataset.

Ages = { 11, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50 }

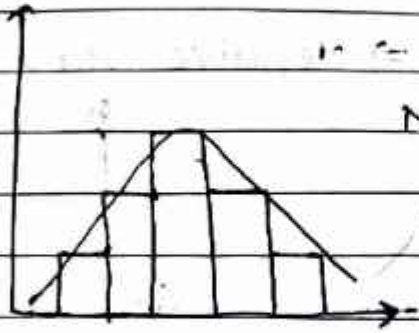
→ No. of bins = 10
→ Range = [0 - 50]
→ bin size = $\frac{50}{10} = 5$

Bins → [0-5, 5-10, 10-15, 15-20, ---, 45-50]



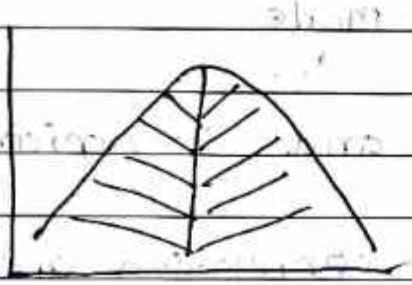
SUKALA
Page
Date

- Skewness

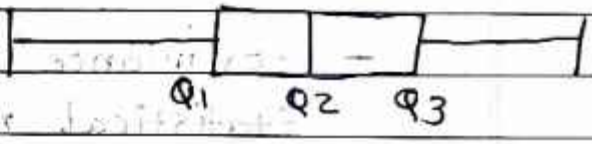


Normal / Gaussian distribution

symmetrical distribution
(No skewness)

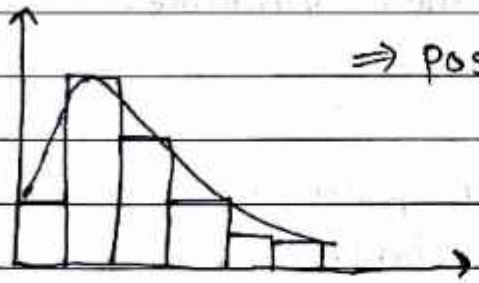


The mean, median & mode are perfectly at the center



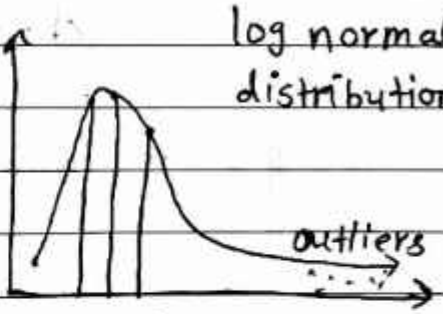
Mean = Median = Mode $\quad Q_3 - Q_2 \approx Q_2 - Q_1$

- Right skewed



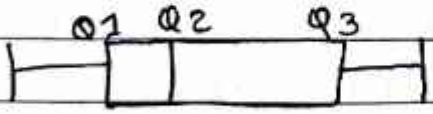
⇒ positive skewed

log normal distribution



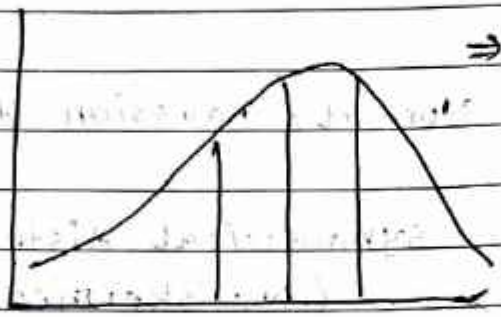
mean ≥ median ≥ mode

Box plot:

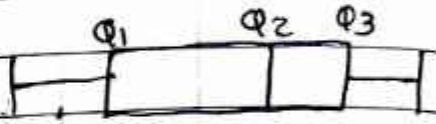


$Q_3 - Q_2 \geq Q_2 - Q_1$

- Left skewed.



⇒ Negative skewed



$Q_2 - Q_1 \geq Q_3 - Q_2$

mean ≤ median ≤ mode

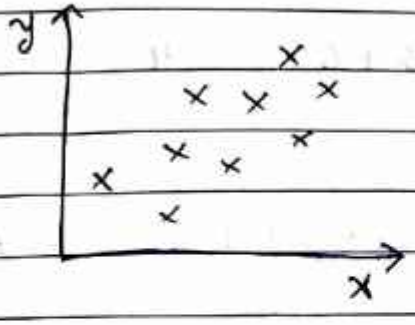
(2.6) Correlation and Covariance

- covariance and correlation are two statistical measures used to determine the relationship between two variables. Both are used to understand how changes in one variable are associated with changes in another variable.

∴ covariance :

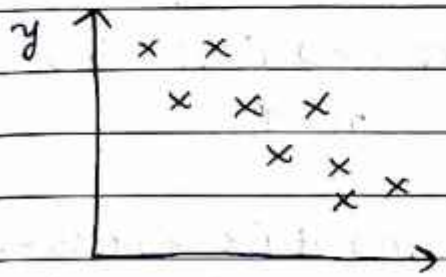
covariance is a measure of how much two random variables change together. If the variable tend to increase and decrease together, the covariance is positive. If one tends to increase and other decrease then the covariance is negative.

size of house ↑↓	price ↑↓	
1200	45 lakhs	x↑ y↑
1300	50 lakhs	x↓ y↑
1500	75 lakhs	x↑ y↓
		x↓ y↓



$x\uparrow \; y\uparrow$
$x\downarrow \; y\downarrow$

⇒ Positive (+ve) covariance
⇒ +ve value



$x\downarrow \; y\uparrow$
$x\uparrow \; y\downarrow$

⇒ Negative (-ve) covariance
⇒ -ve value.

Formula :

$$cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

x_i → datapoint of random variable x
$\bar{x}$ → sample mean of n
y_i → Datapoints of random variable y
$\bar{y}$ → sample mean of y

eg : Students

Hour of studied (x)	Exam score (Y)	
2	50	
3	60	$x\uparrow \; y\uparrow$
4	70	$x\downarrow \; y\downarrow$
5	80	
6	90	

(i) $\bar{x} = \frac{2+3+4+5+6}{5} = 4$

(ii) $\bar{y} = \frac{50+60+70+80+90}{5} = 70$

$$cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$cov(x,y) = \frac{(2-4)(50-70) + (3-4)(60-70) + (4-4)(70-70) + (5-4)(80-70) + (6-4)(90-70)}{4}$$

$\therefore cov(x,y) = \underline{20}$

positive covariance indicates the no. of hours studied increased the exam score also.

- Advantage:
  - Quantify the relationship between x & y.

- Disadvantage:
  - covariance does not have a specific unit value

$cov(x,y) \Rightarrow -\infty$ to $+\infty$

SUKALA
Page
Date

- correlation :

→ Pearson correlation coefficient
→ Spearman Rank correlation

(i) pearson correlation coefficient. ⇒ [-1 to 1]

- Linear relationship between two continuos variable
- It checks how well a streight line Fits the data.

Formula : $r = \frac{cov(x,y)}{\sigma x \cdot \sigma y}$

Standard deviation of x & y

- The more the value toward +1 the more +ve correlated x & y
- The more the value toward -1 the more -ve correlated it is (x,y).
- IF r = 1 then all variables perfectly Fit to streight line.

(ii) Spearman Rank correlation

- Good when data is not linear or ordinal
- use rank instead of raw values.

SUKALA
Page
Date

Formula: $P = \frac{cov(rank(x), rank(y))}{\sigma(rank(x)) \cdot \sigma(rank(y))}$

eg:

x	y	Rank(x)	Rank(y)
1	2	2	1
3	4	3	2
5	6	4	3
7	8	5	5
0	7	1	4

- Realworld Application in datascience.

(i) Feature selection.
- Remove irrelevant Features (correlation)

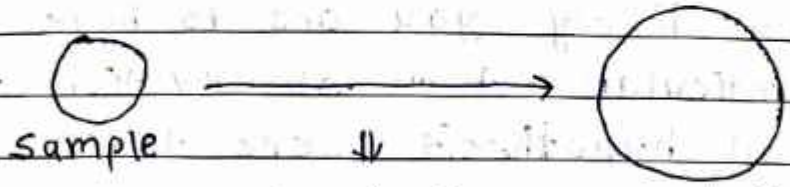
(ii) Linear regression.
- Detecting multicollinearity
- IF two features are highly co-related then remove one feature of combine them using PCA. (correlation)

(iii) PCA.
- reduce the number of features while retaining most of the variance in the data. It relies on the covariance matrix.

SUKALA
Page
Date

(3) Inferential Statistics and hypothesis testing.

(3.1) Hypothesis testing and it's mechanism



Hypothesis testing

- Hypothesis testing mechanism :

Person → crime
court

(1) Null hypothesis (H_0)
- person is not guilty
- The assumption you are beginning with

(2) Alternate hypothesis (H_1)
- The person is guilty
- opposite of null hypothesis

(3) Experiment → Statistical Analysis
Different type of test.

(4) Accept the null hypothesis or reject the null hypothesis.

SUKALA
Page
Date

(3.2) P value and hypothesis testing

- The P value is a number, calculated From a statistical test, that describes how likely you are to have found a particular set of observation if the null hypothesis were true.

- P values are used in hypothesis testing to help decise whether to reject the null hypothesis.

- Simple example:

A company claims their battery lasts 10 hours on average. you test 30 batteries and find the average life is 9.5 hours.

Question ?
is this difference (10 vs 9.5) due to random chance or it claim not true ?

Ho (Null Hypothesis): $\mu = 10$ hours.
H1 (Alt Hypothesis): $\mu \neq 10$ hours.

you calculate P-value = 0.03

P-value	Interpretation
$P < 0.01$	Strong evidence (Reject Ho)
$0.01 < P < 0.05$	Moderat evidence (Reject Ho)
$P > 0.05$	Weak evidence (Accept Ho)

In our example $p = 0.03$

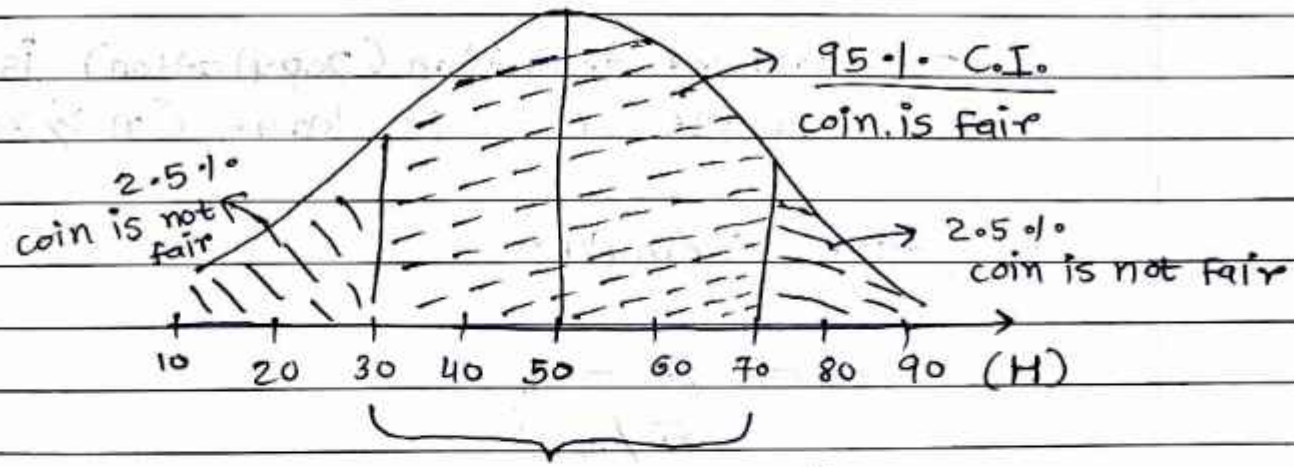
therefore Reject H_0

conclusion : Average battery life is likely not 10 hours.

- Example :

coin is fair or not, we want to test.
we toss 100 times

(i) Null Hypothesis (H_0) = coin is fair
(ii) Alternative hypothesis (H_1) = coin is not fair
(iii) Experiment : 100 times



(iv) Significance value : $\alpha = 0.05$

$C.I. = 1 - 0.05$

$\therefore C.I = 0.95$

(iv) conclusion : $p < \alpha$ (significance value) then Reject the Null hypothesis

else failed to reject null hyp.

SUKALA
Page
Date

(3.3) z test and hypothesis testing.

(i) z test } Average ⇒ z table
(ii) t test } t table
(iii) Chi square ⇒ categorical data
(iv) Anova ⇒ variance

- **z test :**

z test is a statistical test used to determine whether there is a significant difference between:
- A sample mean and population mean
- or a two sample means.

- Standard deviation (population) is known & sample size is large ($n \geq 30$).

- **z test Formula:**

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size.

SUKALA
Page
Date

Problem : company claim that their energy drink increase alertness and people stay focused for 8 hours on average.
you think it's less. you test it with a sample of 36 peoples and get:
- sample mean = 7.5 hours.
- population std = 1.2
- population mean = 8 hours.
- sample size = 36
you want to check if people actually stay focused less than 8 hours. (left tailed test)

→ (i) Null hypothesis (H_0) = $\mu = 8$
(ii) Alt hypothesis (H_1) = $\mu < 8$

(iii) calculate z score:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{7.5 - 8}{1.2/\sqrt{36}} = \frac{-0.5}{0.2}$$

$\therefore z = -2.5$

(iv) Now find p value from z table.
(look at -2.5 & 0.00 value in z table)

$\therefore p(z < -2.5) = 0.0062$

$\therefore$ p value = 0.0062

(v) compare p value with significance level (α)
let take $\alpha = 0.05$
$\therefore 0.0062 < 0.05 \Rightarrow$ We Reject H_0.
$(p < \alpha) \Rightarrow$ People stay focused less than 8 hr.

SUKALA
Page
Date

→ Two types of hypothesis tests

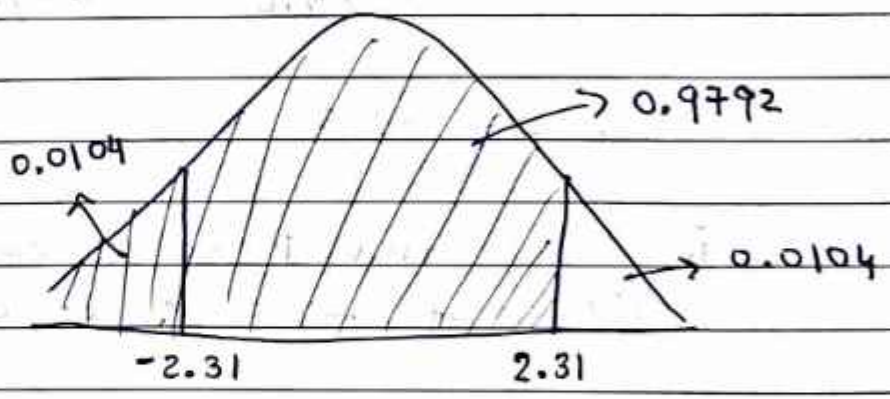
(i) Left tailed or Right tailed test (one tailed)
- we test, if the mean is less than or greater than the population mean.
- use only one side of distribution
- No need to multiply P value by 2.
- left tailed : use $P(z < z)$
- Right tailed : use $1 - P(z < z)$

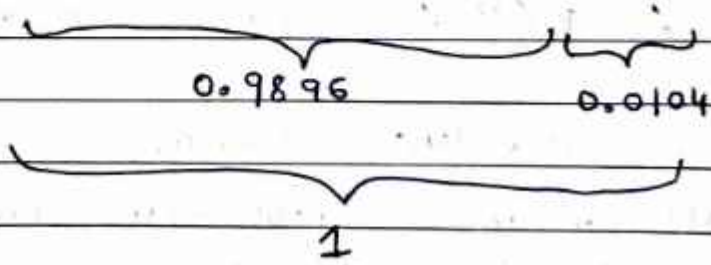
(ii) Two tailed test.
- We test, if the mean is simply different
- you have to check both side of the curve.
- So you multiply by 2 to get full P value.

Suppose : $z = 2.31$.

& z table shows 0.9896

$\therefore P(z < 2.31) = 0.9896$





$1 - 0.9896 = 0.0104$

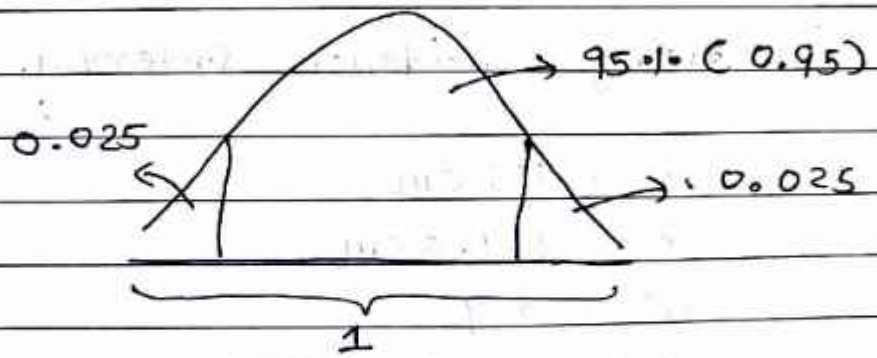
0.0104 area of Right side (tail)
& also 0.0104 area of left tail.
∴ both is 2 × 0.0104
P = 0.0208

→ Two method use for conclusion in z test.

(i) use confidence Interval (C.I) & Find critical z value and compare with original z value.

- In many case: default C.I. = 95%
∴ α = 0.05
∴ two tailed test : critical z = ± 1.96
∴ one tailed test : critical z = + 1.645

→ how can i find z using C.I.



1 - 0.025 = 0.9750 ⇒ Now check
Area z-table

∴ 0.9750 slows in table + 1.96
hence proved.

(ii) calculate the p-value and compare with α.

- compute z score : $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
- find p value using z table.
- if p value $< 0.05 \Rightarrow$ Reject null hypothesis
  if p value $\geq 0.05 \Rightarrow$ do not reject null.

- Problem :

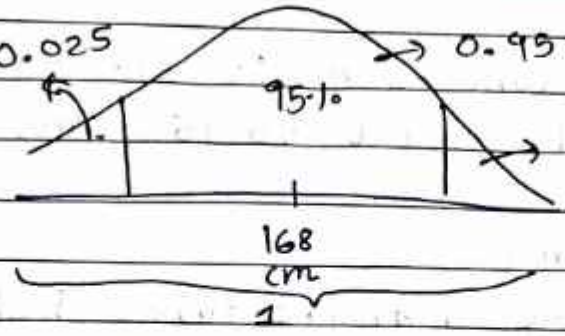
A average heights of all residents in a city is 168 cm with standard deviation $\sigma = 3.9$
A doctor believes the mean to be different. he measure the height of 36 individuals and found the average height to be 169.5 cm
Is doctor's believes is real or not?

we solve using both methods :-

(i) using confidence Interval.

$\mu = 168$ cm
$\bar{x} = 169.5$ cm
$\sigma = 3.9$
C.I (default) = 95%
$\alpha = 0.05$ ($\alpha = 1 - C.I. = 1 - 0.95 = 0.05$)

(i) Null hypothesis $H_0 = \mu = 168$ cm
Alt hypothesis $H_1 = \mu \neq 168$ cm.



$\therefore 1 - 0.9025 = 0.9750$

z score for 0.9750 in z-table is ± 1.96

$\therefore$ critical z score $= \pm 1.96$

Now calculate z score for given data.

$$\therefore z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \therefore z = \frac{169.5 - 168}{3.9/\sqrt{36}}$$

$$\therefore z = \frac{1.5}{0.65} = 2.31$$

conclusion :

$z >$ critical z

$\therefore 2.31 > 1.96$

$\therefore$ Reject the null hypothesis

(ii) using P value method :

$P(z < 2.31) = 0.98956$ (~~Right tail~~).

Now, $1 - 0.98956 = 0.01044$

(Right tail area).

$\therefore$ Both tail $= 2 \times 0.01044$

Both tail $= 0.02088$

$\rightarrow$ $0.02088 < 0.05 \Rightarrow \therefore$ Reject Null Hypothesis (H_0).

∴ Final conclusion is the average height is not equal to 168 cm.

- Problem:

A Factory manufectures bulbs with avg. warranty of 5 years with std of 0.50. A worker believes that the bulb will defected in less than 5 years. he tests 40 bulbs and find the average time of bulb is 4.8 years.

(i) State null & alt hypothesis

(ii) At a 2.1. significant level, is there enough evedence to support the idea that the warranty should be revised?

→ Null hypothesis $(H_0) = \mu = 5$

Alt hypothesis $(H_1) = \mu < 5$ (left tailed)

Z test:

$\mu = 5$, $\bar{x} = 4.8$, $n = 40$, $\sigma = 0.50$

$$\therefore z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$\therefore z = \frac{4.8 - 5}{0.5 / \sqrt{40}}$$

$$\therefore z = (-2.53)$$

z score = (-2.53)

see z table and get
P-value = 0.00570

here α = 2% = 0.02

compare p value with α
∴ 0.00570 < 0.02

∴ we reject the null hypothesis

conclusion:
- company should Revise their warranty timeline.

(3.4) t test and hypothesis testing

- t test uses when:
  - sample size is small ($n \leq 30$)
  - population standard deviation (σ) is unknown

Formula: $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\bar{x}$ = sample mean
μ = population mean
s = sample standard deviation
n = sample size

SUKALA
Page
Date

Steps For performing t test :

(i) State hypotheses
null (H_0)
Alt (H_1)
(ii) calculate t-score using formula
(iii) Find critical t-value from t-table.
(based on degrees of Freedom : $df = n-1$)
(iv) compare your t-score with
critical value or compute the P-value.

Problem:
In a population, the average IQ is 100. A team of researchers to test a new meditation to see if it has eighter positive or negative effect on IQ, or no effect at all. A sample of 30 participants who have taken the meditation has a mean of 140 IQ with a standard deviation of 20. Did the meditation effect intelligence ?

→ $\mu = 100$, $n = 30$, $\bar{x} = 140$, $s = 20$,
$CI = 95\%$, $\alpha = 0.05$

(i) null hypothesis (H_0) = $\mu = 100$
Alt hypothesis (H_1) = $\mu \neq 100$ (2 tail test)

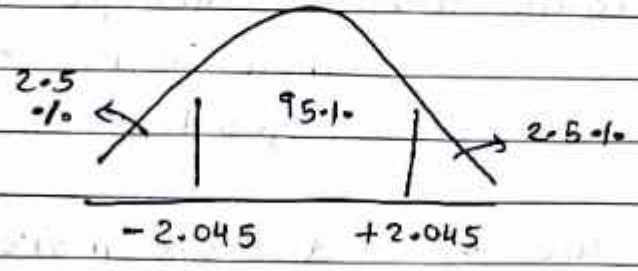
(ii) $\alpha = 0.05$

(iii) Degree of freedom = $n-1$
$= 30-1$
$= 29$.

→ Now use t-table,

dof = 29 & $\alpha = 0.05$ (two tail test)

∴ critical t-score = ± 2.045



If t-score is less than -2.045 or greater than +2.045 than reject the null hypothesis.

(iv) calculat t-score

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{140 - 100}{20/\sqrt{30}} = 10.96$$

conclusion:

t > critical t (10.96 > 2.045)

∴ we reject the null hypothesis

∴ IQ is increasing with meditation.

(3.5) Type 1 & Type 2 Error

- outcome 1 : we reject the null hypothesis when in reality it is False → good.

- outcome 2 : we reject the null hypothesis when in reality it is true → Type 1 error.

- outcome 3 : we Accept the null hypothesis when in reality it is false
→ Type 2 error.

- outcome 4 : we Accept the null hypothesis when in reality it is true
→ good.

Note: we will discuss more about in confusion matrix topic. (stay tuned)

(3.6) Baye's Theorem :

Probability
- → Independent events
- → Dependent events

(i) Independent Events
eg: Rolling dice
toss a coin.
$P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$
etc.

(ii) Dependent Events
- one box have two Red & 3 white balls
- we picking ball one by one.

$P(\text{Red}) = \frac{2}{5}$

Now
$P(\text{yellow}) = \frac{3}{4}$

Next event depends on Previous events.

$\therefore P(\text{Red and white}) = P(R) * P(W/R)$

Now, $P(A \text{ and } B) = P(B \text{ and } A)$

$\therefore P(A) * P(B/A) = P(B) * P(A/B)$

$$\therefore P(B/A) = \frac{P(B) * P(A/B)}{P(A)}$$ Baye's theorm

$$\therefore P(A/B) = \frac{P(A) * P(B|A)}{P(B)}$$

→ It is use in Naiv Baye's algorithm we will deep dive in that algorithm in ML part.

(3.7) CHI Square test

- The chi-square test is used to check if there is significant relationship between two categorical values or
- IF the observed data fits the expected data.
- It is use For ordinal & nominal data.

SUKALA
Page
Date

Example: (chi square goodness of fit).

In a science class of 75 students, 11 are left handed. Does this class fit the theory that 12.1. of people are left handed

	Observation	Expectation
Lef hand	11	9
Right hand	64	66
	75	75

- chi-square test formula:

$$\chi^2 = \sum \frac{(\text{Observat}^n - \text{Expect}^n)^2}{\text{Expectation}}$$

- Solved Example:

In 2010, the weight of the individuals in a small city were found to be Following.

< 50 kg	50-75	> 75
20-1.	30-1.	50-1.

In 2020, weights of 500 individuals were sampled, below is result.

< 50 kg	50-75	> 75
140	160	200

using $\alpha = 0.05$, would you conclude the population difference of weight has changed in the last 10 years?

– Ans :

Expected	<50	50 - 75	>75
	100	150	250

observed.	<50	50 - 75	>75
	140	160	200

– (1) Null hypothesis (Ho) : The data meets the expectation.
Alt. (H1) : The data doesn't meet expectation.

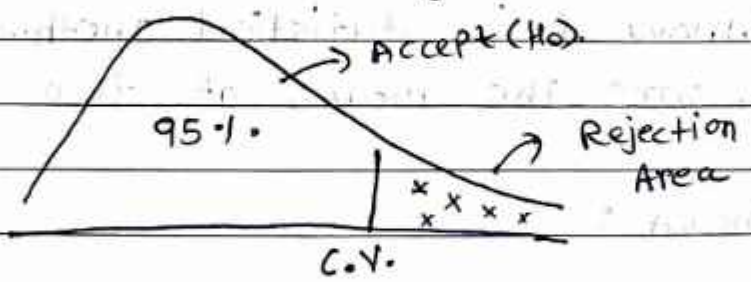
(2) $\alpha = 0.05$, CI = 95%

(3) Degree of freedom

$df = k - 1$ (categories - 1)

$df. = 3 - 2 = 2$

(4) Decision Boundry



check in chi-square table with $df = 2$ & $\alpha = 0.05$

$\therefore$ C.V. = 5.991

$\therefore$ If $\chi^2 > 5.99$, Reject Ho.
else failed to Reject Ho.

(5) calculate chi-square test score

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}$$

$$= \frac{1600}{100} + \frac{100}{150} + \frac{2500}{250}$$

$$\chi^2 = 26.66$$

Conclusion:

$\chi^2 > 5.99$

∴ we Reject (Ho).

- The weight of 2020 population are different than those expected in the 2010 population.

(3.8) ANNOVA Test : (Analysis of variance):

- Annova is a statistical method used to compare the means of two or more groups.

- Annova :

(1) Factors (variable)
(2) Levels

eg: Medicine (factor)
[Dosage] 5 mg 10mg 15 mg → Levels

SUKALA
Page
Date

mode of payment (factor)

[Levels] → GPay, PhonePay, IMPS, NEFT

- Assumption in ANOVA

(i) Normality of sampling distribution of mean
- The data in each group shoud be approximately normally distributed.

(ii) Absense of outliers
- outlying score need to be removed From the dataset

(iii) Homoginity of variance
- All groups should have equal variance (spread of data). This is called homoscedasticity
[$\sigma_1^2 = \sigma_2^2 = \sigma_3^3$]

(iv) Independence
- samples must be independent of each other (no overlap in group)

- Types of ANOVA :
- There are main three types of ANOVA depending on the number of factors and group involved.

(i) One way ANOVA
(ii) Two way ANOVA
(iii) Repeated measures ANOVA.

SUKALA
Page
Date

(i) one way ANOVA :

- use to compare means of 3 or more groups (Levels) for 1 Independent Variable (factor).

eg : compare weight loss acrross 3 different diets.

weight loss (Independent variable or. factor):

keto	Vegan	Protein
5.1	4.1	6.2
6.0	3.5	7.0

→ Now, compare mean weight loss across 3 grps

(ii) Two way ANOVA :

- use when we have 2 Independent variable (Factors), and want to check:
  - effect of each group
  - Interaction effect between them

eg : Does diet type & exercise type affact weight loss ?

factor (Diet) — Keto (level 1), Vegan (level 2), Protein (level 3)

Exercise (factor) — Cardio (level 1), Strength (level 2)

Diet type	Exercise type	weight loss (kg).
keto	cardio	5.0
keto	Strength	6.1
vegan	cardio	4.0
Vegan	strength	4.8
Protein	cardio	6.3
protein	strength	6.9

(iii) Repeated Measures ANOVA:

- Measure the same group or individual multiple times over times or under different condition.

eg: does a weight loss program change weight over time?

Time	Weight (kg)
Before	80
After 1 months	78
After 2 months.	75

⇒ Solved Example of ANOVA Test: (One way ANOVA).

A Doctor want to test a new medication which reduce headache. They split the participant into 3 condition (15 mg, 30mg, 45 mg). Later on the doctor ask the patient to rate the headache between [1-10]. Are there any different between the three conditions., alpha = 0.052

SUKALA
Page
Date

Ans :

15 mg	30mg	45mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

(i) $H_0 = \mu_{15} = \mu_{30} = \mu_{45}$

H_1 = not all μ are equal or any one are different.

(ii) Significant $\alpha = 0.05$, $CI = 0.95$

(iii) calculate degree of freedom.

$N = 21$, $n = 7$, $a = 3$

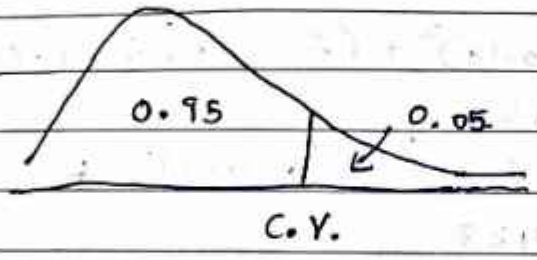
$df_{between} = a - 1 = 3 - 1 = 2$

$df_{within} = N - a = 21 - 3 = 18$

$df_{total} = N - 1 = 20$

Now, df_1 df_2 $(2, 18)$ & $\alpha = 0.05$
check F table and Find critical value (C.V).

(iv) Decision Boundry.



$\therefore$ C.V. = 3.5546

if F is greater than 3.5547, than reject the null hypothesis

(v) Group mean & overall mean

$\bar{x}_1$ (15 mg) = (9+7+8+8+8+9+8)/7 = 8.14
$\bar{x}_2$ (30 mg) = (7+6+6+7+8+7+6)/7 = 6.71
$\bar{x}_3$ (45 mg) = (4+3+2+3+4+3+2)/7 = 3

$$\bar{x} \text{ (overall)} = \frac{57+47+21}{21} = \frac{125}{21} = 5.95$$

(vi) sum of square between group (SSB).

$$SSB = \sum n_i (\bar{x}_i - \bar{x})^2$$

- 15 mg = $7 \times (8.14-5.95)^2 = 7 \times 4.82 = 33.74$
- 30 mg = $7 \times (6.71-5.95)^2 = 7 \times 0.58 = 4.06$
- 45 mg = $7 \times (3-5.95)^2 = 7 \times 8.70 = 60.89$

SSB = 33.74 + 4.06 + 60.89
SSB = 98.69

(vii) sum of square within groups (SSW)

- 15 mg $= (9-8.14)^2 + (8-8.14)^2 + (7-8.14)^2 + \dots$
$= 2.853$
- 30 mg $= (7-6.71)^2 + (6-6.71)^2 + \dots$
$= 3.428$
- 45 mg $= (4-3)^2 + (3-3)^2 + (2-3)^2 + \dots$
$= 4$

$\therefore SSW = 2.853 + 3.428 + 4$

$\therefore SSW = 10.281$

(viii) Mean Squares.

$MSB = SSB / dFB = 98.69 / 2 = 49.345$

$MSW = SSW / dFW = 10.281 / 18 = 0.571$

(ix) F-statistics.

$$F = \frac{MSB}{MSW} = \frac{49.345}{0.571} = 86.42$$

conclusion:

$86.42 > 3.55$

$\therefore$ we, Reject the null hypothesis (H_0)

All medicines effects are different.

ch : 3

Pro Baby

SUKALA
Page
Date

Index :

1. Introduction to probability
   1.1 Addition and Multiplication rule
2. Probability Distribution Function.
   2.1 Probability Mass function (PMF)
   2.2 cumulative distribution function (CDF)
   2.3 Probability density function (PDF)
3. Types of Probability distribution.
   3.1 Bernoulli distribution
   3.2 Binomial distribution
   3.3 Poisson distribution
   3.4 Normal Gaussian distribution
   3.5 Standard Normal distribution
   3.6 Uniform distribution
   3.7 Log-Normal distribution
   3.8 Power Law distribution
   3.9 Pareto distribution.
   3.10 Central Limit theorem
   3.11 Estimates.

SUKALA
Page
Date

(1) Introduction to Probability :

(1.1) Addition and Multiplication Rule

(i) Intro
(ii) Addition rule (For mutually exclusive event)
(iii) Addition rule (for non mutual ex. event)
(iv) Multiplication rule (Independent & Dependent eve

(i) **Probability :** It is about determining the likelihood of an event.

eg : toss coin {H, T}

$P(H) = \frac{1}{2} = 50\%$

$P(T) = 1/2 = 50\%$

(ii) Mutually Exclusive Events

- Two events are mutually exclusive if they cannot occure at the same time

eg : togsing a coin (Head & tail cannot both come at same time).

- $P(H \text{ or } T) = P(H) + P(T)$ (Addition Rule for mutually exclusive)
$= \frac{1}{2} + \frac{1}{2} = 1$

- Rolling a dice: {1, 2, 3, 4, 5, 6}

$P(1 \text{ or } 5) = P(1) + P(5)$

$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

SUKALA
Page
Date

(iii) Non- mutual Exclusive events.

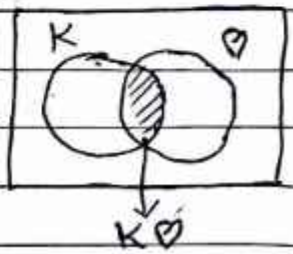
eg: Taking a card From the deck.

$$P(K \text{ or } \heartsuit) = P(K) + P(\heartsuit) - P(K \text{ and } \heartsuit)$$

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$= \frac{17}{52} - \frac{1}{52}$$

$$\therefore P(K \text{ or } \heartsuit) = \frac{16}{52}$$



(iv) Multiplication Rule (Independent and Dependent events)

- Two events are Independent, if they do not affect one another.

eg: Tossing a coin {H or T}
Rolling a dice {1, 2, 3, 4, 5, 6}

- Two events are dependent, if they affect each other

eg: Take a king From the deck and then the queen card From the deck.

$$P(K) = \frac{4}{52}, \quad P(Q) = \frac{4}{51}$$

← multiplication rule :

(1) Independent Event : {Tossing coin}

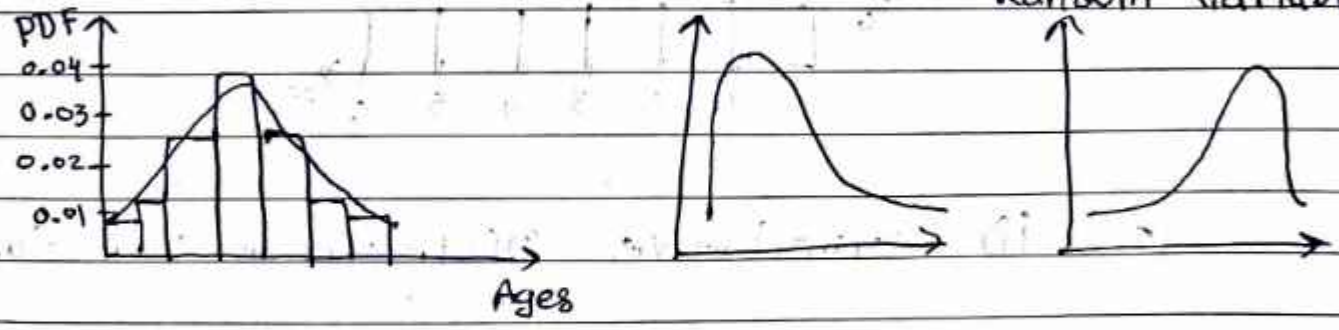
$$P(H \text{ and } T) = P(H) * P(T)$$
$$= \frac{1}{2} \times \frac{1}{2}$$
$$= \frac{1}{4}$$

(2) Dependent Event : → conditional Pro.

$$P(K \text{ and } Q) = P(K) * P(Q/K)$$
$$= \frac{4}{52} * \frac{4}{51}$$

(2) Probability Distribution Function :

- Age = {5, 7, 9, 12, 11, 25, ...} ⇒ continuous Random variable



(i) Probability mass Function (PMF): used for discrete random variable.

(ii) Probability density Function (PDF): used for continuous random variable.

(iii) cumulative distribution fⁿ (CDF): used for both discrete & ~~Rant~~ continuous rand. var.

SUKALA

(2.1) (i) Probability Mass Function : (PMF)
[Descrete Random variable]

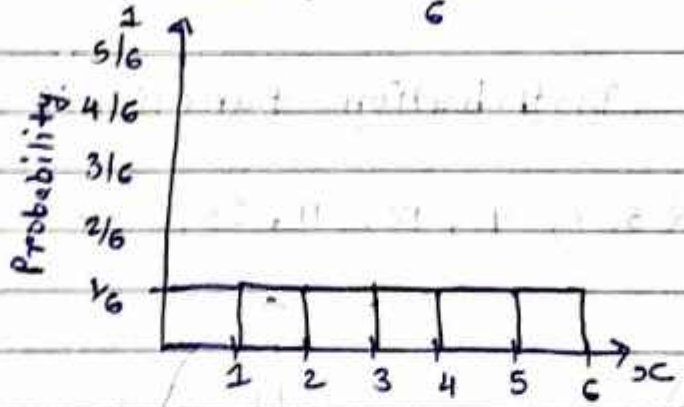
- Discrete random variable takes countable value like {0, 1, 2, 3, ..}
- PMF gives the probability that a discrete random variable equal to a specific value.

$$P(X = x) = PMF(x)$$

eg Rolling a dice $X = \{1, 2, 3, 4, 5, 6\}$

$$P(X = 1) = \frac{1}{6}$$

$$P(X = 5) = \frac{1}{6} \quad \text{etc.....}$$



(2.2).(ii) Cumulative Distribution Function : (CDF)

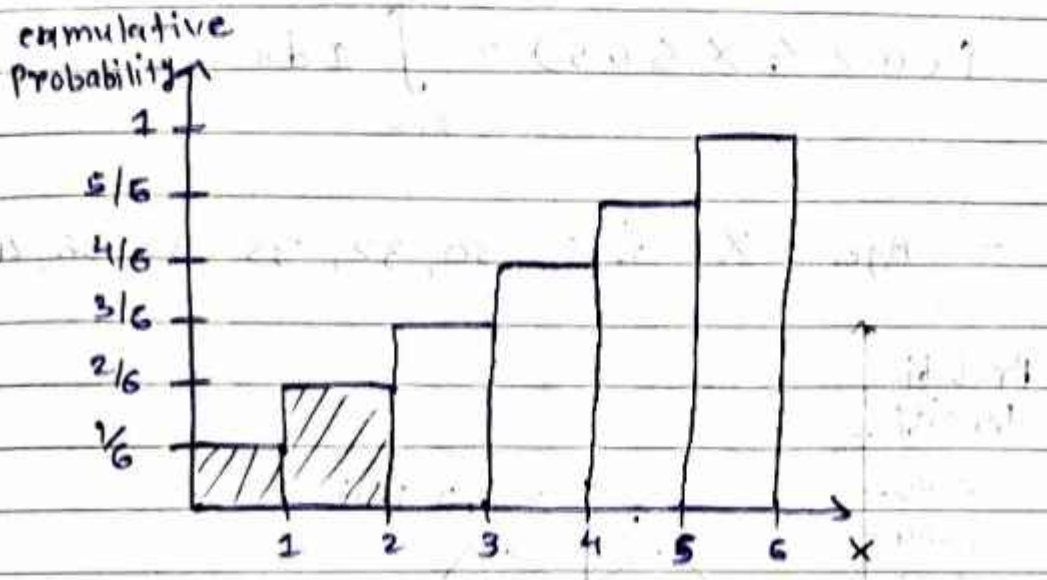
- Used for both discrete and continuous random variable
- CDF gives the probability that a variable is less than or equal to a value x.

$$F(x) = P(X \leq x)$$

eg :- Rolling a dice :

$$P(X \leq 3) = P(1) + P(2) + P(3)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

$$P(X \leq 3) = \frac{3}{6} = 0.5$$



$$P(X \leq 2) = P(1) + P(2)$$
$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

(2.3) (iii) Probability Density Function : (PDF)

- used for continuous random variables.
- PDF is a function such that the area under the curve between two values gives the probability that the variable lies in the range.

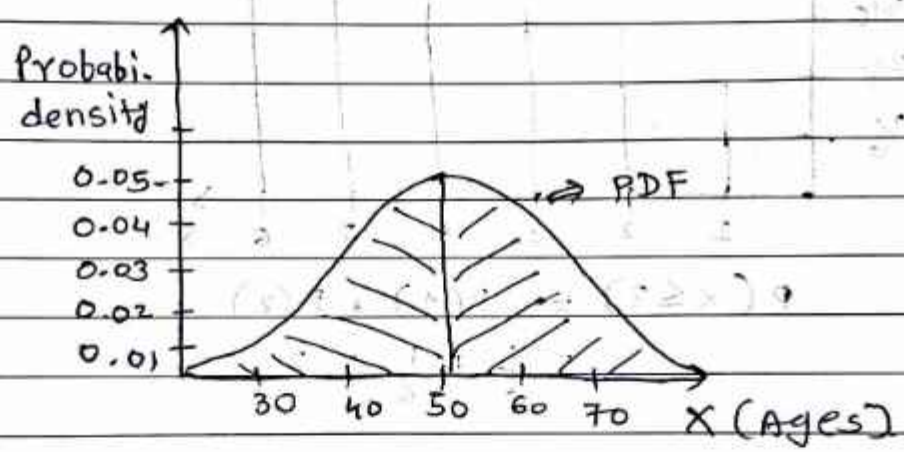
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

eg : Let X be a continuous variable with uniform distribution between 0 & 1.

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

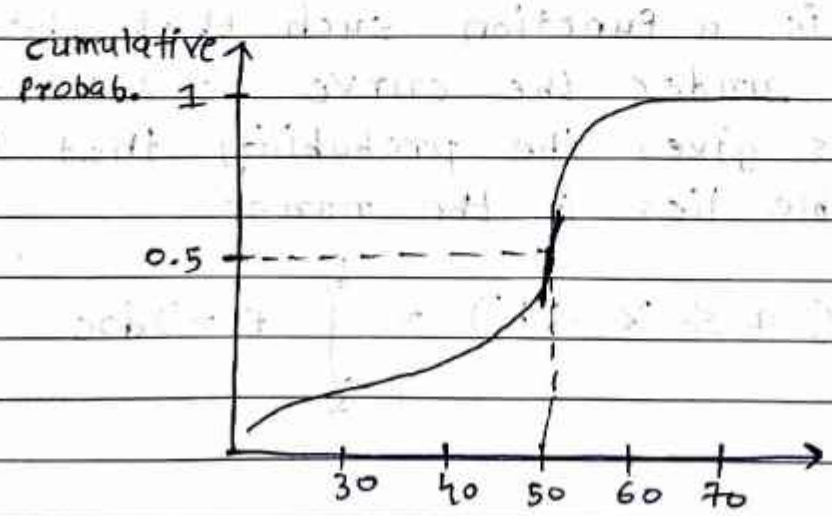
$$P(0.2 \le X \le 0.5) = \int_{0.2}^{0.5} 1 \, dx = 0.5 - 0.2 = 0.3$$

- Age = X = {..25, 30, 32, 35, 37, 38, 40, 45, ...}



$$P(X \le 50) = 0.5 = 50\%$$

CDF graph :



- PDF Properties :

(i) Non negativity $f(x) \geq 0$ for all x.

(ii) The total area under the PDF curve is equal to 1.

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

With respect to different distribution $f(x)$ function is going to change.

(3) Types of Probability distribution.

There are many types of Probability distribution:

(i) Bernoulli distribution
(ii) Binomial distribution
(iii) Poisson distribution
(iv) Normal gaussian distribution
(v) Standard Normal distribution
(vi) Uniform distribution
(vii) Log normal distribution
(viii) Power Law distribution
(ix) Pareto distribution
(x) Central limit theorem
(xi) Estimates.

SUKALA
Page
Date

(3.1) Bernoulli distribution :

- The bernoulli distribution is the simplest discrete probability distribution. It represent the probability distribution of a random variable that has exactly two possible outcomes: success (with probability p) and failure (withe probability 1-p). It is used to model binary outcomes, such as a coin flip or yes/no questions.

- Tossing a coin → Head (1), Tail (0).
- Student passes (1) or failes (0) an exam.
- customer buys (1) or doesn't buy (0).

- If a random variable X follows bernoulli distribution, it means:

$$X \sim \text{Bernoulli}(p)$$

where:
- p = Probability of success ($X = 1$)
- $(1-p)$ = probability of failure ($X = 0$)

- Probability Mass function (PMF):

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

$$PMF = p^k \cdot (1-p)^{1-k}$$

if $k = 1$ (success)

$$P(k=1) = p^1 \cdot (1-p)^{1-1} \Rightarrow p$$

if $k = 0$ (failure)

$$P(k=0) = p^0 \cdot (1-p)^1 \Rightarrow (1-p) \Rightarrow q$$

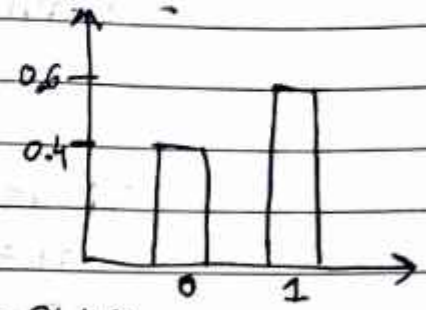
- Mean of Bernoulli distribution :

$$E(X) = \sum_{k=0}^{1} k \cdot P(k) \qquad \rightarrow K = \{0, 1\}$$

$$= (0) \cdot (0.40) + (1) \cdot (0.60)$$

$$= 0 + 0.60$$

$$E(X) = 0.60 = P$$



$$\therefore q = 1 - P = 1 - 0.60 = 0.40$$

- Variance :

$P(K=0) = 0.4 = q$

$P(K=1) = 0.6 = P$

$$\sigma^2 = P \cdot q$$

$$\sigma^2 = 0.4 * (0 - 0.6)^2 + 0.6 (1 - 0.6)^2$$

$$\sigma^2 = 0.4 * 0.36 + 0.6 * 0.16$$

$$\sigma^2 = 0.24 \quad (P * q)$$

$\therefore$ variance $\sigma^2 = P \cdot q$ or $P \cdot (1 - P)$

(3.2) Binomial Distribution :

- The binomial distribution is used to model the number of successes in a fixed number of independent experiments (called trials), where each trial has only two outcomes :
  - success (1)
  - failure (0)

SUKALA
Page
Date

eg: - flipping a coin 10 times. (H or T)
- Sending 100 emails (opened or not)

- Formula for binomial distribution:

$$X \sim Binomial(n, p)$$

where,
- n = number of trials.
- p = Probability of success in one trial
- X = number of success in n trial

(PMF):

$$P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

or

$$P(k, n, p) = {}^{n}C_{k} \cdot p^k \cdot (1-p)^{n-k}$$

where,
- $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ is binomial coefficient
- k = number of success $(0 \leq k \leq n)$

eg: What is probability of getting exactly 3 heads in 5 coin tosses?

- $n = 5$
- $p = 0.5$
- $k = 3$

$\therefore P(X=3) = \binom{5}{3} (0.5)^3 \cdot (0.5)^2$

$= 10 \times 0.125 \times 0.25$

$P(X=3) = 0.3125$

So, there are 31.25 % chance of getting exactly 3 heads in 5 tosses.

- Mean : $E(X) = n \cdot p$
- Variance : $Var(X) = n \cdot p \cdot q$
- Std : $\sqrt{n \cdot p \cdot q}$

(3.3) Poisson Distribution :

- The poisson distribution is used to model the number of times an event happens in a fixed interval of time, space, distance, when events happen randomly and independently at constant average rate.

- Poisson answers the question:
"How likely is it that an event happens k times in fixed period, given the average number of times it usually happens?"

eg :
- Calls at a call center per hour
- Emails received per day
- Road accidents per month in a city
- Typos per page in book.

SUKALA
Page
Date

- If random variable X follows a poisson distribution, we write.

$$X \sim \text{Poisson}(\lambda)$$

where,

λ = avg. number of events per interval

X = actual number of events in that interval

- Probability Mass Function (PMF):

$$P(X = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

- k = no. of events
- $e \approx 2.718$
- λ = avg. rate (mean no. of events)

eg. A call center receive 5 calls per hour on average.
What is the probability, it receive exactly 3 calls in one hour?

Let,

$\lambda = 5$, $k = 3$

$$P(X = 3) = \frac{5^3 \cdot e^{-5}}{3!} = \frac{125 \cdot e^{-5}}{6} \approx 0.1404$$

So, there is a 14.04% chance of getting exactly 3 calls in an hour.

(3.4) Normal gaussian distribution :

- Normal distribution is a bell shaped curve that represents how data is distributed around the mean (avg).

Real life example :

- Height of people. (most people are avg. height)
- Exam score. (most student score around the avg)

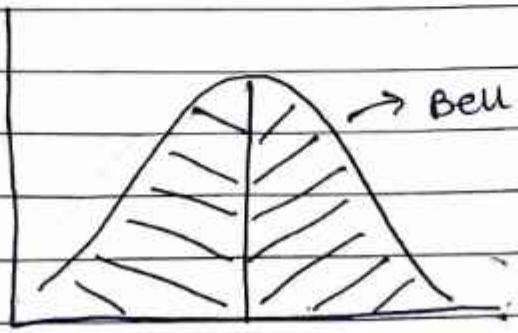
Shape:

- symmetrical around the mean
- has a peak at the mean
- Tails of equally on both sides.

Formula of Normal distribution :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$$

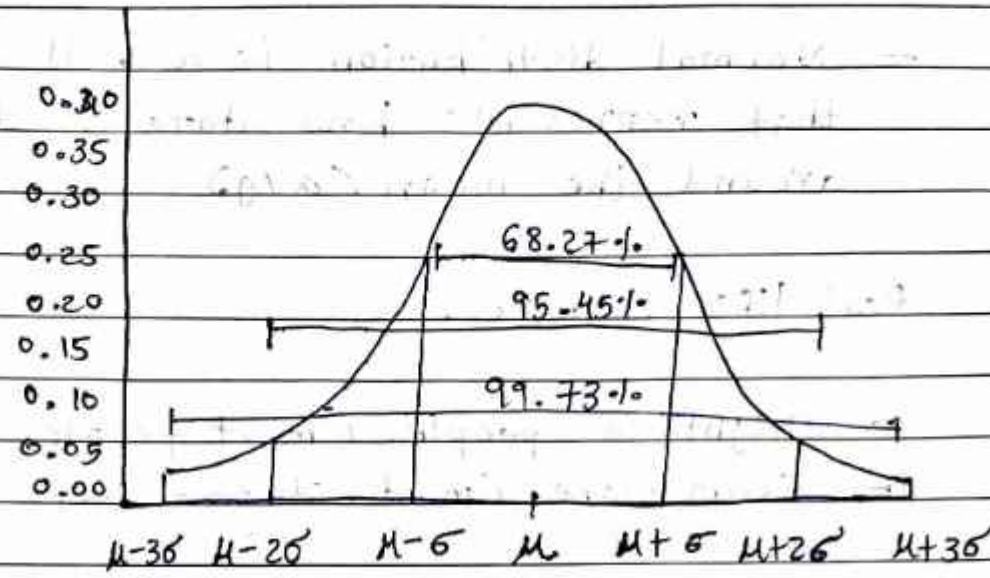
where, x = value
μ = mean
σ = standard deviation



μ = median = mode

SUKALA
Page
Date

- Emperical rule of Normal distribution.

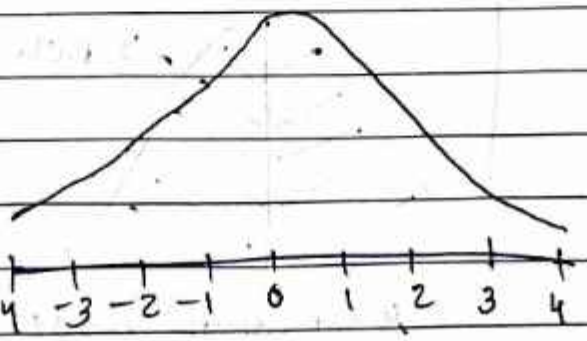


- 68 % of data lies within 1σ of the mean
- 95 % with in 2σ
- 99.7 % within 3σ

(3.5) Standard Normal Distribution

- standard normal distribution is just a special case of the normal distribution where:
  - mean $\mu = 0$
  - Standard deviation $\sigma = 1$

It also bell-shaped, but centered at 0.



- Why we use standard Normal?

To compare different datasets or values, we standardize them using the z-score formula:

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

∴ X = your actual data point
μ = mean of the data
σ = Standard deviation.

eg: Exam score are normally distributed
mean = 70, SD = 10
You scored 85, what is your z-score?

$$z = \frac{85 - 70}{10} = 1.5$$

You score is 1.5 standard deviation above mean

(3.6) Uniform Distribution

- Uniform distribution is a type of probability distribution where all outcomes are equally likely.

eg:
- Tossing a fair dice
- Picking a card randomly
- choosing random number between 1 and 10.

- Types of Uniform distribution-

(i) Discrete uniform distribution
- Finite set of outcomes
- eg: Rolling a die → {1, 2, 3, 4, 5, 6}

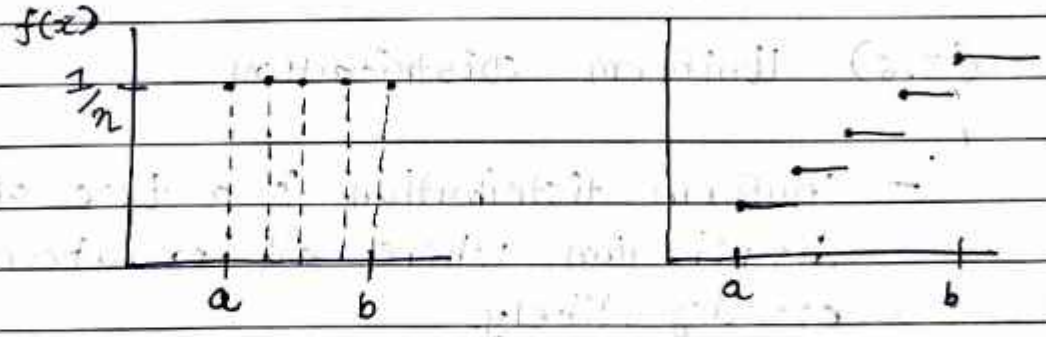
(ii) continuous uniform distribution
- Infinite number of outcomes within range
eg: Picking real number between 0 & 1.

(i) Discrete uniform distribution

- every one of n value has equal probability 1/n.

If a random variable X can take n value equally likely, then:

$$P(X = x_i) = \frac{1}{n}, \text{ for all } x_i$$



(PMF) (CDF)

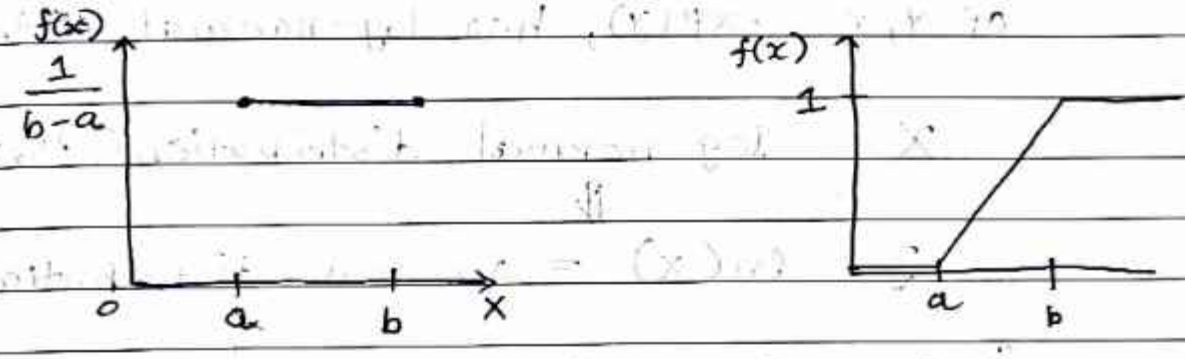
SUKALA
Page
Date

(ii) Continuos uniform distribution

If X is a continuous random variable between a and b, then the probability density function (PDF) is :

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \le x \le b$$

and, $P(a \le X \le b) = 1$



(PDF) (CDF)

eg: choose random number between 0 & 10. what is prob. picking number between 2 & 6?

$a = 0, \; b = 10$

$$f(x) = \frac{1}{10-0} = 0.1$$

$$P(2 \le X \le 6) = (6-2) \times 0.1 = 0.4$$

∴ So, there are 40.1. chance the number lies between 2 & 6.

- Mean $= \frac{1}{2}(a+b)$
- Variance $= \frac{1}{12}(b-a)^2$

(3.7) Log normal distribution.

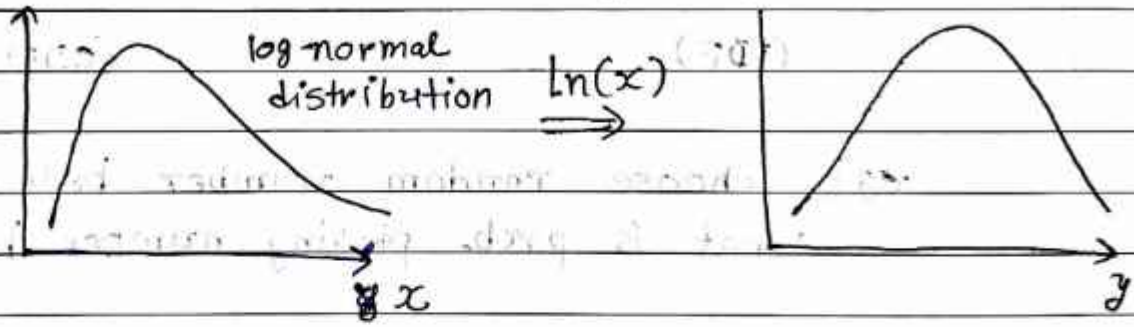
- A Log-normal distribution is a probability distribution of a random variable whose logarithm is normally distributed

- If the random variable X is log-normally distributed, then $y = \ln(x)$ has a normal distribution, equivalently, if Y has a normal distribution, then the exponential function of Y, $X = \exp(y)$, has log-normal distribution.

X = log normal distribution (μ, σ)

$\Downarrow$

$Y = \ln(X)$ = Normal distribution.



eg: - Wealth distribution of the world.
- length of comments (only few long)
- Salary of employees in a company.

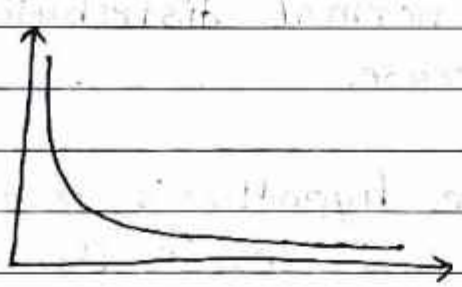
SUKALA
Page
Date

(3.8) Power law distribution.

- Power law is a mathematical relationship where small occurences are extremely common and large instances are extremly rare, but still more likely than in normal distribution.

$$P(x) \propto x^{-\alpha}$$

x = value of interest
α = exponent (also called pareto index)
$\propto$ = proportional to



(3.9) Pareto distribution.

- Pareto distribution is a special case of power law, used to describe situations where small number of things account for a large portion of the effects.

- 80/20 Rule (Pareto principle).

- 80% Result outcome From 20% of causes.

- Real life examples :

- 20% peoples have 80% of wealth
- 80% of peoples leaving in 20% cities.
- 80% of entire project is done by 20% of the team.

(3.10) Central Limit Theorem

- If you take many random samples From any population (no matter its shape), the average (mean) of those samples will form a normal distribution as the sample size increase.

- used for hypothesis testing, confidence Interval etc, even if the original data is not normal

Real life example :

- you take random groups of 30 peoples
- calculate the average brushing time for each group.
- Repeat this 1000 times.
- The distribution of these average will look like bell curve.

$$\text{Standard Error} = \frac{\sigma}{\sqrt{n}}$$

where:

σ = population standard deviation

n = sample size;

As n increase, standard error gets smaller, meaning sample means get closer to the true population mean.

eg: - you have population data of people's montly spending (not normal distribut^n).

- population mean $\mu = 10,000$ ₹

  standard deviation $\sigma = 2,000$ ₹

Now,

- Take 50 samples ($n = 50$), each of 30 peoples.
- Find mean spending of each sample.
- plot the 50 means.
- Distribution will be normal distribution.

= x = x =

ch : 4

Calculus

SUKALA Page Date

Index :

1. Differential calculus
2. Power rules and derivative rules
3. Product rules
4. Chain rule of derivatives.

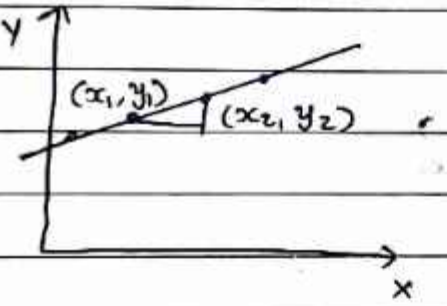
SUKALA
Page
Date

(1) Differential calculus :

- slope :

The slope of a line is a measure of how steep the line is, and it represents the rate of change of one variable with respect to another.

The slope indicates the ratio of the vertical change (rise) to the horizontal change (run) between two points on a line.



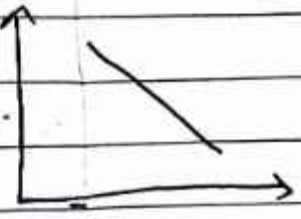
$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x} = \frac{\text{rise}}{\text{run}}$$

since the line is straight
Rate of change = constant.

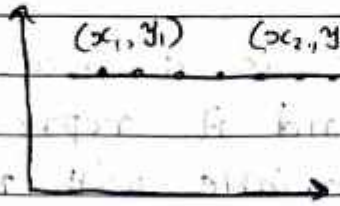
- Interpretation of slope :

(i) Positive slope : if slope > 0, the line rise as it moves from left to right, the larger the slope, the steeper the line.

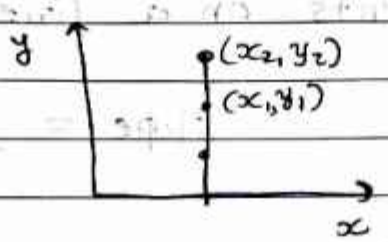
(ii) Negative slope : if slope < 0, the line falls as it move from left to right, the more nagative the slope, the steeper the line in downward direction.



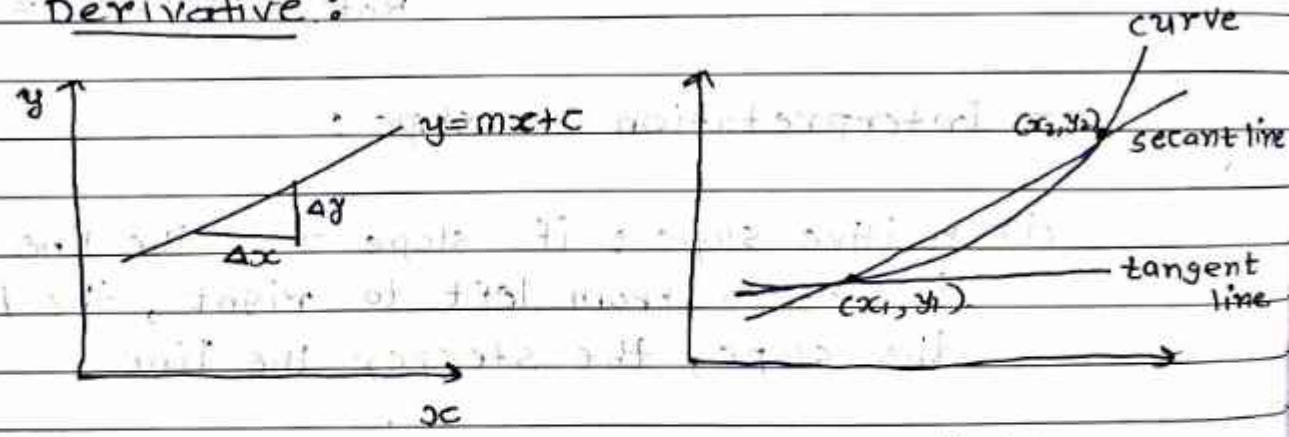
(iii) Zero slope :: if slope = 0, the line is horizontal, means there is no vertical change as the line move from left → Right

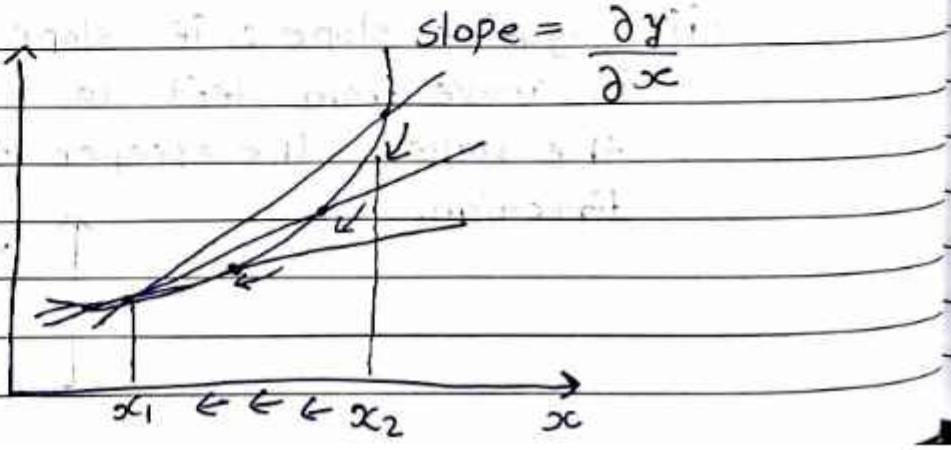


(iv) Undefined slope : if $x_2 = x_1$, the line is vertical, and the slope is undefine.



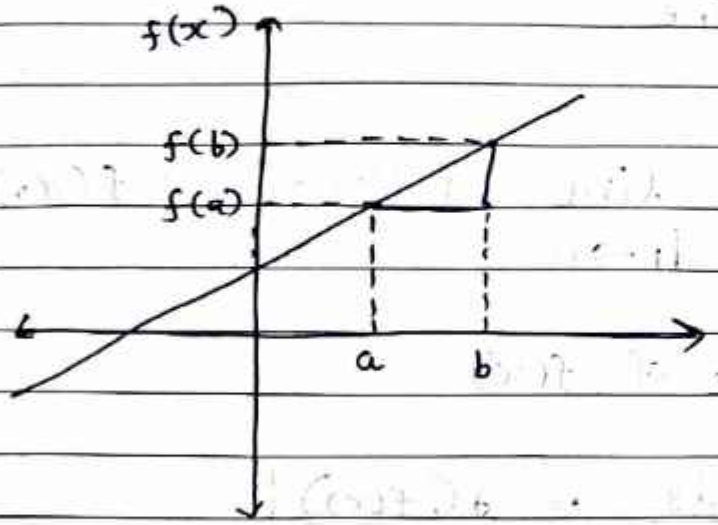
Derivative :





SUKALA
Page
Date

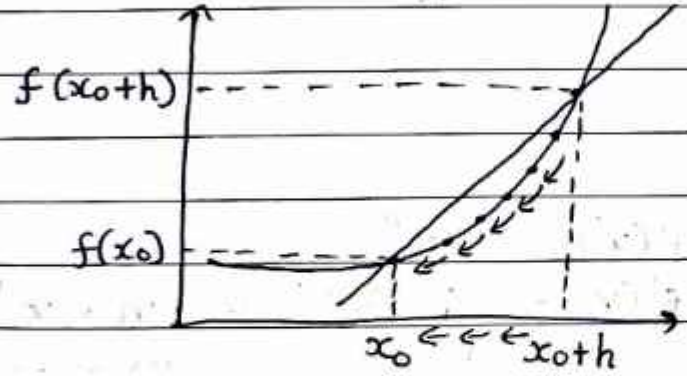
– Mathematical Notation of Derivative with Limits



$$f(x) = mx + c \qquad (m = \text{slope},\ c = \text{Intercept})$$

$$\text{slop} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

$$\boxed{\frac{\Delta y}{\Delta x} = \frac{f(b) - f(a)}{b - a}}$$



Now, $(h \to 0)$

$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(x_0+h) - f(x_0)}{(x_0+h) - x_0} = \frac{\Delta y}{\Delta x}$$

slope of secent line $= \frac{f(x_0+h) - f(x_0)}{h}$

$$f'(x) = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$$

↑ derivative of $f(x)$

$$f'(x) = \frac{dy}{dx} = \frac{\partial(f(x))}{\partial x}$$

(2) Power rules and derivative rules :

- Power rule :

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Formula:

If : $f(x) = x^n$

Then: $f'(x) = \frac{d}{dx}(x^n) = n \cdot x^{n-1}$

if : $n = 0$ then : $f(x) = x^0 = 1 = \text{constant}$

↓

slope = 0 $f'(x) = 0$

(Derivative of constant = 0)

SUKALA
Page
Date

eg : (1) $f(x) = x^5$

$f'(x) = 5x^4$

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

$f'(x) = -2x^{-3}$

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

$f'(x) = \frac{1}{2} x^{-1/2} = \frac{1}{2\sqrt{x}}$

(4) $f(x) = 7x^3$

$f'(x) = 21x^2$

- sum of two functions ($f(x)$, $g(x)$)

$$\frac{\partial}{\partial x}[f(x) + g(x)] = \frac{\partial(f(x)}{\partial x} + \frac{\partial(g(x))}{\partial x}$$

$$\frac{\partial}{\partial x}[x^4 + x^{-2}] = \frac{\partial(x^4)}{\partial x} + \frac{\partial(x^{-2})}{\partial x}$$

" $= 4x^3 + (-2x)$

" $= 4x^3 - 2x$

- Derivative of trignometric, Logarithmic, exponential, power and constant function.

(i) Trignometric function :

function	derivative
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$\sec^2(x)$
$\cot(x)$	$-\text{cosec}^2(x)$

eg: $f(x) = \sin(x) + \cos(x)$

$f'(x) = \cos(x) - \sin(x)$

(ii) Logarithmic function.

$f(x) = \ln(x) \longrightarrow f'(x) = \frac{1}{x}$

$f(x) = \ln_a(x) \longrightarrow f'(x) = \frac{1}{x \cdot \ln(a)}$

(iii) Exponential function.

$f(x) = e^x \longrightarrow f'(x) = e^x$

(iv) constant function

$f(c) = 0$

(v) Polynomial function.

$f(x) = x^n \longrightarrow f'(x) = n x^{n-1}$

(3) Product Rule :

$$\frac{\partial}{\partial x}[h(x) \cdot f(x)] = h'(x) \cdot f(x) + h(x) \cdot f'(x)$$

eg : $\frac{\partial}{\partial x}(x^2 \cos x) = \frac{\partial}{\partial x}(x^2) \cdot \cos x + x^2 \cdot \frac{\partial(\cos x)}{\partial x}$

$$= 2x \cdot \cos x + x^2(-\sin x)$$

$$\frac{\partial}{\partial x}(x^2 \cdot \cos x) = 2x \cos x - x^2 \sin x$$

(4) chain Rule of Derivative : (use in NN)

- The chain rule is a fundamental theorem in calculus that is used to find the derivative of a composite function.
- When a function is composed of other function, the chain rule allow us to differentiate it with respect to the innermost variable.

If $y = f(g(x))$
when $y = f(u)$ and $u = g(x)$, then derivative of y with respect to x.

$$\boxed{\frac{\partial y}{\partial x} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x}}$$

In simple term:

$$\frac{\partial y}{\partial x} = f'(g(x)) \cdot g'(x)$$

This mean that to differentiate a composite function, you first differentiate the outer function with respect to inner function and then multiply by derivative of inner function.

eg: (1) $y = (3x^2 + 2x + 1)^5$

Identify the outer, and Inner function

outer f^n : $f(u) = u^5$

Inner f^n : $u = g(x) = 3x^2 + 2x + 1$

Differentiat the outer function.

$$f'(u) = \frac{\partial f}{\partial u} = u^5 = 5u^4$$

Differenttiat the Inner function

$$g'(x) = \frac{\partial u}{\partial x} = \frac{\partial (3x^2 + 2x + 1)}{\partial x} = 6x + 2$$

$$\therefore \frac{\partial y}{\partial x} = f'(g(x)) \cdot g'(x)$$

$$= 5u^4 \cdot (6x + 2)$$

$$\frac{\partial y}{\partial x} = 5(3x^2 + 2x + 1)^4 \cdot (6x + 2)$$

(ii) $y = \sin(4x^3 + x)$

outer function = $\cos(4x^3 + x)$.
Inner function = $12x^2 + 1$

$$\frac{\partial y}{\partial x} = \cos(4x^3 + x) \cdot (12x^2 + 1)$$

(iii) $y = \sqrt{\sin(3x)}$

$\therefore$ outer f^n : $\sqrt{u} = u^{\frac{1}{2}} = \frac{1}{2\sqrt{u}}$

$\therefore$ middle f^n : $\sin(3x) \rightarrow \cos(3x)$

$\therefore$ Inner f^n : $3x \rightarrow 3$

$$\frac{\partial y}{\partial x} = \frac{1}{2\sqrt{\sin(3x)}} \cdot \cos(3x) \cdot 3$$

$$\therefore \frac{\partial y}{\partial x} = \frac{3\cos(3x)}{2\sqrt{\sin(3x)}}$$

- Application of chain rule :

(i) Backpropogation In NN
(ii) Gradient decent optimization
(iii) Regularization technique (overfitting, underfitting)

= x = x =