Data Science and Machine Learning: Basics of Mathematics

Dr. Manojkumar C. Ramteke, Dr. Hariprasad Kodamana, Dr. Agam Gupta

Department of Chemical Engineering

IIT Delhi

2023

Outline

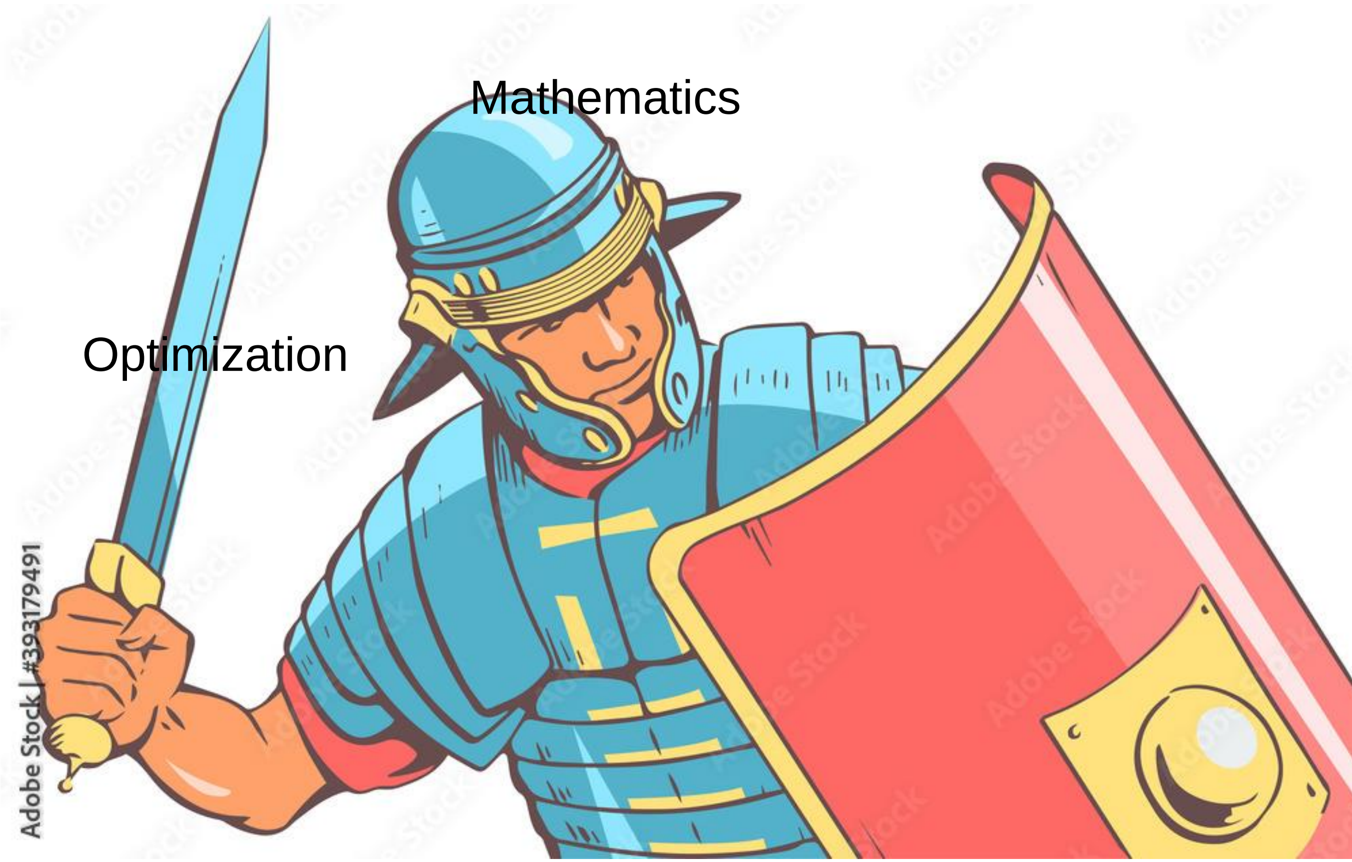
1) Algebraic Equations
2) Determinants
3) Linear equations and Matrices
4) Probability
5) Statistics
6) Discussion

Reference Book: Advanced Engineering Mathematics by Erwin Kreyszig

Pillars of Machine Learning

Mathematics

Optimization



Data Science & Statistics

Battle (Machine Learning)

Source: https://stock.adobe.com/ca/images/roman-soldier-legionnaire-in-battle-in-the-hands-of-a-sword-and-a-large-shield-there-is-a-helmet-on-the-head-armor-on-the-body-vector-illustration/393179491

What is Machine Learning?

Real Systems with Complex Models (Differential Equations with few parameters)

Represented using Simple Models (Mostly Linear Equations with more parameters)

Data used for Optimizing the Parameters (Training)

Trained Model with Simple Equations is used for Faster Prediction (Testing)

Accurate but Response is Time Consuming

Approximate but Quick Response

Algebraic Equations

An expression of the form: $$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$$

where a's are constants ($a_0 \neq 0$) and n is positive integer, is called a polynomial in x of degree n. The above polynomial $f(x) = 0$ is called an algebraic equation of degree n. The value of x which satisfies $f(x) = 0$ is called root. Geometrically, a root is the value where the graph y = f(x) crosses the x-axis. The process of finding the roots of an equation is known as solution of that equation. This is extremely important in applications.

General Properties

1) If α is a root of the equation $f(x) = 0$, then the polynomial $f(x)$ is exactly divisible by x- α. For instance, 3 is a root of equation $x^4 - 6x^2 - 8x - 3 = 0$, because x = 3 satisfies this equation i.e., x – 3 divides the equation completely and it represents its factor.
2) Every equation of nth degree has n roots (real or imaginary) i.e., if $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the roots on nth degree equation $f(x) = 0$, then $f(x) = A(x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_n)$.
3) If f(a) and f(b) have different signs, then the equation $f(x) = 0$ has at least one root between x = a and x = b.
4) In an equation with real coefficients, imaginary roots occur in conjugate pairs i.e., if α + iβ is a root of equation $f(x) = 0$, then α - iβ must also be its root.
5) The equation $f(x) = 0$ cannot have more positive roots than the changes of signs in $f(x)$; and more negative roots than the changes of signs in $f(-x)$

Algebraic Equations

Interesting Properties

If $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the roots on nth degree equation $f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0$.

Then,

$\alpha_1 + \alpha_2 + \ldots + \alpha_n = - a_1/a_0$

$\alpha_1 \alpha_2 + \alpha_2 \alpha_3 + \ldots + \alpha_n \alpha_1 = a_2/a_0$

$\alpha_1 \alpha_2 \alpha_3 + \alpha_2 \alpha_3 \alpha_4 + \ldots + \alpha_n \alpha_1 \alpha_2 = -a_3/a_0$

.

.

.

$\alpha_1 \alpha_2 \ldots \alpha_n = (-1)^n a_n/a_0$

Example: Solve the equation $x^4 - 2x^3 + 4x^2 + 6x - 21 = 0$ given that the sum of two of its roots is zero

Let the roots be $\alpha, \beta, \gamma, \delta$ such that $\alpha + \beta = 0$

$\alpha + \beta + \gamma + \delta = 2 \Rightarrow \gamma + \delta = 2$

Thus, the quadratic factor corresponding to α, β is of the form $x^2 - 0x + p$, and that of γ, δ is of the form $x^2 - 2x + q$.

Algebraic Equations

Example: Solve the equation $x^4 - 2x^3 + 4x^2 + 6x - 21 = 0$ given that the sum of two of its roots is zero

Let the roots be $\alpha, \beta, \gamma, \delta$ such that $\alpha + \beta = 0$

$\alpha + \beta + \gamma + \delta = 2 \Rightarrow \gamma + \delta = 2$

Thus, the quadratic factor corresponding to α, β is of the form $x^2 - 0x + p$, and that of γ, δ is of the form $x^2 - 2x + q$.

$x^4 - 2x^3 + 4x^2 + 6x - 21 = (x^2 + p)(x^2 - 2x + q) = x^4 - 2x^3 + (p + q)x^2 - 2px + pq$

Equating the coefficients: $p + q = 4$; $-2p = 6 \Rightarrow p = -3, q = 7$

Therefore, the equation is $(x^2 - 3)(x^2 - 2x + 7)$

Roots are x = $\pm\sqrt{3}, 1 \pm i\sqrt{6}$

Determinants

Definition: The expression $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ is called determinant of the second order and leads to $a_1b_2 - a_2b_1$. There are four numbers a_1, b_1, a_2, b_2 called elements which are arranged in two rows and two columns.

nth order determinant: $\begin{vmatrix} a_1 & b_1 & \cdots & l_1 \\ a_2 & b_2 & \cdots & l_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_n & b_n & \cdots & l_n \end{vmatrix}$ The diagonal through the left-hand top corner which contains the elements a_1, b_2, …, l_n is called the leading or principal diagonal

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

For example, in $\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$ minor of b_2 is $\begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix}$

Determinants

The **cofactor** of any element in a determinant is its minor with sign $(-1)^{i+j}$ where, i is the row number and j is the column number. It is usually denoted by the capital letter corresponding element.

Cofactor of b_2 is B_2 = $$(-1)^{2+2}\begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix}$$

Cofactor of c_2 is C_2 = $$(-1)^{2+3}\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} = -\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}$$

A determinant can be expanded in terms of any row (or column) by multiplying each element of the row (or column) by its cofactor and then add up all these terms.

For example, $$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$ Can be expanded in terms of row 1 as

$$\Delta = a_1A_1 + b_1B_1 + c_1C_1 = a_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}$$

Properties of Determinants

1) A determinant remains unaltered by changing its rows into column and columns into rows.

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \qquad \Delta' = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Then Δ = Δ'

2) If two adjacent parallel lines of a determinant are interchanged, the determinant retains its value but changes in sign.

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \qquad \Delta' = \begin{vmatrix} b_1 & a_1 & c_1 \\ b_2 & a_2 & c_2 \\ b_3 & a_3 & c_3 \end{vmatrix}$$

Then Δ = -Δ'

3) A determinant vanishes if two parallel lines are identical.

$$\Delta = \begin{vmatrix} a_1 & a_1 & c_1 \\ a_2 & a_2 & c_2 \\ a_3 & a_3 & c_3 \end{vmatrix} = 0$$

Properties of Determinants

4) If each element of a line be multiplied by the same factor, the whole determinant is multiplied by the factor.

$$\Delta = \begin{vmatrix} a_1 & b_1 & pc_1 \\ a_2 & b_2 & pc_2 \\ a_3 & b_3 & pc_3 \end{vmatrix} = p \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

5) If each element of a line consists of m terms, the determinant can be expressed as the sum of m determinants.

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 - d_1 + e_1 \\ a_2 & b_2 & c_2 - d_2 + e_2 \\ a_3 & b_3 & c_3 - d_3 + e_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} - \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} + \begin{vmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \\ a_3 & b_3 & e_3 \end{vmatrix}$$

Multiplication of Determinants:

$$\Delta_1 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}; \Delta_2 = \begin{vmatrix} c_1 & d_1 \\ c_2 & d_2 \end{vmatrix} \Rightarrow \Delta_1 \Delta_2 = \begin{vmatrix} a_1 c_1 + b_1 d_1 & a_1 c_2 + b_1 d_2 \\ a_2 c_1 + b_2 d_1 & a_2 c_2 + b_2 d_2 \end{vmatrix}$$

Matrices

Definition: A system of m × n numbers arranged in a rectangular formation along m rows and n columns and bounded by [ ] is called an m by n matrix. It is written as m × n matrix and is usually denoted by single capital letter.

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

A is a matrix of order mn. It has m rows and n columns where each ($a_{i,j}$) of the mn numbers are called as element of the matrix.

Types:

$$A = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

Column Matrix

$$A = \begin{bmatrix} 1 & 5 & 6 \end{bmatrix}$$

Row Matrix

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

Square Matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Diagonal Matrix

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

Unit Matrix

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

Symmetric matrix where, $a_{i,j} = a_{j,i}$

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

Skew Symmetric matrix where, $a_{i,j} = -a_{j,i}$

Matrices

Equality: Two matrices A and B are said to be equal if their order is same and each elements are equal.

Addition and Subtraction: Two matrices A and B are added or subtracted (if these are of same order) to give A + B and A – B where corresponding elements of A and B are added and subtracted, respectively.

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}; B = \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix} \Rightarrow A + B = \begin{bmatrix} a_1 + c_1 & b_1 + d_1 \\ a_2 + c_2 & b_2 + d_2 \end{bmatrix}$$

$$A - B = \begin{bmatrix} a_1 - c_1 & b_1 - d_1 \\ a_2 - c_2 & b_2 - d_2 \end{bmatrix}$$

Multiplication by scalar: For product of matrix A and scalar k, all elements of matrix A are multiplied by k

$$k \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} ka_1 & kb_1 \\ ka_2 & kb_2 \end{bmatrix}$$

Matrices

Matrix Multiplication: Two matrices can be multiplied only when the number of columns in the first is equal to the number of rows in the second.

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}_{3\times 2} \begin{bmatrix} c_1 & d_1 & e_1 \\ c_2 & d_2 & e_2 \end{bmatrix}_{2\times 3} = \begin{bmatrix} a_1c_1 + b_1c_2 & a_1d_1 + b_1d_2 & a_1e_1 + b_1e_2 \\ a_2c_1 + b_2c_2 & a_2d_1 + b_2d_2 & a_2e_1 + b_2e_2 \\ a_3c_1 + b_3c_2 & a_3d_1 + b_3d_2 & a_3e_1 + b_3e_2 \end{bmatrix}$$

For example:

$$A = \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}; B = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}; AB = \begin{bmatrix} 1\times 2 + 3\times 1 & 1\times 3 + 3\times 2 \\ -1\times 2 + 2\times 1 & -1\times 3 + 2\times 2 \end{bmatrix} = \begin{bmatrix} 5 & 9 \\ 0 & 1 \end{bmatrix}$$

$$BA = \begin{bmatrix} 2\times 1 + 3\times -1 & 2\times 3 + 3\times 2 \\ 1\times 1 + 2\times -1 & 1\times 3 + 2\times 2 \end{bmatrix} = \begin{bmatrix} -1 & 12 \\ -1 & 7 \end{bmatrix}$$

$$AB \neq BA$$

Matrices

Transpose of a Matrix: If the rows of a matrix A are changed to column or vice versa, the resultant matrix is called as transpose of A and is denoted by A'.

$$A=\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix} \Rightarrow A'=\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

Transpose of product of two matrices is the product of their transpose taken in the reverse order i.e., (AB)' = B'A'

Every square matrix can be uniquely expressed as a sum of a symmetric and skew-symmetric matrix.

$$A=\underbrace{\frac{1}{2}\left(A+A'\right)}_{\text{Symmetric matrix}}+\underbrace{\frac{1}{2}\left(A-A'\right)}_{\text{Skew Symmetric matrix}}$$

Matrices

Adjoint of a Square Matrix: Adjoint of A is the transposed matrix of cofactors of A

Determinant of the square matrix $A = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \Rightarrow \Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

The matrix formed by the cofactors of Δ is $\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{bmatrix}$

The transpose of this matrix is referred as adjoint of the matrix A and is written as Adj. A

$$Adj.A = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{bmatrix}$$

Matrices

Inverse of a Matrix: For any matrix A, if there exists a matrix B such that AB = BA = I, then B is an inverse of matrix A. Since AB and BA are defined and are equal, A and B should be square matrix of the same order. Also, $|AB| = |A||B| = |I| =$ 1, both $|A|$ and $|B|$ must be non-zero i.e., both the matrix and its inverse must be non-singular.

The inverse of a matrix A is denoted by A^{-1} so that $AA^{-1} = A^{-1}A = I$

$$A^{-1} = \frac{Adj.A}{|A|}$$

Proof:

$$A(Adj.A) = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{bmatrix} = \begin{bmatrix} |A| & 0 & 0 \\ 0 & |A| & 0 \\ 0 & 0 & |A| \end{bmatrix} = |A| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A(Adj.A) = |A| I \Rightarrow \frac{A(Adj.A)}{|A|} = I \Rightarrow \frac{(Adj.A)}{|A|} = A^{-1}$$

The reciprocal of the product of two matrices is the product of their reciprocals taken in the reverse order i.e. $(AB)^{-1} = B^{-1} A^{-1}$

Solution of Linear Equations

Cramer's Rule

Consider the linear equations

$$a_1x + b_1y + c_1z = d_1$$

$$a_2x + b_2y + c_2z = d_2$$

$$a_3x + b_3y + c_3z = d_3$$

Consider the determinant of the coefficients for above equations $\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

$$x\Delta = \begin{vmatrix} xa_1 & b_1 & c_1 \\ xa_2 & b_2 & c_2 \\ xa_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} xa_1 + yb_1 + zc_1 & b_1 & c_1 \\ xa_2 + yb_2 + zc_2 & b_2 & c_2 \\ xa_3 + yb_3 + zc_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}$$

Multiplying x in Δ

Operating C1 + yC2 + zC3

$$x = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

Similarly $y = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

Similarly, z can also be calculated

Solution of Linear Equations

Matrix Method

Consider the linear equations

$$a_1x+b_1y+c_1z=d_1$$

$$a_2x+b_2y+c_2z=d_2$$

$$a_3x+b_3y+c_3z=d_3$$

If $A=\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}; X=\begin{bmatrix} x \\ y \\ z \end{bmatrix}; D=\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$

Such that AX = D

Multiplying both sides by A^{-1}

$A^{-1}AX = A^{-1}D \Rightarrow IX = A^{-1}D \Rightarrow X = A^{-1}D$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{Adj.A}{|A|}D = \frac{1}{\Delta}\begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{bmatrix} \times \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Clearly $\Delta \neq 0$

Transpose of the cofactors of A

Rank of the matrix is the largest order of any non-zero minor of the matrix

→ Important to know the independent variables

Gauss –Jordan Method of Matrix Inverse

Elementary Transformation:

1)interchange of any two rows (columns)

2)multiplication of any row (column) by non-zero number

3)addition of constant multiple of the elements of any row (column) to the corresponding elements of any other row (column)

Step 3: Make c_3 = 1 and c_1 and c_2 = 0

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A^{-1} = \begin{bmatrix} e_1 & f_1 & g_1 \\ e_2 & f_2 & g_2 \\ e_3 & f_3 & g_3 \end{bmatrix}$$

Step 2: Make b_2 = 1 and b_1 and b_3 = 0

Step 1: Make a_1 = 1 and a_2 and a_3 = 0

Gauss –Jordan Method of Matrix Inverse

Example:

Operate $R_2 - R_1$ and $R_3 + 2R_1$

Operate $R_2/2$ and $R_3/2$

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & -3 \\ -2 & -4 & -4 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & -6 \\ 0 & -2 & 2 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

Operate $R_1 - R_2$ and $R_3 + R_2$

Operate $R_1 + 3R_3$, $R_2 - 3R_3/2$ and $-R_3/2$

$$\begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & -3 \\ 0 & -1 & 1 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1/2 & 0 \\ 1 & 0 & 1/2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & -2 \end{bmatrix} A^{-1} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

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

Consistency of Linear Equations

Consider a system of m linear equations

$$
\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = d_1 \\
&a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = d_2 \\
&\vdots \\
&a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = d_m
\end{aligned}
\qquad (1)
$$

$$
\Rightarrow A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}; D = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & d_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & d_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & d_m \end{bmatrix}
$$

Coefficient Matrix — Augmented Matrix

System of equations (1) is consistent if and only if the rank (r) of matrix A and that (r’) of D is same

1) If r = r’ = n, the equations are consistent and there is a unique solution present
2) If r ≠ r’, the equations are inconsistent and there is no solution present
3) If r = r’ < n, the equations are consistent and there are infinite solutions present

Consistency of Linear Homogeneous Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \qquad (1)$$

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

Coefficient Matrix

1) If r = n, the equations have only a trivial solution $x_1 = x_2 = \ldots = x_n = 0$
2) If r < n, the equations have (n – r) linearly independent solutions i.e., if arbitrary values are assigned to (n-r) of the variables, the values of remaining variables can be found uniquely.
3) If m < n (i.e., the number of equations is less than number variables), the solution is always other than $x_1 = x_2 = \ldots = x_n = 0$. The number of solutions is infinite.
4) If m = n (i.e., the number of equations is same as number of variables, the necessary and sufficient condition for solution other than $x_1 = x_2 = \ldots = x_n = 0$, is that the determinant of the coefficient matrix is zero. The solution is non-trivial.

Linear Transformations

$$Y = AX\,;Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix};\, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Linear Transformation

Coefficient Matrix

Orthogonal Transformation: The linear transformation i.e., Y = AX, is said to be orthogonal if it transforms $y_1^2 + y_2^2 + \ldots + y_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2$

$$X'X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2 \qquad \Rightarrow Y'Y = y_1^2 + y_2^2 + \cdots + y_n^2$$

$Y'Y = (AX)'(AX) = X'A'AX$ This is equals to $X'X$ if $A'A = I \Rightarrow A' = A^{-1}$

A is orthogonal matrix only if A’ = A^{-1}

Example Of Orthogonal Transformations

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

$$AA' = \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \\ 2/3 & 1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & 2/3 & 1/3 \\ -2/3 & 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Characteristic Equation

If A is any square matrix of order n, a matrix A – λI can be formulated. The determinant of this matrix equated to zero is called the characteristic equation of A.

$$|A-\lambda I| = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0$$

On expanding the determinant, the characteristic equation takes the form

$$(-1)^n \lambda^n + k_1\lambda^{n-1} + k_2\lambda^{n-2} + \cdots + k_n = 0$$

where, k_1, k_2, …, k_n are the expressible in terms of a_{ij}

$$= (-1)^n A^n + k_1 A^{n-1} + k_2 A^{n-2} + \cdots + k_n = 0$$

Cayley – Hamilton Theorem: Every square matrix satisfies its own characteristic equation.

The roots of this equation are called as **eigen-values** of matrix A.

Eigen Vectors

Linear Transformation

Coefficient Matrix

$$Y = AX\,;Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix};X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

In the above linear transformation, the column vector X is transformed to Y using square matrix A. In practice, often it is required to transform the such vector into themselves or scalar multiple of themselves. Let X be such vector which transform into λX.

$$\lambda X = AX \Rightarrow [A-\lambda I]X = 0 \quad \Rightarrow \begin{cases} (a_{11}-\lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + (a_{22}-\lambda)x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn}-\lambda)x_n = 0 \end{cases}$$

The above system of homogeneous equations will have non-trivial solution only if the coefficient matrix is singular, i.e., $|A-\lambda I|=0$. This equation has n roots and corresponding to each root the non-zero solution is $X = [x_1, x_2, \ldots, x_n]'$ is obtained. This solution is called as Eigen vector. For n values of λ, n Eigen vectors are obtained.

Properties of Eigen Values

1) The sum of the eigen values of a matrix is the sum of the elements of the principal diagonal.

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

$$|A - \lambda I| = -\lambda^3 + \lambda^2 (a_{11} + a_{22} + a_{33}) - \cdots$$

If λ_1, λ_2, λ_3 are the eigen values of a matrix A

$$|A - \lambda I| = (-1)^3 (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) = -\lambda^3 + \lambda^2 (\lambda_1 + \lambda_2 + \lambda_3) - \cdots + \lambda_1 \lambda_2 \lambda_3 \quad (1)$$

$$\Rightarrow \lambda_1 + \lambda_2 + \lambda_3 = a_{11} + a_{22} + a_{33}$$

2) The product of Eigen values of a matrix A is equal to its determinant i.e., $\lambda_1 \lambda_2 \ldots \lambda_n = |A|$ [this can be obtained by putting $\lambda = 0$ in (1)].
3) If λ is an eigen value of a matrix A, then $1/\lambda$ is the eigen value of A^{-1}.
4) If λ is an eigen value of an orthogonal matrix, then $1/\lambda$ is also its eigen value.
5) If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigen values of a matrix A, then A^m has the eigen values $\lambda_1^m, \lambda_2^m, \ldots, \lambda_n^m$ (m is the positive integer).

Probability

1) Very often we are not sure about certain outcome but wish to assess the chances of our predictions coming true. The study of probability provides a mathematical frame-work for such assertions and is essential in every decision-making process.
2) A set of events is said to be **exhaustive**, if it includes all the possible events. For instance, while tossing the coin either we get head or tail. There is no third option.
3) If the occurrence of one of the events precludes the occurrences of all others, then such a set of events is said to be **mutually exclusive**. While tossing the coin either head comes or tail. But both cannot occur simultaneously therefore these are two mutually exclusive events.
4) If one of the events cannot be expected to happen in preference to another then such events are said to be **equally likely**. For example, while tossing the coin both the outcome head and tail are equally likely.
5) **Definition:** If there are n exhaustive, mutually exclusive and equally likely cases of which m are favourable to an event A, then probability of the happening of A is P(A) = m/n.
6) The number of **Permutation** of n different things taken r at a time is ${}^nP_r = n!/[(n-r)!] = [n(n-1)(n-2)\dots(n-r+1)]$. Order is important (i.e., AB and BA are different).
7) The number of **combination** of n different things taken r at a time is ${}^nC_r = n!/[r!(n-r)!] = [n(n-1)(n-2)\dots(n-r+1)]/r!$. Order is not important (i.e., AB and BA are same).

Probability

Example: A bag contains 50 tickets numbered 1, 2, 3, …, 50, of which four are drawn at random and arranged in ascending order ($t1 < t2 < t3 < t4$). Find the probability of t3 being 25?

The exhaustive number of cases = ${}^{50}C_4$

If t3 = 25, then t1 and t2 must come out of 24 tickets numbered 1 to 24. This can be done in ${}^{24}C_2$ ways.
Further, t4 must come out from the 25 tickets numbering 26 to 50. This can be done in ${}^{25}C_1$ ways.

Favourable number of cases = ${}^{24}C_2 \times {}^{25}C_1$

Probability of t3 being 25 = $[{}^{24}C_2 \times {}^{25}C_1] / {}^{50}C_4 = 276 \times 25 / 230300 = 69/2303$

Probability

Addition Law of Probability: If the probability of an event happening as a result of a trial is P(A) and the probability of a mutually exclusive event B happening is P(B), then the probability of either of the events happening as a result of trial is P(A + B) or $P(A \cup B) = P(A) + P(B)$.

Example: A bag contains 8 white and 6 red balls. Find the probability of drawing two balls of the same colour.

Two balls can be drawn from 14 with = ${}^{14}C_2$ ways.

Probability of drawing two white balls = ${}^{8}C_2 / {}^{14}C_2 = 28/91$

Probability of drawing two red balls = ${}^{6}C_2 / {}^{14}C_2 = 15/91$

Probability of drawing two balls of same colour = 28/ 91 + 15/ 91 = 43/91

Probability

Multiplication Law of Probability: If the probability of an event happening as a result of a trial is P(A) and after A has happened, the probability of an event B happening as a result of another trial (i.e., conditional probability of B for given A) is P(B/A), then the probability of both the events A and B happening as a result of two trials is P(AB) or $P(A \cap B) = P(A)\,P(B/A)$.

Example: A box A contains 2 white and 4 black balls. Another box B contains 5 white and 7 black balls. A ball is transferred from the box A to the box B. Then the ball is drawn from the box B. Find the probability that it is white.

The probability of drawing a white ball from box B will depend on whether the transferred ball is white or black.

Case 1: If the black ball is transferred from A, then its probability is 4/6

With this, the box B now has 5 white and 8 black balls. Thus, probability of taking white balls is = 5/13

Therefore, the probability for this case for taking a white ball is = $(4/6)\times(5/13) = 10/39$

Case 2: If the white ball is transferred from A, then its probability is 2/6

With this, the box B now has 6 white and 7 black balls. Thus, probability of taking white balls is = 6/13

Therefore, the probability for this case for taking a white ball is = $(2/6)\times(6/13) = 2/13$

Probability for taking a white ball considering both cases = $(10/39) + (2/13) = 16/39$

Probability

Baye's Theorem: An event A corresponds to a number of exhaustive events B_1, B_2, …, B_n. If $P(B_i)$ and $P(A/B_i)$ are given, then

$$P(B_i / A) = \frac{P(B_i)P(A/B_i)}{\sum P(B_i)P(A/B_i)}$$

Proof:

$$P(A \cap B_i) = P(A)P(B_i / A) = P(B_i)P(A/B_i) = P(B_i \cap A)$$

$$P(B_i / A) = \frac{P(B_i)P(A/B_i)}{P(A)}$$

Since the event A corresponds to events B_1, B_2, …, B_n.

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

$$= \sum P(A \cap B_i) = \sum P(B_i)P(A/B_i) \Rightarrow P(B_i / A) = \frac{P(B_i)P(A/B_i)}{\sum P(B_i)P(A/B_i)}$$

Probability

Example: Three machines U1, U2 and U3 produces same items. Of their respective output 5 %, 4 % and 3 % items are faulty. On a given day, U1 has produced 25 % of the total output, U2 has produced 30 % output and U3 produced the remaining. An item selected at random is found to be faulty. What are the chances that it was produced by U3?

	U1	**U2**	**U3**	**Comment**
$P(B_i)$	0.25	0.30	0.45	Total = 1
$P(A/B_i)$	0.05	0.04	0.03	
$P(B_i)\ P(A/B_i)$	0.0125	0.012	0.0135	Total = 0.038
$P(B_i\ /A)$	0.0125/0.038	0.012/0.038	0.0135/0.038	

Probability of U3 producing the faulty item on that day = 0.0135/0.038 = 0.355

Statistics

Statistics deals with the method for collection, classification, and analysis of numerical data for drawing valid conclusions and making reasonable decisions.

1) The collection of data constitutes the starting point of any statistical investigation.
2) Data may be collected for each and every unit of the whole population. This ensures the higher accuracy.
3) But complete enumeration is prohibitively expensive and time consuming. Therefore, instead of taking very large number of data points, **a few of them** are selected and conclusions drawn on the basis that these samples hold for the entire population.
4) The data collected in the course of an inquiry is not in an easily assimilable form and its proper classification is necessary for making intelligent inferences. The classification is done by dividing the raw data into a convenient number of groups according to the values of the variable and finding the frequency of the variable in each group i.e., grouping the marks of students in the range of 0 – 4, 5 – 9, 10 – 14, and so on. Each of these groups are called **classes** and the number in that interval is called **frequency**.
5) A set of raw data summarised by distributing it into number of classes along with their frequencies is known as **frequency distribution**.
6) In some investigation, the number of items less than particular value is needed. For this, the frequencies of the classes are added up to that value which is called as **cumulative frequency**.

Statistics

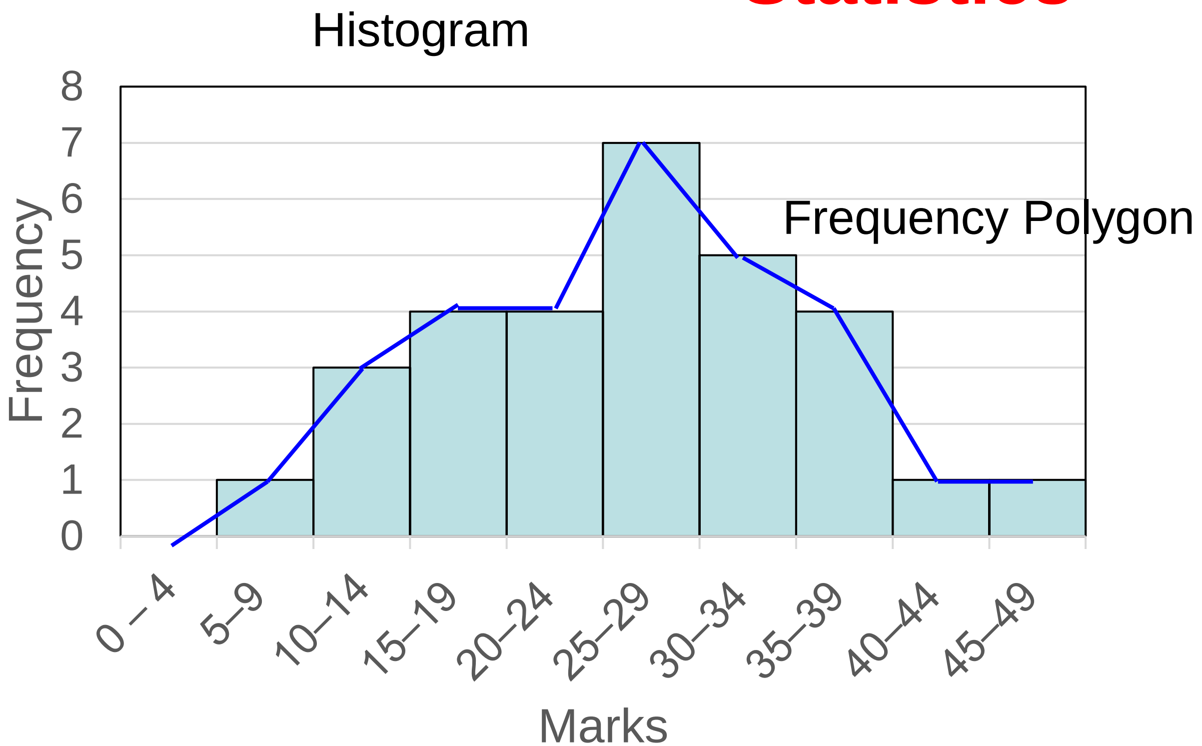
Raw data of marks obtained out of 50 by 30 students

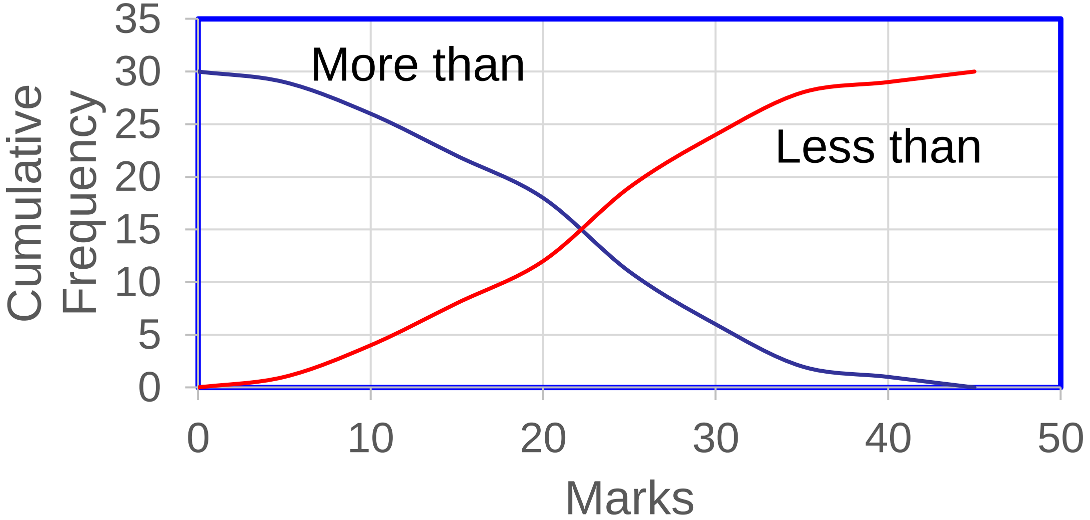
29	30	25	39	31	18
34	12	32	25	07	24
15	28	23	32	47	41
28	38	23	37	11	22
17	25	17	25	10	38

Classified Data

Class	Frequency	Cumulative Frequency
0 – 4	0	0
5 – 9	1	1
10 – 14	3	4
15 – 19	4	8
20 – 24	4	12
25 – 29	7	19
30 – 34	5	24
35 – 39	4	28
40 – 44	1	29
45 – 49	1	30

Statistics


Histogram
Frequency Polygon
Frequency
8
7
6
5
4
3
2
1
0
0 – 4
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
Marks


More than
Less than
Cumulative Frequency
35
30
25
20
15
10
5
0
0
10
20
30
40
50
Marks

Statistics

Mean: If x_1, x_2, …, x_n are a set of *n* values of a variate, then the arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

In a frequency distribution, if x_1, x_2, …, x_n be the mid values of the class intervals having frequencies f_1, f_2, …, f_n respectively then mean is given by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n}$$

Median: If the values of a variable are arranged in the ascending order of magnitude, the median is the middle item if the number is odd and is the mean of the two middle items if the number is even. Thus, the median is the mid value.

Statistics

Mode: The mode is defined as that value of the variable which occurs most frequently. i.e., the value of maximum frequency

$$Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$$

Where, L is lower limit of the class containing mode, Δ_1 is the excess of modal frequency over frequency of preceding class, Δ_2 excess modal frequency over following class, and h is the width of the modal class.

Geometric and Harmonic Mean: If x_1, x_2, …, x_n are a set of *n* observations, then the geometric mean (GM) and harmonic mean (HM) are given by

$$GM = \left(x_1 x_2 \cdots x_n\right)^{1/n}$$

$$HM = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)}$$

In a frequency distribution, if x_1, x_2, …, x_n be the mid values of the class intervals having frequencies f_1, f_2, …, f_n respectively then mean is given by

$$GM = \left[\left(x_1\right)^{f_1}\left(x_2\right)^{f_2} \cdots \left(x_n\right)^{f_n}\right]^{1/n}; n = \sum f_i$$

$$HM = \frac{1}{\frac{1}{n}\left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}\right)}$$

Thank You