Date: 14-03-2025

Module 1 Report

Foundational Mathematics & AI Fundamentals

By,

Vaishnavi R Saralaya

Sabha Ambrin

Foundational Mathematics & AI Fundamentals

1. Introduction

Artificial Intelligence (AI) is a multidisciplinary field that relies heavily on mathematical concepts to build models, analyze data, and make predictions. This report explores the foundational mathematical topics essential for AI, including linear algebra, calculus, probability, and statistics. Each topic is explained with theoretical examples to provide a clear understanding of its relevance and application in AI.

2. Linear Algebra

Linear Algebra is the branch of mathematics aimed at solving systems of linear equations with a finite number of unknowns. Linear algebra is essential for many machine learning algorithms and techniques. It helps in manipulating and processing data, which is often represented as vectors and matrices. These mathematical tools make computations faster and reveal patterns within the data.

It simplifies complex tasks like data transformation, dimensionality reduction (e.g., PCA), and optimization. Key concepts like matrix multiplication, eigenvalues, and linear transformations help in training models and improving predictions efficiently.

Fundamental Concepts in Linear Algebra for Machine Learning

In machine learning, vectors, matrices, and scalars play key roles in handling and processing data.

- Vectors are used to represent individual data points, where each number in the vector corresponds to a specific features of the dataset (like age, income, or hours).
- Matrices are considered as data storage units used to store large datasets, with rows representing different data points and columns representing features.
- Scalars are single numbers that scale vectors or matrices, often used in algorithms like gradient descent to adjust the weights or learning rate, helping the model improve over time.

3.Matrix operations

A matri*x* is a rectangular array of numbers that is usually named by a capital letter: A, B, C ,and so on. Each entry in a matrix is referred to as aij, such that i represents the row and j represents the column. Matrices are often referred to by their dimensions: m×n indicating m rows and n columns.

Finding the Sum and Difference of Two Matrices

To solve a problem like the one described for the soccer teams, we can use a matrix, which is a rectangular array of numbers. A row in a matrix is a set of numbers that are aligned horizontally. A column in a matrix is a set of numbers that are aligned vertically. Each number is an entry, sometimes called an element, of the matrix. Matrices (plural) are enclosed in [ ] or ( ), and are usually named with capital letters. For example, three matrices named AA, BB, and CC are shown below.

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

$$B = \begin{bmatrix} 1 & 2 & 7 \\ 0 & -5 & 6 \\ 7 & 8 & 2 \end{bmatrix}$$

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

A matrix is often referred to by its size or dimensions: m×n indicating mm rows and nn columns. Matrix entries are defined first by row and then by column. For example, to locate the entry in matrix AA identified as aij, we look for the entry in row i, column j. In matrix A, shown below, the entry in row 2, column 3 is a23.

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

- A **square matrix** is a matrix with dimensions n×n, meaning that it has the same number of rows as columns. The 3×3 matrix above is an example of a square matrix.
- A **row matrix** is a matrix consisting of one row with dimensions 1×n.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}$$

- A **column matrix** is a matrix consisting of one column with dimensions m×1.

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}$$

Scalar Multiplication

Scalar multiplication involves multiplying each element of a vector or matrix by a scalar.

Dot Product (Scalar Product)

The dot product of two vectors tells us how similar their directions are. To calculate it, you multiply the matching elements of the vectors and then add them together.

Example: For example, given two vectors

(u=[u1,u2,u3] and v=[v1,v2,v3]) their dot product is calculated as:

$$u \cdot v = u1 \cdot v1 + u2 \cdot v2 + u3 \cdot v3$$

Cross Product (Vector product)

The **cross product** of two 3D vectors makes a new vector that points **at a right angle** to the two original vectors. It is used less frequently in machine learning compared to the dot product.

Example: Given two vectors **u** and **v,** their cross product **u×v** is calculated as:

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}$$

Key Concepts

- A matrix is a rectangular array of numbers. Entries are arranged in rows and columns.
- The dimensions of a matrix refer to the number of rows and the number of columns. A 3×2 matrix has three rows and two columns.
- We add and subtract matrices of equal dimensions by adding and subtracting corresponding entries of each matrix.
- Scalar multiplication involves multiplying each entry in a matrix by a constant.

- Scalar multiplication is often required before addition or subtraction can occur.
- Multiplying matrices is possible when inner dimensions are the same—the number of columns in the first matrix must match the number of rows in the second.
- The product of two matrices, A and B,is obtained by multiplying each entry in row 1 of A by each entry in column 1 of B; then multiply each entry of row 1 of A by each entry in column 2 of B, and so on.
- Many real-world problems can often be solved using matrices.
- We can use a calculator to perform matrix operations after saving each matrix as a matrix variable.

4. Eigenvalues and Eigenvectors

Eigenvalues are scalar values associated with a square matrix that measure how a matrix transforms a vector. If a matrix A multiplies a vector v, and the result is a scalar multiple of v, then that scalar is the eigenvalue corresponding to the eigenvector v. Eigenvalues are widely used in fields like physics, engineering, and data science.

Eigenvectors for square matrices are defined as non-zero vector values which when multiplied by the square matrices give the scaler multiple of the vector, i.e. we define an eigenvector for matrix A to be "v" if it specifies the condition, $Av = \lambda v$

The equation for eigenvalue is given by

$$\mathbf{Av = \lambda v}$$

Where,

- A is the matrix,
- v is associated eigenvector, and
- λ is scalar eigenvalue.

The scaler multiple λ in the above case is called the eigenvalue of the square matrix. We always have to find the eigenvalues of the square matrix first before finding the eigenvectors of the matrix.

5. Calculus | Differential and Integral Calculus

Calculus was founded by Newton and Leibniz. Calculus is a branch of mathematics that helps us study change. It is used to understand how things change over time or how quantities grow, shrink, or accumulate. There are two main parts of calculus**:**

- Differential Calculus : It helps us calculate the rate of change of one quantity concerning another. This rate of change is called the derivative.
    - Example: Finding how fast a balloon inflates as you pump air into it.
    - Calculating the slope of a hill (steepness).
- Integral Calculus : helps us calculate the total accumulation of change. This accumulation is called the integral.
    - Example: Calculating the area under a curve (e.g., finding the distance travelled by a car when you know its speed at every moment).
    - Determining the total rainfall collected in a reservoir.

Note: The process of finding the value of a derivative is called differentiation, and the process of finding the value of an integral is called integration.

6. Probability

Probability defines the likelihood of occurrence of an event.

Probability can be defined as the ratio of the number of favourable outcomes to the total number of outcomes of an event.

For an experiment having 'n' number of outcomes, the number of favourable outcomes can be denoted by x. The formula to calculate the probability of an event is as follows.

Probability(Event) = Favourable Outcomes/Total Outcomes = x/n

Probability of an Event

The probability of an event E, denoted by P(E), is a number between 0 and 1 that represents the likelihood of E occurring.

- If $P(E) = 0$, the event E is impossible.
- If $P(E) = 1$, the event E is certain to occur.
- If $0 < P(E) < 1$, the event E is possible but not guaranteed.

Experiment: A trial or an operation conducted to produce an outcome is called an experiment.

Sample Space: All the possible outcomes of an experiment together constitute a smaple. For example, the sample space of tossing a coin is {head, tail}.

Favourable Outcome: An event that has produced the desired result or expected event is called a favourable outcome. For example, when we roll two dice, the possible/favourable outcomes of getting the sum of numbers on the two dice as 4 are (1,3), (2,2), and (3,1).

Events: In Probability, an event is an outcome or set of outcomes resulting from an experiment.

<u>Types of events:</u>

- **Mutually Exclusive events:** Two events A and B are said to be mutually exclusive if they cannot occur simultaneously. Note If A and B represent mutually exclusive events then they are disjoint, that is AB where is the null set.

Example 1. when we toss a coin, either head or tail can be up, but both cannot be up at a time.

3. when we throw a dice the outcomes getting 1, 2, 3 . . .. 6 are mutually exclusive events.

- **Equally likely events:** The events are said to be equally likely if none of them is expected to occur in preference to the other. i.e. each one of them has an equal chance of happening.

Example: In tossing of a coin, getting a head and tail are equally likely events.

- **Exhaustive Events:** Outcomes are said to be Exhaustive when they include all possible outcomes.

Example: In the case of throwing two dice the exhaustive number of cases is 36.

- **Independent Event:** Two or more events are considered to be independent if the occurrence or non-occurrence of an event does not affect the occurrence or non-occurrence of the other.

Example: In successive tossing of a coin, the event of getting a head or tail in the first toss does not affect the event of getting a head or tail in the second toss.

- **Dependent Events:** The events are said to be dependent if the occurrence or non-occurrence of one event in any trial affects the occurrence of other events in other trails.

Mathematical classical or a priority definition of probability:

If a trail results in n exhaustive mutually exclusively and equally likely cases and m of them are favourable to the happening of the event A, then the probability of happening of A is given by,

$$P(A) = p = \frac{\text{number of favaourable cases}}{\text{Total number of exhaustive cases}} = \frac{n(A)}{n(S)}:$$

Axioms of Probability:

Let us say S is the sample space and E is an event then,

1. $0 \le P(E) \le 1$
2. $P(S) = 1$
3. Let us consider 2 events A and B then

   $P(AUB) = P(A) + P(B) - P(A \cap B)$.

   In this case id A and B are mutually exclusive then $P(A \cap B) = 0$,

   Then $P(AUB) = P(A) + P(B)$.

   For any number of mutually exclusive events,

   $P(E_1UE_2UE_3U, \ldots .. E_n) = P(E_1)+P(E_2)+P(E_3)+ \ldots \ldots +P(E_n)$
4. If A ,B and C are three events then,

   $P(AUBUC) = P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)-P(C \cap B)+P(A \cap B \cap C)$
5. For A and B are events of random experiment then,

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

   $P(A \cap B) = P(B)*P(A/B)$. This is known as Multiplication Rule given $P(A) \neq 0$ and $P(B) \neq 0$.

   If A and B are independent then,

   $P(A \cap B) = P(A) *P(B)$.
6. $P(A) = 1- P(\bar{A})$

Concepts of Probability are used in various real life scenarios :

- Stock Market : Investors and analysts often study these parameters and use probabilistic models to understand trends and patterns for the movement of stock price.
- Insurance: Insurance companies use probability models to estimate the likelihood of various events to manage this risk, and set premiums accordingly.
- Weather Forecasting : Meteorologists use probability to predict the likelihood of various weather events, such as rain, snow, storms, or temperature changes.

Formula for probability

Probability of Event P(E) = [Number of Favorable Outcomes] / [Total Number of Outcomes]

Example 1: There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of randomly picking a yellow pillow?

Solution:

Probability is equal to the number of yellow pillows in the bed divided by the total number of pillows, i.e.

2/6 = 1/3

Conditional probability

If the probability of the event A provided the event B has already occurred is called the conditional probability and is defined as

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided} \quad P(B) \neq 0$$

If the probability of the event B provided the event A has already occurred is given by

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{provided} \quad P(A) \neq 0$$

Note

i) if the events A and B are independent then

$$P(A/B) = \frac{P(A).P(B)}{P(B)} = P(A)$$

$$P(B/A) = \frac{P(A).P(B)}{P(A)} = P(B)$$

ii) If A and B are mutually exclusive events then

$$P(B/A) = 0 \text{ and } P(A/B) = 0 \text{ since } P(A \cap B) = 0$$

Multiplication law of probability

If A and B are dependent events then

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

Total Probability Theorem

If $B_1, B_2, \ldots, B_n$ are mutually exclusive and exhaustive set of events of a sample space S and A is any event associated with the events $B_1, B_2, \ldots, B_n$ then

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \ldots + P(B_n)P(A/B_n).$$

i.e., $$P(A) = \sum_{i=1}^{n} P(B_i)P(A/B_i).$$

Baye's Theorem

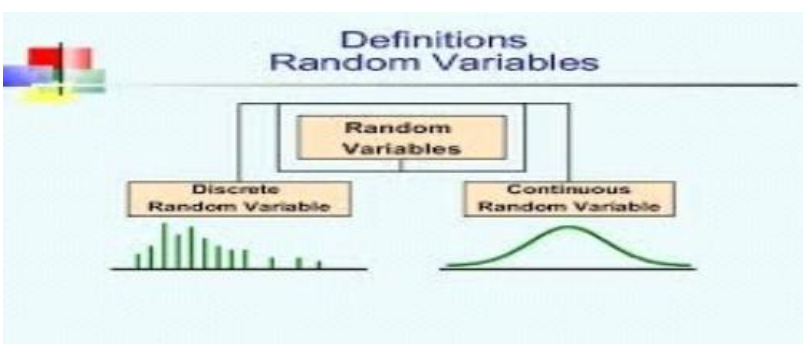
If $B_1, B_2, \ldots, B_n$ are a set of exhaustive and mutually exclusive events of a sample space S and A is any event associated with $B_1, B_2, \ldots, B_n$ such that

$$A \subseteq {}_{i=1}^{\;n} B_i$$ then $$P(B_i / A) = \frac{P(B_i)P(A/B_i)}{\sum_{i=1}^{n} P(B_i)P(A/B_i)}$$

Random variables:

A random variable represents the outcome of statistical experiments on numerical values. There are two types of Random Variable:

- Discrete Random Variable
- Continuous Random Variable



Discrete Random Variable

A random variable X which can assume countable number of isolated values (integers) is called discrete random variable. A random variable X which contains finite number of discrete random values in an interval is called discrete is called discrete random variable. The meaning of the word 'discrete' is separate and individual. Discrete random variables are typically represented by integers or whole numbers, and their probability distribution is a discrete **probability mass function.** Then probability mass function is given by,

- $f(x) \geq 0$ for each real number x.

Note: Any function satisfying the above 2 conditions is said to be discrete probability function or probability mass function.

Example: Consider we toss 3 coins, let X be the random variable given as number of heads then,

X=x	0	1	2	3
f(x)=P(X=x)	1/8	3/8	3/8	1/8

<u>Examples</u>:

- Number of kids in a family
- Number of attempts to hit a target
- Number of heads in tossing a coin thrice
- Number of stars in the sky

Cumulative Distribution Function (CDF):
The **Cumulative Distribution Function**, of a real-valued random variable X, evaluated at x, is the probability function that X will take a value less than or equal to x.

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

- The CDF defined for a discrete random variable and is given as,

$F_X(x) = P(X <= x)$

- The CDF defined for a continuous random variable is given as;

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Expected value for discrete random variables:
If X is a discrete random variable with possible values x1, x2…, xi, …
and probability mass function p(x),
Then the expected value (or mean) of X is denoted E[X]
and given by

$E[X] = \sum x.p(x) =$ (Arithmetic mean)

Then the expected value of the previous example is given by E(x)

Continuous Random Variable:

A random variable X which can assume all real values within a given interval is known as continuous random variable. Thus, the possible values of a continuous random variable are uncountably infinite. Continuous random variables are typically represented by real numbers, and their probability distribution is a continuous **probability density function.** Examples:

- Price of a commodity
- Height of a person in cm
- Life of an electric component (in hours)

Probability distributions:

1. Bernoulli distribution
2. Binomial distribution
3. Poisson distribution
4. Geometric distribution
5. Normal distribution
6. Exponential distribution

a. Bernoulli distribution:

A random experiment with 2 outcomes or binary outcomes is called the Bernoulli trails. Bernoulli trails associated with a random experiment have the following properties:

- Outcomes are denoted as successes or failures 2.
- The probability of success is denoted as p and the probability of failure is denoted $q = 1 - p$.
- The trails are repeated independently and the probability of success, p, remains the same in all trails.

The random variable associated with Bernoulli trails is called the Bernoulli random variable. The probability distribution of the Bernoulli random variable is called the Bernoulli distribution.

Example:

• A student will pass or fail an exam.

• Rolled dice will show either 6 or any other number.

Probability Mass Function of Bernoulli distribution

$$P(X = x) = p^x(1-p)^{1-x}, \qquad x = 0,1$$

p = probability of success and q = 1- p = probability of failure

b. Binomial distributions:

If the Bernoulli trial is repeated n times, then the random variable associated with this experiment is called the binomial random variable.
Binomial distribution is simply defined as the probability of success or failure outcome in an experiment or survey that is repeated multiple times (say n).
Example: Tossing a coin 5 times.

Probability Mass Function of binomial distribution:

$$P(X = x) = \binom{n}{x} p^x(1-p)^{1-x}, \qquad x = 0,1 \ldots n \text{ and } q = 1-p$$

n = number of times Bernoulli trial repeated
p = Probability of success, q = 1-p = probability of failure

c. Poisson distribution:

Poisson distribution definition is used to model a discrete probability of an event where independent events are occurring in a fixed interval of time and have a known constant mean rate. In other words, Poisson distribution is used to estimate how many times an event is likely to occur within the given period of time. λ is the Poisson rate parameter that indicates the expected value of the average number of events in the fixed time interval.

Probability mass function for Poisson distribution is:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0,1,2,3,\ldots \quad \text{where} \lambda > 0$$
$$= 0, \text{otherwise}$$

then X is said to follow Poisson distribution with the parameter.

d. Geometric Distribution:

In a series of trials, if you assume that the probability of either success or failure of a random variable in each trial is the same, geometric distribution gives the probability of achieving success after N number of failures.

geometric distribution defines the probability that first success occurs after k number of trials. If p is the probability of success or failure of each trial, then the probability that success occurs on the k^{th} trial is given by the formula,

$$Pr(X = k) = (1 - p)^{k-1}p$$

Continuous Distribution:

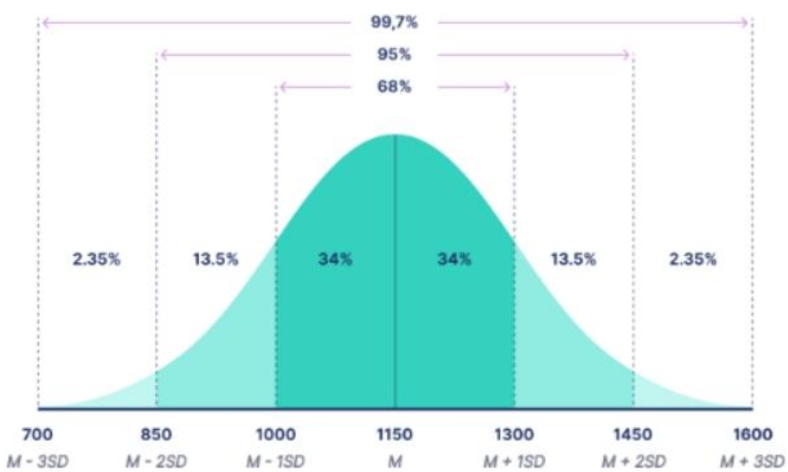
a. **Normal distribution**:

A continuous random variable X is said to have normal distribution if its probability density function is given by ,

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

$f(x)$ = probability density function
σ = standard deviation
μ = mean

Z-score Formula	Explanation
$z = \frac{x - \mu}{\sigma}$	• x = individual value • μ = mean • σ = standard deviation



b. **Exponential Distribution:**

A continuous random variable X is said to have an *exponential* distribution with parameter $\lambda > 0$, shown as $X \sim Exponential(\lambda)$, if its PDF is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

7. Statistics

Statistics is the science of collecting, organizing, analyzing, and interpreting information to uncover patterns, trends, and insights.

Types of data

Various types of Data used in statistics are,

- Qualitative Data –qualitative data is the descriptive data of any object. For example, Kabir is tall, Kaira is thin, etc.
- Quantitative Data –quantitative datais the numerical data of any object. For example, he ate three chapatis, and we are five friends.

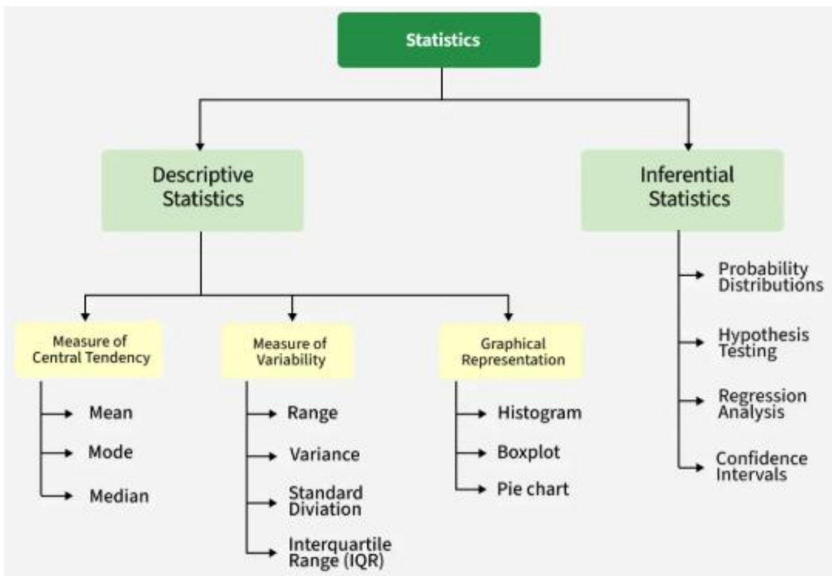
Quantitative data

We have two types of quantitative data that include,

1. Discreate Data: The data that have fixed value is called discreate data, discreate data can easily be counted.
2. Continuous Data: The data that has no fixed value and has a range of data is called continuous data. It can be measured.

- Statistics allows us to see the bigger picture and tackle real-world problems like measuring the popularity of a new product, predicting the weather, or tracking health trends with clarity and confidence by using numbers and data.

- It helps in making informed decisions by transforming raw data into meaningful conclusions, reducing uncertainty, and providing a solid foundation for research, business strategies, and policy-making.
- From sports analytics and financial forecasting to medical research and artificial intelligence, statistics plays a crucial role in nearly every field, enabling us to optimize processes, test hypotheses, and make data-driven predictions.



Introduction to Statistics

Descriptive Statistics

Descriptive statistics uses data that provides a description of the population either through numerical calculated graphs or tables. It provides a graphical summary of data.

It is simply used for summarizing objects, etc. There are two categories in this as follows.

- Measure of Central Tendency
- Measure of Variability

Measure of Central Tendency

Measure of central tendency is also known as summary statistics that are used to represent the center point or a particular value of a data set or sample set. In statistics, there are three common measures of central tendency that are:

- Mean
- Median

- Mode

Mean

The arithmetic average of a set of values. It's calculated by adding up all the values in the data set and dividing by the number of values. Formula= $\bar{x}=\sum x/n$

Example: Consider the data set {2, 4, 6, 8, 10}.
(2+4+6+8+10/ 5) = (30/5)
Mean = 6

Arithmetic mean: Arithmetic mean (or, simply, "mean") is nothing but the average. It is computed by adding all the values in the data set divided by the number of observations in it. If we have the raw data, mean is given by the formula

Mean $\bar{X} = \frac{\Sigma X}{n}$

Mean $\bar{X} = \frac{\Sigma f X}{n}$

Where, *f* is the frequency and *X* is the midpoint of the class interval and n is the number of observations.

GEOMETRIC MEAN:

The Geometric Mean (GM) is the average value or mean which signifies the central tendency of the set of numbers by taking the root of the product of their values. Basically, we multiply the 'n' values altogether and take out the nth root of the numbers, where n is the total number of values.

For example: for a given set of two numbers such as 8 and 1, the geometric mean is equal to $\sqrt{8 * 1}$ = √8 = 2√2.

Thus, the geometric mean is also defined as the nth root of the product of n numbers.

Note that Geometric mean (for two values) = $\sqrt{a.b}$

Geometric Mean Formula

The Geometric Mean (G.M) of a data set containing n observations is the n^{th} root of the product of the values. Consider, if $x_1, x_2, ..., x_n$ are the observations, for which we aim to calculate the geometric mean. The formula to calculate the geometric mean is given below:

Geometric mean or Geometric Average Formula

$$\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

n : number of terms that are multiplied

It is also represented as:

$GM = (x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{1/n}$

Taking logarithm on both sides,

$\log GM = \log (x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{1/n}$

$\log GM = (1/n) \log (x_1 \cdot x_2 \cdot \ldots \cdot x_n)$

$\log GM = (1/n) [\log x_1 + \log x_2 + \ldots + \log x_n]$

$\log GM = (\sum \log x_i) / n$

Therefore, geometric mean**,**

$GM = \text{Antilog} ((\sum \log x_i) / n)$ for ungrouped data

$GM = \text{Antilog} ((\sum f_i \log x_i) / n)$ for grouped data

Harmonic Mean:

The harmonic mean is a numerical average calculated by dividing the number of observations, or entries in the series, by the reciprocal of each number in the series. Thus, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

$$\text{H. M.} = \frac{n}{\sum\left(\frac{1}{x_i}\right)}$$

For raw data.

$$\text{H. M.} = \frac{N}{\sum_{i=1}^{n} f_i\left(\frac{1}{x_i}\right)}$$

For grouped data.

Relation Between AM, GM and HM:

Consider two numbers 'a' and 'b' such that a and b are greater than 0. Terms in the sequence are 'a' and 'b' and the number of terms in the sequence 'n = 2'. If AM GM HM formula is used, AM GM HM can be found as follows.

$$AM = \frac{(a+b)}{2}$$

$$HM = \frac{2}{\frac{1}{a}+\frac{1}{b}} = \frac{2}{\frac{b+a}{ab}} = \frac{2ab}{a+b}$$

$$GM = \sqrt[2]{ab}$$

The above equation gives the relation between AM,GM and HM. The equation can also be written as,

$$AM \times HM = GM^2$$

Combined mean:

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Median:

The middle value of a data set when it is arranged in ascending or descending order. If there's an odd number of observations, it's the value at the center position. If there's an even number, it's the average of the two middle values**.**

Example: Data set {3, 6, 9, 12, 15}. Median = 9

Median Formula for Ungrouped Data:

Median for Odd Data

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

Median for Even Data

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

Median Formula for grouped Data:

Median for Grouped Data

$$\text{Median} = l + \left[\frac{\frac{n}{2} - c}{f}\right] \times h$$

Here, l=lower frequency of the median

f = frequency of the median class

h = size of the median class

c = cumulative frequency of the class preceding the median class

n = Total number of observations, $\sum f_i = n$.

Mode

The value that occurs most frequently in a data set. There can be one mode (unimodal), multiple modes (multimodal), or no mode if all values occur with the same frequency.

Example: Data set {2, 3, 4, 4, 5, 6, 6, 6, 7}. Mode = 6

Mode formula for grouped data is given as,

$$\text{Mode} = \mathbf{L} + \left(\frac{\mathbf{f_1 - f_0}}{\mathbf{2f_1 - f_0 - f_2}}\right) \cdot \mathbf{h}$$

Where,

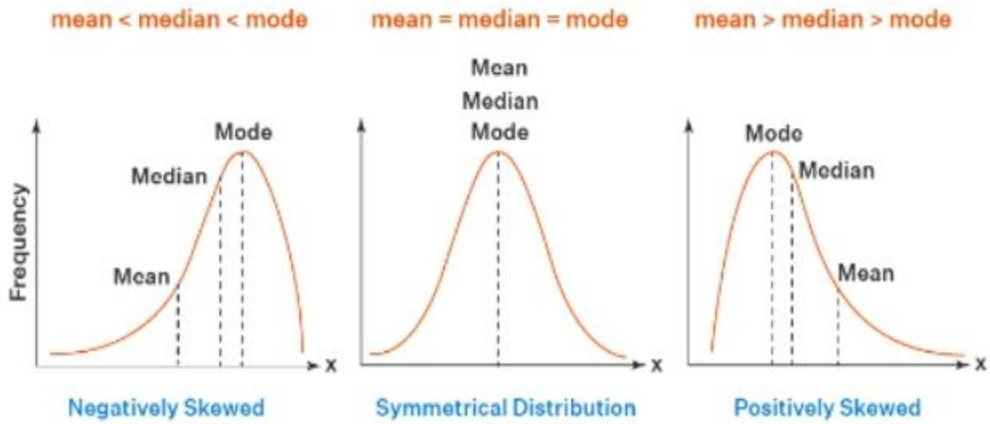
- ***L*** *is the lower limit of the modal class,*
- ***f_1*** *is the Frequency of the modal class,*
- ***f_2*** *is the Frequency of the class succeeding the modal class,*
- ***f_0*** *is the Frequency of the class preceding the modal class, and*
- ***h*** *is the size of the class interval.*

Relation between mean median and mode:

Mean - Mode = 3(Mean - Median)

Or

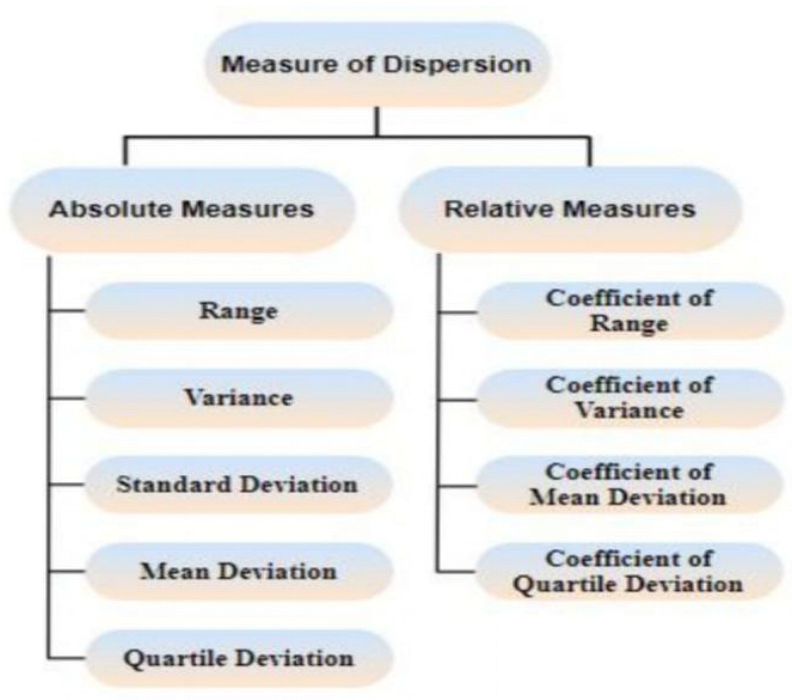
3Median = Mode + 2Mean



Measures of dispersion:

Dispersion in statistics is a way to describe how spread out or scattered the data is around an average value. It helps to understand if the data points are close together or far apart.

Dispersion shows the variability or consistency in a set of data. There are different measures of dispersion like range, variance, and standard deviation.



Range of Data

It is a given measure of how to spread apart values in a sample set or data set.

Range = Maximum value – Minimum value

Variance and Standard deviation

Variance is a measurement value used to find how the data is spread concerning the mean or the average value of the data set. It is used to find how the distribution data is spread out concerning the mean or the average value. The symbol used to define the variance is σ2.

Standard Deviation Formula for Grouped Data

$$(\sigma) = \frac{1}{N}\sqrt{N\sum_{i=1}^{n} f_i x_i^{\,2} - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$$

$$\sigma = \sqrt{\frac{\sum f x^2}{N} - \left(\frac{\sum f x}{N}\right)^2}$$

Variance Formula for grouped data:

$$\sigma^2 = \left(\frac{\sum f x^2}{N} - \left(\frac{\sum f x}{N}\right)^2\right)$$

There are two types of variance:

- Population variance: Often represented as σ^2
- Sample variance: Often represented as s^2.

Note: The standard deviation is the square root of the variance.

Combined Standard deviation:

$$\sigma_{12} = \sqrt{\frac{N_1 \cdot \left(\sigma_1^2 + d_1^2\right) + N_2 \cdot \left(\sigma_2^2 + d_2^2\right)}{N_1 + N_2}}$$

Coefficient of Variation:

Coefficient of variation is a type of relative measure of dispersion. It is expressed as the ratio of the standard deviation to the mean. The coefficient of variation is a dimensionless quantity and is usually given as a percentage. It helps to compare two data sets on the basis of the degree of variation.

Coefficient of variation formula:

	Coefficient of Variation	Standard Deviation
Population	$\frac{\sigma}{\mu} \times 100$	$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$
Sample	$\frac{s}{\mu} \times 100$	$s = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N-1}}$

Quartiles:

A quartile divides the set of observation into 4 equal parts. The middle term, between the median and first term is known as the first or Lower Quartile and is written as Q_1. Similarly, the value of mid-term that lies between the last term and the median is known as the third or upper quartile and is denoted as Q_3. Second Quartile is the median and is written as Q_2.

Quartiles formulae:

The following formula to calculate quartiles for grouped data:

$$Q_1 = L + \left(\frac{\frac{n}{4} - c}{f}\right) h$$

$$Q_3 = L + \left(\frac{\frac{3n}{4} - c}{f}\right) h$$

For Q_1 , first find n/4 For Q_3 ,find 3n/4

$$\text{Quartile Deviation} = \frac{Q3 - Q1}{2}$$

$$\text{Coefficient of Quartile deviation} = \frac{Q3 - Q1}{Q3 + Q1}$$

Interquartile Range (IQR)

The Interquartile Range is a measure of statistical dispersion, or how spread out the data points are in a data set. It is the range between the first quartile (Q1) and the third quartile (Q3) of a data set, which represents the middle 50% of the data.

Formula for IQR: $IQR = Q3 - Q1$

Representation of data

we can easily represent the data using various graphs, charts or tables. Various types of representing data set are:

- Bar graph
- Pie chart
- Line graph
- pictograph
- histogram
- frequency distribution

Conclusion:

Fundamental mathematics concepts such as linear algebra, calculus, probability, and statistics forms the backbone of Artificial Intelligence. These concepts enable the representation, optimization, and analysis of data, providing the tools necessary to build and refine AI models. A strong grasp of these mathematical foundations is essential for understanding and advancing the field of AI, empowering the development of intelligent systems that can learn, predict, and make decisions effectively.

References:

1. **https://www.geeksforgeeks.org/**
2. **Fundamentals of Mathematical Statistics by S C Gupta and V K Kapoor**
3. **Principles of Real Analysis by S C Malik**
4. **https://www.wikipedia.org/**

GitHub Repository link: https://github.com/VaishnaviRSaralaya/Basic_maths_for_AI