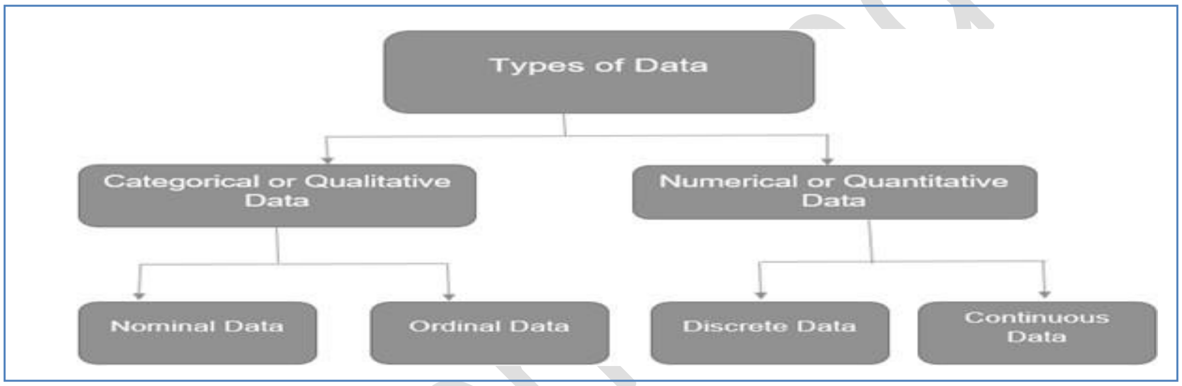
Statistics

- Types of Data
- Graphical Representation
- Frequency Distribution
- Sample Population
- Central Tendency
- Measure of Dispersion

What is Data?

Data is any information collected through observation for the purpose of analysis containingboth numerical or characteristics data points.

It is broadly categorized into 2 categories



Categorical Data

Categorical can be understood of that data where there are characteristics and have some meaning for each option like language preference, food menu, movie ratings, most of the option we see while filing forms.

It is of 2 types Nominal and Ordinal

Nominal Data

They are option based data that contains options but the order of the options don't matter muchlike

What is your mother tongue?

- English
- Hindi
- Telugu
- Tamil

Ordinal Data

Ordinal is more or less same as nominal with small difference that order of options matter, and specific.

Example:- Rating of the movies

- 
- 
- 
- 
- 

Numerical Data

Discrete Data

Discrete data can be understood of those quantities which can be counted and represented ina discrete or ungrouped Frequency distribution.

Example:-Number of people using different mobile phone in your respective locality

Mobile Brand	Number of user
Apple	80
Samsung	54
Xaomi	66
Realme	45
One Plus	75

2. Continuous Data

Continuous Data represents measurements and therefore their values **can't be counted butthey can be measured** or are has too much variability.

Example:-Students scoring marks in test out of 50marks by 100 students

Marks Scored	No of Students
0-10	5
10-20	14
20-30	18
30-40	9
40-50	4

TOTAL **50**

Frequency Distributions

There are 5 types of distribution in which the data is sorted

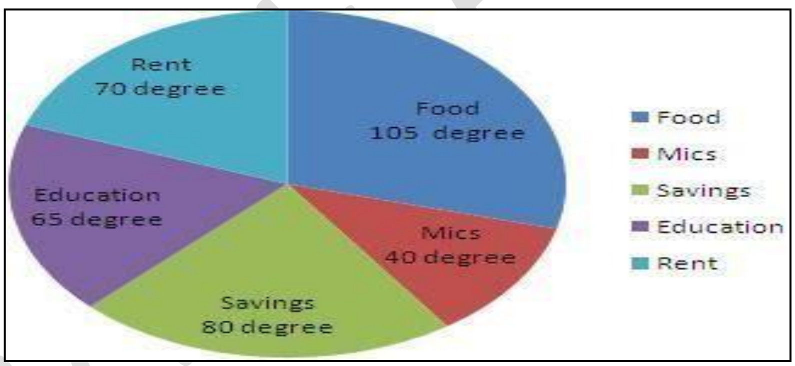
I) Discrete Frequency Distribution
II) Grouped Continuous Frequency distribution
III) Cumulative Frequency Distribution
IV) Relative Frequency Distribution
V) Relative Cumulative Frequency Distribution

Based on different frequency Distribution graphical representations are used.

Graphical Representations

There are 4 Graphical representations used in data science

I) Pie Chart

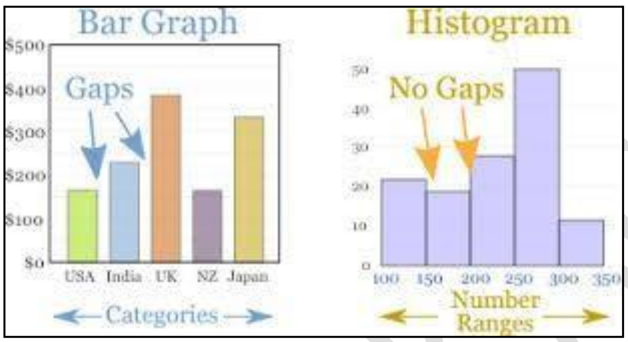


II) Bar Graph & Histogram

Bar Graph is used for discrete data set where there is data set is not continuous likeThe current example of education spent by individual counties

Whereas

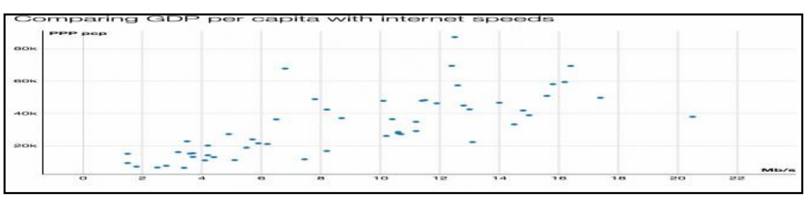
Histogram is used for Data set which continuous grouped data where there are intervals is having continuous limits, such as marks scored bu students out of 100 grouped in intervals of 10 each.



III) Scatter Plots (correlations)

A **scatter plot** also called a scatterplot, scatter graph; scatter chart, scatter gram,or scatter diagram is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data

The below example give the comparison between the per capita Income andinternet speed of countries.



They are used to plot data when representation is required to be made of aparticular data set based on 2 constraints or factors.

Correlation

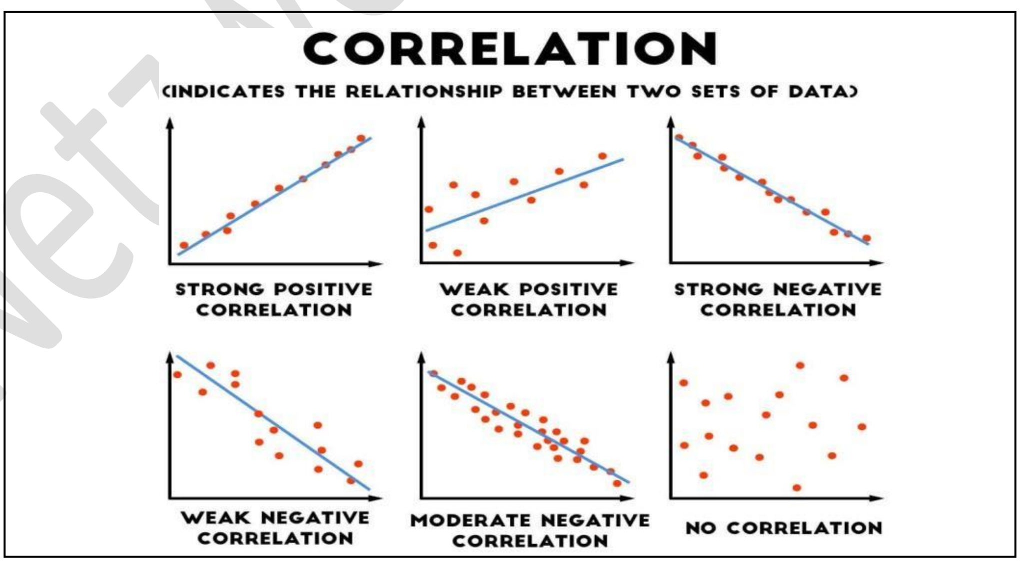
The **Scatter Diagram Method** is the simplest method to study the correlation between two variables wherein the values for each pair of a variable is plotted on a graph in the form of dots thereby obtaining as many points as the number of observations. Then by looking at the scatter of several points, the degree of correlation is ascertained.

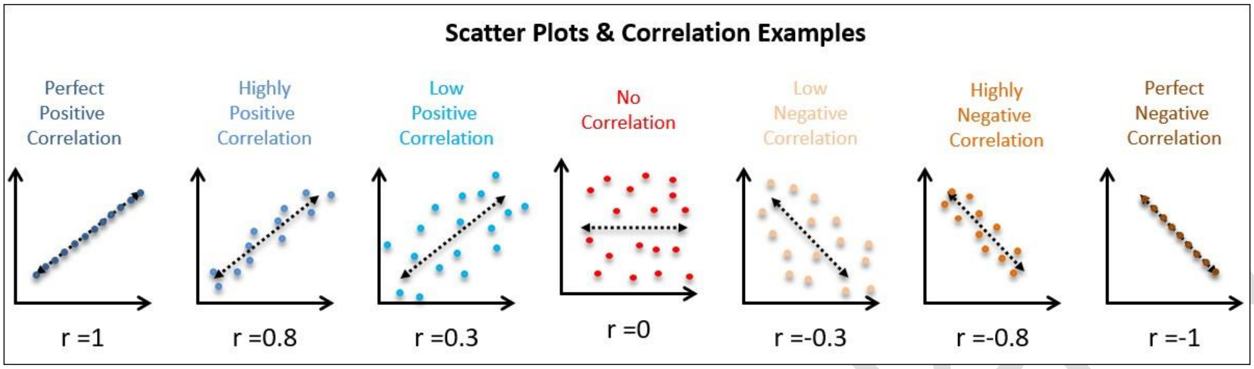
The degree to which the variables are related to each other depends on the manner in which the points are scattered over the chart. The more the points plotted are scattered over the chart, the lesser is the degree of correlation between the variables. The more the points plotted are closer to the line, the higher is the degree of correlation. The degree of correlationis denoted by **"r".**

The following types of scatter diagrams tell about the degree of correlation between variableX and variable Y.

1. **Perfect Positive Correlation (r=+1)**
2. **Perfect Negative Correlation (r=-1)**
3. **High Degree of +Ve Correlation (r= + High)**
4. **High Degree of –Ve Correlation (r= – High):**
5. **Low degree of +Ve Correlation (r= + Low):**
6. **Low Degree of –Ve Correlation (r= + Low):**
7. **No Correlation (r= 0): r = 0**

Thus, the scatter diagram method is the simplest device to study the degree of relationship between the variables by plotting the dots for each pair of variable values given





Central tendency

A measure of Central Tendency is the single value that describes the way in which a group of data clusters around the central value. We have learnt about methods of representing data graphically and in tabular form. Such representations exhibit certain characteristics or salientfeatures of the data.

We have also studied various methods of finding a representative value of the given data. This value is called the central value for the given data and various methods for finding thecentral value are known as the measures of central tendency.

The measures of central tendency are mean (arithmetic mean), median and mode. We have learnt that the measures of central tendency give us one single figure that represents the entire data 1.e., they give us one single figure around which the observations are concentrated.
In other words, measures of central tendency give us a rough idea where observations are centered.

Mean

The "mean" is the "average" you're used to, where you add up all the numbers and then divide by the number of numbers.

Direct method :

$$\bar{X} = \frac{\sum fx}{\sum f}$$

Assumed mean method :

$$\bar{X} = A + \frac{\sum fd}{\sum f}$$

$$d = x - A$$

Step deviation method :

$$\bar{X} = A + \left[\frac{\sum fd}{\sum f} \times c\right], \text{ where } d = \frac{x - A}{c}$$

Mode

The "mode" is the value that occurs most often. If no number in the list is repeated, thenthere is no mode for the list.

$$\text{Mode} = l_1 + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$$

Where, l_1 = The lower limit of the model class

f_1 = The frequency of the model class

f_0 = The frequency of the class preceding the model class

f_2 = The frequency of the class succeeding the model class

i = The size of the model class.

Median

The "median" is the "middle" value in the list of numbers. To find the median, your numbers have to be listed in numerical order from smallest to largest, so you may have to rewrite yourlist before

Median

n is odd,

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

n is even,

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

you can find the median.

<u>Measure of Dispersion</u>

But the central values are inadequate to give us a complete idea of the distribution as they donot tell us the extent to which the observations vary from the central value. In order to make better interpretation from the data, we should also have an idea how the observations are scattered or how much they are bunched around a central value. There can be two or more distributions having the same central value but still there can be wide disparities in the formation of the distribution as discussed below.

Consider following three distributions:

(i) 1, 5, 9, 13, 17

(ii) 3, 6,9, 12, 15

(iii) 7,8,9, 10, 11

All the three have same mean, median but there is wide variation between the data points ineach distributions.

If follows from the above discussion that the central values (mean, mode, median) are not sufficient to give complete information about a distribution. Variability in the values of the observations of given data gives us better information about the data. So, variability is another factor which is required to be studied in statistics. Like central value, we have a single number to Describe variability of a distribution. This single number is called the dispersion of distribution and various methods of determining or measuring dispersion arecalled the measures of dispersion.

As discussed above that the dispersion is the measure of variations in the values of thevariable it **measures the degree of scattered ness of the observations in a distribution around the central value.**

Following are commonly used measures of dispersion:

(i) Range
(ii) Quartile deviation
(iii) Mean deviation (iv)
Standard deviation

In study of data Science, more often standard Deviation and variance is used so that must be well versed by Students

Mean Deviation

For Ungrouped Data	***For Grouped Data***
$M.D\ (mean) = \frac{\sum\lvert x-\mu\rvert}{N}$	$M.D\ (mean) = \frac{\sum f\lvert x-\mu\rvert}{N}$

x_i	$\lvert x_i - mean \rvert$
6	4
7	3
10	0
12	2
13	3
4	6
8	2
20	10
TOTAL	30

Sum 80

Mean 10

Mean Deviation 3.75

Standard Deviation & Variance

This is one of most important part and s used extensively in stats and probability.It is denoted by SD or σ called Sigma.
Variance is equal to the square of SD and it is represented by σ^2 (Sigma Square)The formula for SD is

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

where,

σ = population standard deviation
$\sum$ = sum of...
μ = population mean
n = number of scores in sample.

In practice mostly first variance is calculated and then the square root is taken of that value.

x_i	x_i -mean	$(x_i - \text{mean})^2$
8	-6	36
12	-2	4
13	-1	1
15	1	1
22	8	64
TOTAL		106

Sum	70
Mean	14
Variance	**21.2**
SD	**4.604**

Sample and Population

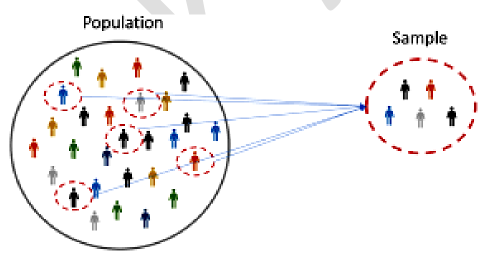
Population: - It is the entire group of people or thing we want to study about.

ample :- It is the part of the population the we actually want to collect data or make some observations about.

Sample Question

A factory overseer selects 50 iPhones produced at random from those produced that week atthe factory, to test their strength.

Sample:- the sample is the 50 iPhones selected.
Population:- is all iPhones produced at the factory that week.



Calculus

- Functions
- Differentiation
- Fundamental Rules of Differentiation
- Slope and Tangent
- Optimization
- Multivariate Functions
- Partial Differentiation
- Gradient

Function, in mathematics, an expression, rule, or law that defines a relationship between one variable (the independent variable) and another variable (the dependent variable). Functions are ubiquitous inmathematics and are essential for formulating physical relationships in the sciences.

This relationship is commonly symbolized as $y = f(x)$. In addition to $f(x)$, other abbreviated symbols such as $g(x)$ and $P(x)$ are often usedto represent functions of the independent variable x, especially when the nature of the function is unknown or unspecified.

There are many types of functions likeAlgebraic

Functions

Polynomial Functions

Exponential Functions and etc.

In Data Science, we use

f(x) = mx + b f(x) = ex f(x) = log(x)

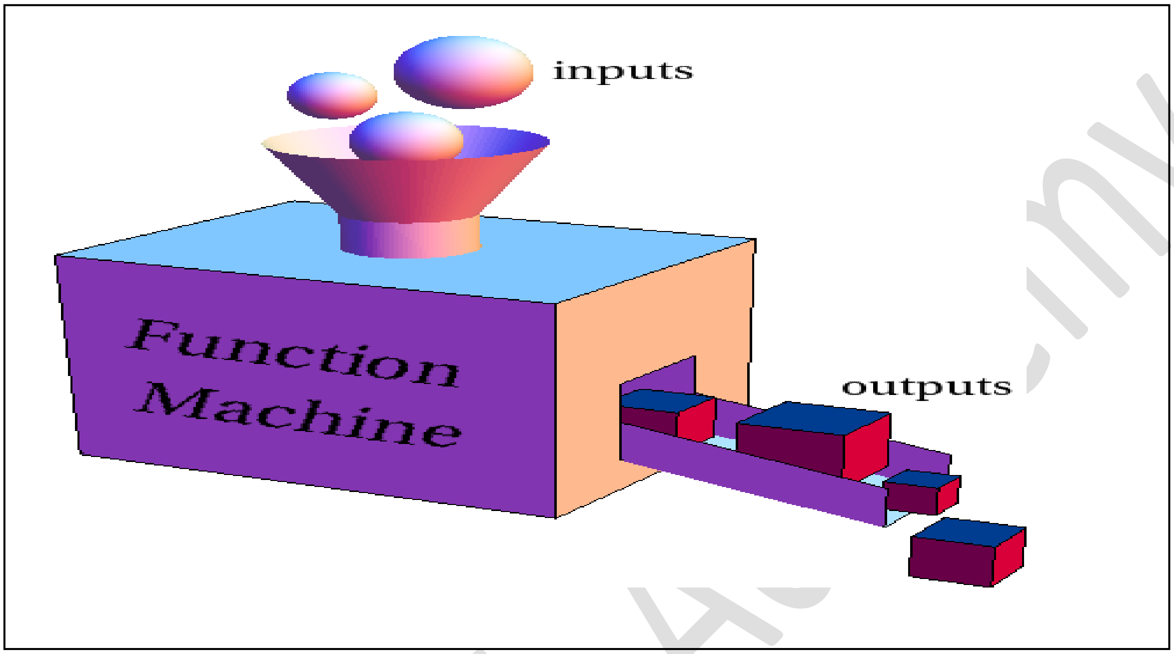
Function, in lay man language can be understood as a machinewhich converts a particular input into a particular output.

Here the

Input is the different values of x

Machine is the function

Output is the value of the y



The set of all inputs x is called Domain **The**

set of all outputs y is called Range.

Example

The function$f(x) = x^2$

x=1 y=1

x=2 y=4

x=3 y=9

x=4 y=16

Differentiation

Differentiation, in mathematics, process of finding the derivative, orrate of change, of a function. In contrast to the abstract nature of the theory behind it, the practical technique of differentiation can be carried out by purely algebraic manipulations, using three basic derivatives, four rules of operation, and a knowledge of how tomanipulate functions.

The abstract meaning of differentiation will be explained later insection of Tangent and slope.

How to differentiate a function

The process of finding the derivate of the function is calleddifferentiation.

There is some notation for the derivative of the function such as

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

the process of finding the derivate of complex functions e will usesome basic formulae and rule of differentiation

The list of basic formulae used is given below:

These formulae are used to find the derivative of the function.

$$\frac{dk}{dx} = 0 \quad \text{where } k = \text{constant}$$

$$\frac{d(x)}{dx} = 1$$

$$\frac{d(kx)}{dx} = k \quad \text{where } k = \text{constant}$$

$$\frac{d(x^n)}{dx} = nx^{n-1}$$

$$\frac{d(\sin x)}{dx} = \cos x$$

$$\frac{d(\cos x)}{dx} = -\sin x$$

$$\frac{d(\tan x)}{dx} = \sec^2 x$$

$$\frac{d(\cot x)}{dx} = -\operatorname{cosec}^2 x$$

$$\frac{d(\sec x)}{dx} = \sec x \tan x$$

$$\frac{d(\operatorname{cosec} x)}{dx} = -\operatorname{cosec} x \cot x$$

$$\frac{d(e^x)}{dx} = e^x$$

$$\frac{d(\ln(x))}{dx} = \frac{1}{x}$$

$$\frac{d(a^x)}{dx} = a^x \log a$$

$$\frac{d(x^x)}{dx} = x^x(1 + \ln x)$$

$$\frac{d(\log_a x)}{dx} = \frac{1}{x} \times \frac{1}{\ln a}$$

Rules of Differentiation

When the function get complicated like $F(x) = \log(\sin^4(e^{\tan(x)}))$

Merely by the above formulae, we will not be able to solve these questions. To extend the notion of differentiation we will introducesome more rules.

Let us consider two functions f(x) and g(x), where f'(x) and g'(x)represent their first derivatives

RULES I: When the sum or difference of 2 functions are given

Sum Rule: $\frac{d}{dx}[f(x)+g(x)] = f'(x)+g'(x)$

Difference Rule: $\frac{d}{dx}[f(x)-g(x)] = f'(x)-g'(x)$

RULES II: When the product of 2 functions are given

Product Rule: $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x)+g(x)f'(x)$

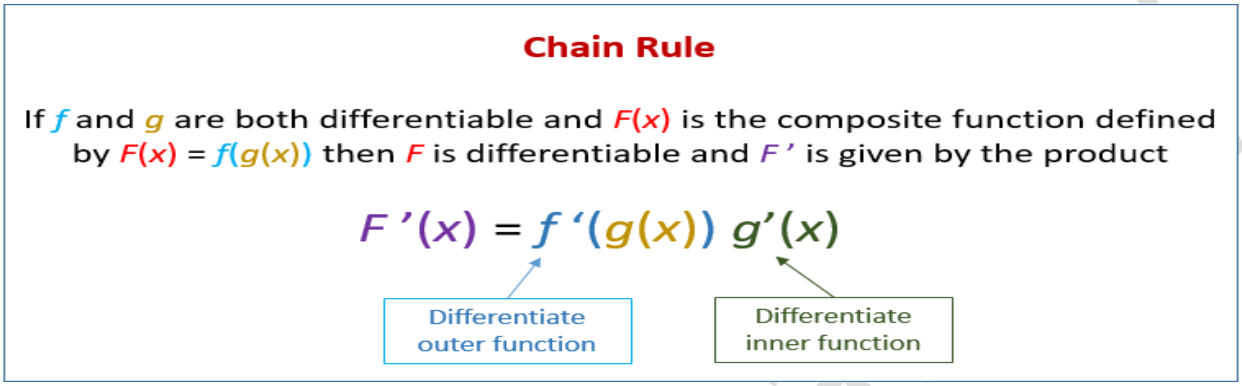
RULES III: When the Quotient of 2 functions are given

Quotient Rule: $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x)-f(x)g'(x)}{[g(x)]^2}$

RULES IV: This rule is applied when the function are of the formf(g(h(x))).

Chain Rule: $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$

Note: - These rules are used in Partial differentiation, so make sure tobe well versed in them.



Optimization

Optimization means to make maximize or minimize a particularfunction using Calculus derivative methods.

In many situations the need of the algorithm or function is to find a value which can either give the value which maximizes or minimizesthe function.

Example:

I) Company in order to maximize the revenue tends to increasethe price of the commodity keeping the fact the some subscribers will discontinue, so what is that price of the commodity that will maximize the revenue?

II) A family wishes to increase its saving so what's the amount ofexpenditure it should minimize to make the increase the savings.

How to optimize the functions

Following steps will help to optimize the functions:-

Step I: Convert the situation to a mathematical function.

Step II: understand the need of the problem whether to maximize orminimize.

of the function and equate it to 0.

Step III: Take the $\frac{dy}{dx}$

Step IV: Obtain the values of x for step III

Step V: Now take the d^2y/dx^2 of the main function.

Step VI: Us e the second order derivative method to check whetherthe points will maximize or minimize.

If

f''(x) >0 Its minima

f''(x) <0 its maxima

f''(x) is the second derivative which is the derivative of the derivative.

Example:- Find two positive numbers whose sum is 300 and whoseproduct is a maximum.

Solution:- The first step is to write down equations describing this situation.

Let's call the two numbers x and y and we are told that the sum is 300

$$x+y=300$$

We are being asked to maximize the product,

$$P=xy$$

We now need to solve the constraint for x or y (and it really doesn't matter which variable we solve for in this case) and plug this into theproduct equation.

$$y=300-x$$

$$\Rightarrow P(x)=x(300-x)=300x-x^2$$

Step 3: Find the first derivative

$$A'(x)=300-2x$$

$$\rightarrow 300-2x=0$$

$$\rightarrow x=150$$

$A''(x) = -2$

Finally, let's actually answer the question. We need to give both values.We already have x so we need to determine the value of y.

$$\mathbf{X} = 300-150=150$$

$$\mathbf{Y} = 300-150=150$$

The final answer is then,

$$\mathbf{x} = 150$$

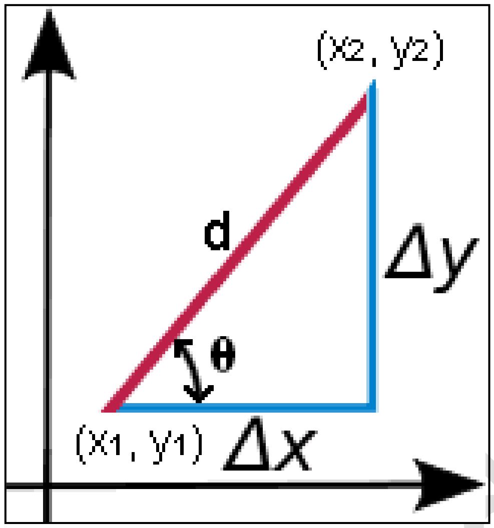
$$\mathbf{y} = 150$$

Tangent and Slope

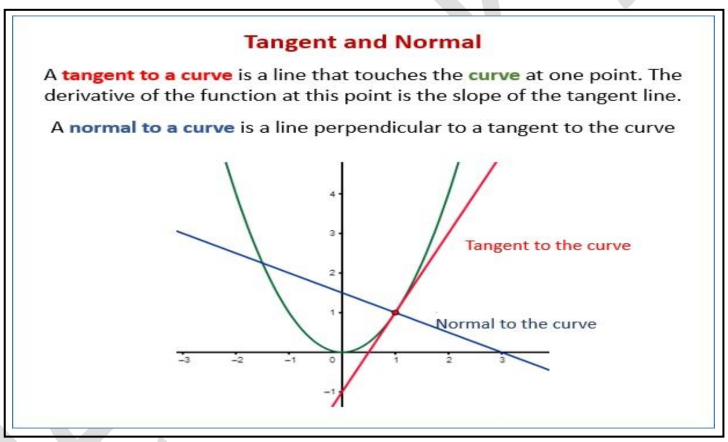
Slope: - The trigonometrically tangent of the angle that lines makeswith the positive direction of the x-axis in anticlockwise direction is called the slope or gradient of the line.

It is denoted by **m**.

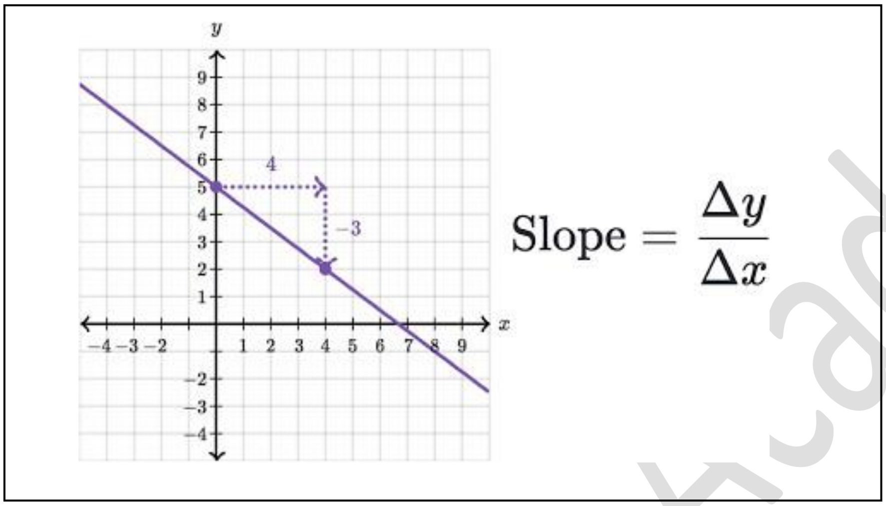
$$\mathbf{m = \tan\theta}$$



(X2, y2)
d
Δy
θ
(X1, y1)
Δx



Tangent and Normal
A tangent to a curve is a line that touches the curve at one point. The derivative of the function at this point is the slope of the tangent line.
A normal to a curve is a line perpendicular to a tangent to the curve
Tangent to the curve
Normal to the curve



The slope of the tangent is calculated by taking the first derivative ofthe function at point x_0.

m = $\frac{dy}{dx}$ at point X_0.

Multivariate Functions

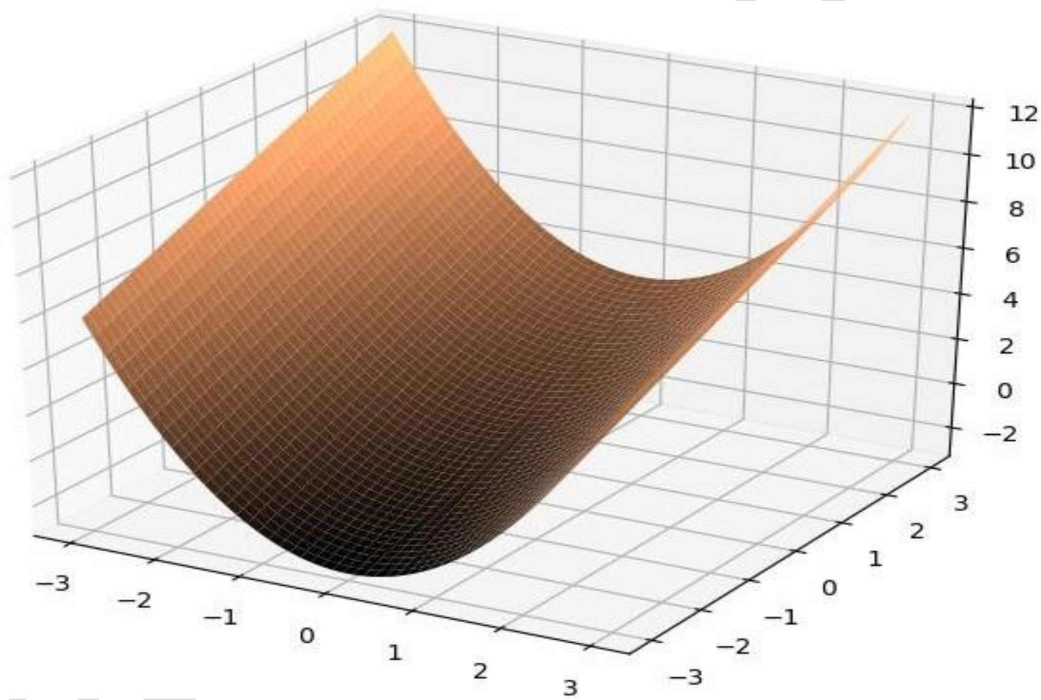
A multivariable function is a function with more than one variable.

$$f(x,y) = x^2 + y^4$$

Such functions have greater use when understanding problemsinvolving multiple factors.

The mapping of such functions, basically involves multiple dimensionswhich is 1 more than the number of the variables used.

Graphically these functions will be a look like these:-



Partial Derivatives

The differentiation of the multivariable functions is done using partialderivation. In this the f(x,y) is differentiated with respect to each variable.

Notation:

- For $f(x,y)$, the partial derivative with respect to x can be written as:

$$\frac{\partial f}{\partial x} = f_x(x,y) = f_x = \frac{\partial}{\partial x}[f(x,y)]$$

- For $f(x,y)$, the partial derivative with respect to y can be written as:

$$\frac{\partial f}{\partial y} = f_y(x,y) = f_y = \frac{\partial}{\partial y}[f(x,y)]$$

$f(x, y) = 2x^7 + 9xy$:

$$\frac{\partial f(x, y)}{\partial x} = \frac{\partial}{\partial x}(2x^7 + 9xy) = \frac{\partial}{\partial x}2x^7 + \frac{\partial}{\partial x}9xy$$

$$= 14x^6 + 9y\frac{\partial}{\partial x}x = 14x^6 + 9y \times 1 = 14x^6 + 9y$$

$$\frac{\partial f(x, y)}{\partial y} = \frac{\partial}{\partial y}(2x^7 + 9xy) = \frac{\partial}{\partial y}2x^7 + \frac{\partial}{\partial y}9xy$$

$$= 0 + 9x \times \frac{\partial}{\partial y}y = 0 + 9x \times 1 = 9x$$

$$\frac{\partial f(x,y)}{\partial x} = \frac{\partial}{\partial x}3x^2y = 6yx$$

$$\frac{\partial f(x,y)}{\partial y} = \frac{\partial}{\partial y}3x^2y = 3x^2$$

Example:-

NOTE:- All the rule of differentiation explained in previous sectionwill be applied to partial differentiation also.

Gradient of the function

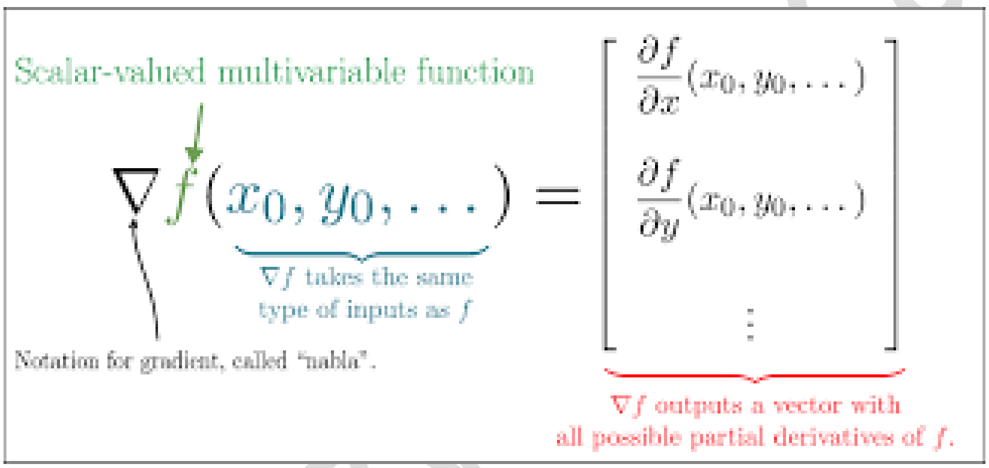
The Gradient of a function is the packaging of all the partial derivativeof the function with respect to all the variables of the functions into a column matrix.

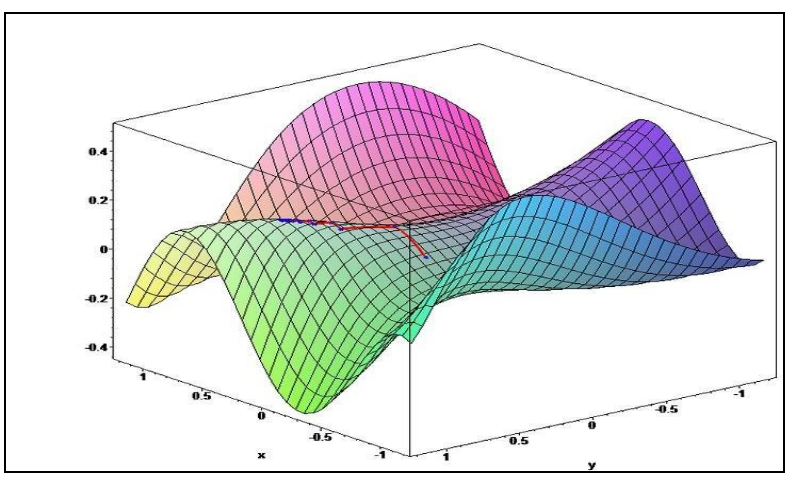
This packaging becomes a vector which gives us the direction inmultidimensional spaces.

The direction in 3D spaces is given by **Gradient.**

The direction or tangent of the 2D plane is given **by Tangent to thecurve.**

The graphical representation of the gradient of multivariate function.





Linear Algebra

1) Matrix

- Types
- Operations
- Inverse

2) Vectors

- Types
- Operations
- Magnitude , Unit vector
- Projection of a vector
- Products

3) Advanced Linear Algebra

- Linear combination of vectors
- Span, Basis
- Linear Dependency and Independency
- Linear Transformation

Matrix

A matrix represents a collection of numbers arranged in an order of rows and columns. It is necessary to enclose the elements of a matrix in parentheses or brackets.
Matrices have also come to have important applications in computer graphics, where they have been used to represent rotations and other transformations of images.

If there are m rows and n columns, the matrix is said to be an "m by n" matrix, written"$m \times n$." For example,

$$\begin{bmatrix} 1 & 3 & 8 \\ 2 & -4 & 5 \end{bmatrix}$$

s a 2×3 matrix. A matrix with n rows and n columns is called a square matrix of order n. An ordinary number can be regarded as a 1×1 matrix; thus, 3 can be thought of as the matrix [3].

In a common notation, a capital letter denotes a matrix, and the corresponding small letter with a double subscript describes an element of the matrix. Thus, a_{ij} is the element inthe

ith row and jth column of the matrix A. If A is the 2×3 matrix shown above, then $a_{11} = 1$, $a_{12} = 3$, $a_{13} = 8$, $a_{21} = 2$, $a_{22} = -4$, and $a_{23} = 5$. Under certain conditions, matrices can be added and multiplied as individual entities, giving rise to important mathematical systemsknown as matrix algebras.

This Matrix [M] has 3 rows and 3 columns. Each element of matrix [M] can be referredto by its row and column number. For example, $a_{23}=6$

Order of a Matrix:

The order of a matrix is defined in terms of its number of rows and columns.Order of a matrix = No. of rows ×No. of columns
Therefore Matrix [M] is a matrix of order 3 × 3.

Transpose of a Matrix:

The transpose $[M]^T$ of an m x n matrix [M] is the n x m matrix obtained by interchanging therows and columns of [M].
if $A= [a_{ij}]$ mxn , then $A^T = [b_{ij}]$ nxm where $b_{ij} = a_{ji}$

Singular and Nonsingular Matrix:

1. Singular Matrix: A square matrix is said to be singular matrix if its determinant is zero i.e. $|A|=0$
2. Nonsingular Matrix: A square matrix is said to be non-singular matrix if its determinantis non-zero.

Here |A| should not be equal to zero , means matrix A should be non-singular.

Types of Matrices:

(1) Row Matrix: Row matrix is a type of matrix which has just one row. It can have multiple columns but there is just a single row present in a row matrix. Example of row matrix can be given as

A= [3 2 1] which has just one row but has three columns.We can mathematically define row matrix as:
Matrix of the form $A=[a_{ij}]_{1\times n}$ where 1 represents just a single row and n represents number of columns.

(2) Column Matrix: Column matrix is a type of matrix which has just one column. It canhave multiple rows but there is just one column present in a column matrix. Example of column matrix can be given as:

which has just one column but has three rows. We
can mathematically define column matrix as:
Matrix of the form $A=[a_{ij}]_{m\times 1}$ where m represents number of rows and 1 represents just a single column.

(3) Null Matrix: Null matrix is a type of matrix which has all elements equal to zero. Example of null matrix can be given as A= [0000]. We can also mathematically define null matrix as**:**
$A=[a_{ij}]_{m\times n}$ where $a_{ij}=0$ for all i,j.

(4) Square Matrix: Square Matrix is a type of **matrix** which has equal numberof **rows** and **columns.** Example of square matrix can be given as

We can define square matrix mathematically as matrix of the form $A=[a_{ij}]_{m\times n}$ where m=n.

(1) Diagonal Matrix: It is type of square matrix which has all the non-diagonal elementsequal to zero.

We can mathematically define diagonal matrix as a matrix of the form $A=[a_{ij}]_{n\times n}$,where $a_{ij}=0$ when $i\neq j$.

---

(2) Identity Matrix: It is a type of square matrix which has all the main diagonal elements equal to 1 and all the non-diagonal elements equal to 0. It is also called unit matrix.
We can mathematically define identity matrix as a matrix of the form $A=[a_{ij}]_{n\times n}$, where$a_{ij}=0$ for $i\neq j$ and $a_{ij}=1$ for i=j.

Operations on Matrices

Addition

- Order of the matrices must be the same
- Add corresponding elements together
- Matrix addition is commutative
- Matrix addition is associative

Subtraction

- The order of the matrices must be the same

- Subtract corresponding elements
- Matrix subtraction is not commutative (neither is subtraction of real numbers)
- Matrix subtraction is not associative (neither is subtraction of real numbers)

Scalar Multiplication

A scalar is a number, not a matrix.

- The matrix can be any order
- Multiply all elements in the matrix by the scalar
- Scalar multiplication is commutative
- Scalar multiplication is associative

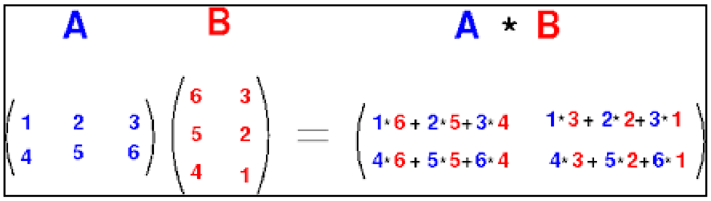
Zero Matrix

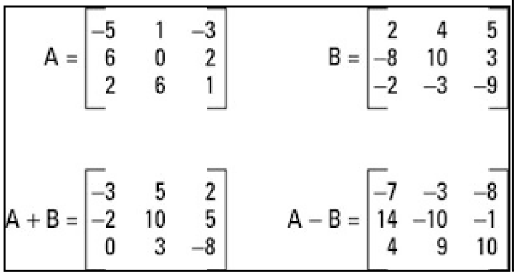
- Matrix of any order
- Consists of all zeros
- Denoted by capital O
- Additive Identity for matrices
- Any matrix plus the zero matrix is the original matrix

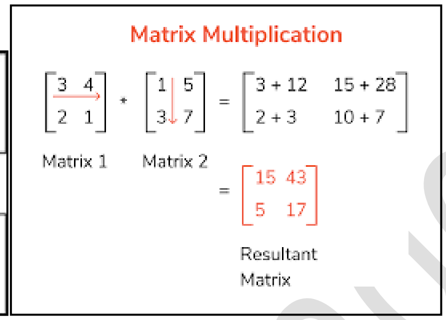
Matrix Multiplication

$$A_{m\times n} \times B_{n\times p} = C_{m\times p}$$

- The number of columns in the first matrix must be equal to the number of rows in the second matrix. That is, the inner dimensions must be the same.
- The order of the product is the number of rows in the first matrix by the number of columns in the second matrix. That is, the dimensions of the product are the outer dimensions.
- Since the number of columns in the first matrix is equal to the number of rows in the second matrix, you can pair up entries.
- Each element in row i from the first matrix is paired up with an element in column j from the second matrix.
- The element in row i, column j, of the product is formed by multiplying these paired elements and summing them.
- Each element in the product is the sum of the products of the elements from rows i ofthe first matrix and column j of the second matrix.
- There will be n products which are summed for each element in the product.







Inverse of a matrix

Suppose **A** is an *n* x *n* matrix. The inverse of **A** is another *n* x *n* matrix, denoted $\mathbf{A^{-1}}$, thatsatisfies the following conditions.

$\mathbf{AA^{-1}} = \mathbf{A^{-1}A} = \mathbf{I_n}$, where $\mathbf{I_n}$ is the identity matrix.

Here |A| should not be equal to zero , means matrix A should be non-singular.

Determinant

The **determinant** is a scalar value that can be computed from the elements of a square matrix and encodes certain properties of the linear transformation described by the matrix.The **determinant** of a matrix A is denoted det(A), det A, or |A|.

Vectors

Vector, in Mathematics, a quantity that has both magnitude and direction but not position. Examples of such quantities are velocity and acceleration.

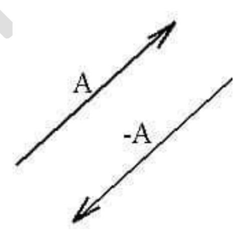
Vectors may be visualized as directed line segments whose lengths are their magnitudes. Since only the magnitude and direction of a vector matter, any directed segment may be replaced by one of the same length and direction but beginning at another point, such as the origin of a coordinate system. Vectors are usually indicated by a boldface letter, such as *v*. A vector's magnitude, or length, is indicated by |*v*|, or *v*, which represents a one-dimensional quantity (such as an ordinary number) known as a scalar

Types

- **Null vector:-** A vector whose magnitude is zero and has no direction, it may have all directions is said to be a null vector. A null vector can be obtained by adding two or more vectors.

- **Negative vector**

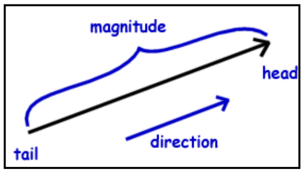
  A vector having the same magnitude as that of the given vector but the opposite direction is called a negative vector.



Unit Vector

A unit vector is that whose magnitude is unity i.e 1 and has any given direction only. A unit vector is obtained by dividing the vector with magnitude.

Magnitude



The magnitude of a vector is the length of the vector. The magnitude of the vector a is denotedas ||a||.
For a two-dimensional vector a= (a1,a2)a=(a1,a2), the formula for its magnitude is

$$\text{Magnitude} = \sqrt{x^2 + y^2} \quad \text{(for 2D vectors)}$$

$$\text{Magnitude} = \sqrt{x^2 + y^2 + z^2} \quad \text{(for 3D vectors)}$$

Unit vector

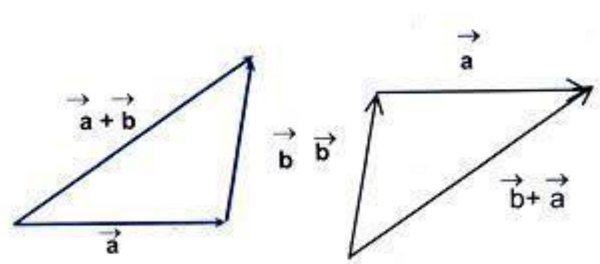
A unit vector is a vector that has a magnitude of 1 unit. A unit vector is also known as a direction vector. It is represented using a lowercase letter with a cap („^') symbol along with it. A vector can be represented in space using unit vectors. Any vector can become a unit vector by dividing it by the vector's magnitude as follows:
Where,
a^ represents a unit vector
|a| represents the magnitude of the vector

$$\hat{a} = \frac{a}{|a|}$$

Operations

Like any other Number set in math, vectors too have the same operation for them :-

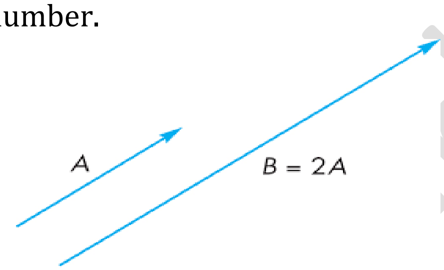
Addition & Subtractions



Adding and subtraction a vector is done in the same way as two matrices are added by component addition. All the vectors can be expressed as the column andcan be added or subtracted using same rules.

Multiplication With a scalar

When vectors are multiplied or divided they are done in the same way as multiplying a matrixwith a scalar number.



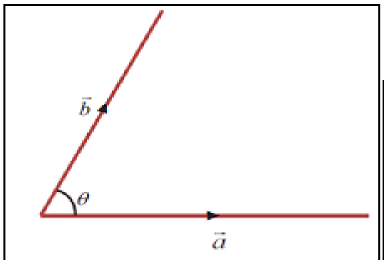
Products

In vector Algebra, their product with other vector is defined in a different way:-
When 2 vectors are multiplied to each other there are 2 possibilities that the result may be ascalar (simple number) or it is vector.
So, there are 2 products namely dot product and cross product.

Dot product

It is also called scalar product as the name suggests that the outcome is a scalar number. Which is given by the following formula, where θ is the angle between a vector and b vector?

$$A \cdot B = |A||B|\cos(\theta)$$



$$\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|\cos\theta$$

$$\begin{bmatrix} A_x & A_y & A_z \end{bmatrix}\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = A_xB_x + A_yB_y + A_zB_z = \vec{A}\cdot\vec{B}$$

The part on the left side is solved using the matrix multiplication as given above. Sometime the angle θ between the vector need to be calculated which is calculated as

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\cdot\|\vec{b}\|}$$

$$\theta = \cos^{-1}\left(\frac{\bar{v}\cdot\bar{w}}{\|\bar{v}\|\|\bar{w}\|}\right)$$

A sample problem has been done to show the calculation.

$$\mathbf{a} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

$$\mathbf{a}.\mathbf{b} = x_1x_2 + y_1y_2 + z_1z_2$$

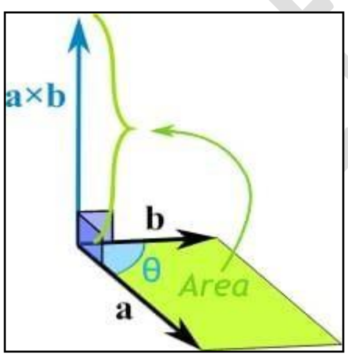
$$\mathbf{a} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}$$

$$\mathbf{a}.\mathbf{b} = 2(3) + 4(5) + 6(7)$$

$$= 68$$

Cross product

It is also called Vector product as the name suggests that the outcome is a vector perpendicular to the product, which is given by the following formula, where θ is the angle between a vector and b vector.



$$\vec{\mathbf{a}} \times \vec{\mathbf{b}} = |\vec{a}|\ |\vec{b}|\ \sin\theta\ \hat{n}$$

This ncap written in RHS is the perpendicular vector to the product of a and b.
The part on the left side is solved like solving a determinant given below.

$$\mathbf{x}_1 \times \mathbf{x}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}$$

$$= \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix}\mathbf{i} - \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix}\mathbf{j} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}\mathbf{k}$$

A sample problem has been done to show the calculation.

$$\mathrm{x}_1 \times \mathrm{x}_2 = \begin{vmatrix} \mathrm{i} & \mathrm{j} & \mathrm{k} \\ 2 & -3 & 1 \\ -2 & 1 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} -3 & 1 \\ 1 & 1 \end{vmatrix}\mathrm{i} - \begin{vmatrix} 2 & 1 \\ -2 & 1 \end{vmatrix}\mathrm{j} + \begin{vmatrix} 2 & -3 \\ -2 & 1 \end{vmatrix}\mathrm{k}$$

$$= [(-3)(1)-(1)(1)]\mathrm{i} - [(2)(1)-(-2)(1)]\mathrm{j} + [(2)(1)-(-2)(-3)]\mathrm{k}$$

$$= -4\mathrm{i} - 4\mathrm{j} + 8\mathrm{k}$$

Advanced Linear Algebra

Linear combination of vectors

If one vector is equal to the sum of scalar multiples of other vectors, it is said to be a linearcombination of the other vectors.

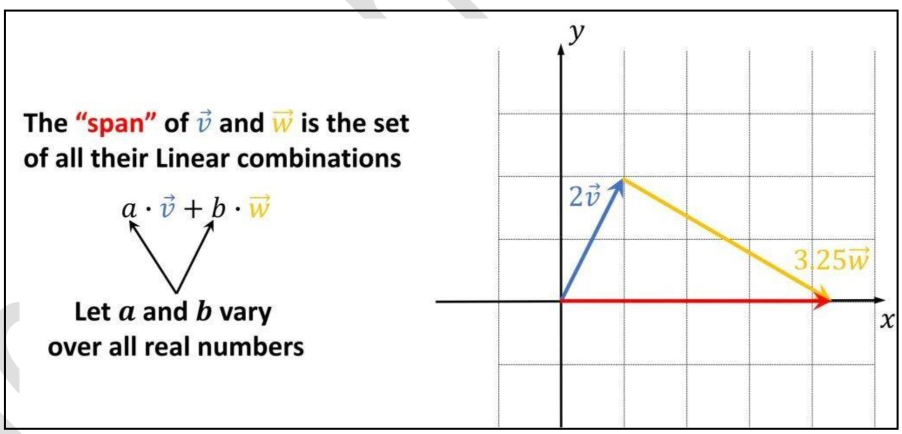
A *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in V$ is a sum of the form $f_1\mathbf{v}_1 + f_2\mathbf{v}_2 + \cdots + f_n\mathbf{v}_n$ where $f_1, f_2, \dots, f_n \in F$ are scalars, a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$.

linear combination of vectors. Let the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \cdots \vec{v}_n$ be vectors in $\mathbb{R}^n$ and $c_1, c_2, \cdots, c_n$ be scalars. Then the vector $\vec{b}$, where $\vec{b} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$ is called a *linear combination* of $\vec{v}_1, \vec{v}_2, \vec{v}_3, \dots \vec{v}_n$. The scalars $c_1, c_2, \dots, c_n$ are commonly called the "weights".

Again, this is stating that $\vec{b}$ is a result of combining the vectors using scalar multiplication (the c's) and addition.

Span

The set of all vector that are the linear combinations of the vectors in the set **V={v1,v2,v3,........,vn}** is called the span of V and is denoted by SpanV or Span**{v1,v2,v3,......,vn}**.The span of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is the set of all linear combinations of the vectors: $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = \{f_1\mathbf{v}_1 + f_2\mathbf{v}_2 + \cdots + f_n\mathbf{v}_n \mid f_1, f_2, \dots, f_n \in F\}$.



Basis

A Basis is the set of all the vectors which constitute the span of linear combinations of vectorsor written in a unique way as a linear combination of vectors.
A basis for a vector space is a **linearly independent** spanning set of the vector space.

The following three-dimensional vectors are linearly independent.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Any vector in R^3 can be expressed as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$.

$\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ form a basis of R^3.

Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = R^3$.

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$$

this set S is the set of Basis

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

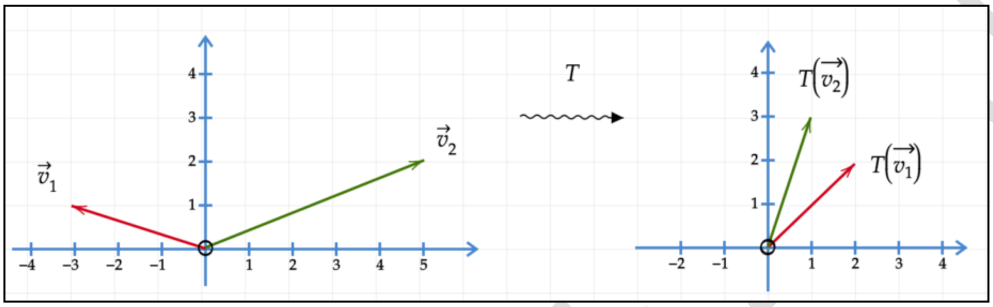
All the vectors produced by the linear combination of the Set S is called span of S.

Linear Dependency and Independency

If at least one of the vectors in the set can be defined as a linear combination of the othervector in the same set, then that set of vectors is said to be **linearly dependent**
 if no such vector in the set can be expressed as the linear combination of the vectors , then theset of vectors are said to be **linearly independent.**

If the determinant of the vectors of the set is equal to **0,** then it is linearly dependent.
Else it is linearly dependent.

Linear Transformation



Transformation is the word used to denote used for functions involving vectors.

$$\mathbf{T} = f(x)$$

A linear transformation (or a linear map) is a function

$$\mathbf{T: R^n \rightarrow R^m}$$

that satisfies the following properties:

1. $T(x + y) = T(x) + T(y)$
2. $T(ax) = aT(x)$

for any vectors x,y ∈ R^n, and any scalar a ∈ R

Probability

- Probability Distribution
- Conditional Probability
- Probability Models
- Hypothesis Tests

Probability Distribution

In probability theory and statistics, a **probability distribution** is the mathematical function that gives the probabilities of occurrence of different possible **outcomes** for an experiment.

More specifically, the probability distribution is a mathematical description ofa random phenomenon in terms of the probabilities of events.
For instance, if the random variable X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X = tails (assuming the coin is fair). Examples of random phenomena can includethe results of an experiment or survey.

A probability distribution is a mathematical function that has a sample space as its **input**,and gives a **probability** as its output.

The sample space is the set of all possible outcomes of a random phenomenon being observed; it may be the set of real numbers or a set of vectors, or it may be a list of non-numerical values.

For example, the sample space of a coin flip would be {heads, tails}.Probability

distributions are generally divided into two classes.

A **discrete probability distribution** (applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice)
can be encoded by a discrete list of the probabilities of the outcomes,known as a probability mass function.

On the other hand, a **continuous probability distribution** (applicable to the scenarioswhere the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by probability density functions (with the probability of any individual outcome actually being 0).

- A function that represents a discrete probability distribution is called a **probability mass function**.
- A function that represents a continuous probability distribution is called a **probability density function.**

- Functions that represent probability distributions still have to obey the rules of probability
- The output of *a* probability mass function is a probability *whereas* the area under the curve produced by a probability density function represents a probability.
- Parameters of a probability function play a central role in defining the probabilities of the outcomes of a random variable.

Conditional Probability

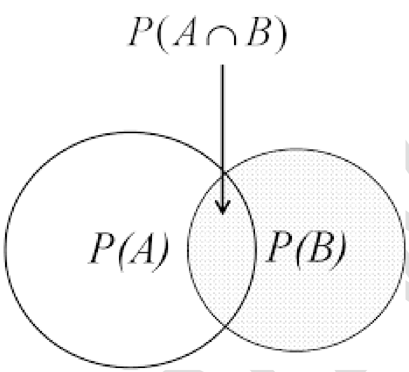
The ***conditional probability*** of an event *B* is the probability that the event will occur given the knowledge that an event *A* has already occurred. This probability is written *P(B|A)*, notation for the *probability of B given A*.

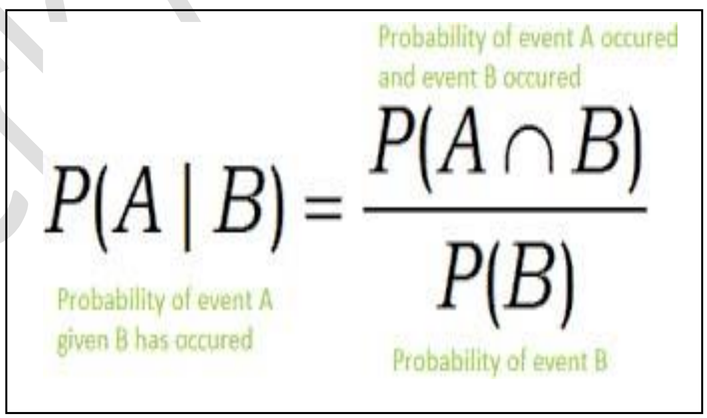
In the case where events *A* and *B* are *independent* (where event *A* has no effect on the probability of event *B*), the conditional probability of event *B* given event *A* is simply theprobability of event *B*, that is *P(B)*.

If events *A* and *B* are not independent, then the probability of the *intersection ofA and B* (the probability that both events occur) is defined by

P(A and B) = P(A)P(B|A).

From this definition, the conditional probability *P(B|A)* is easily obtained by dividingby *P(A)*:





You toss a fair coin three times,
Given that you have observed at least one heads,
What is the probability that you observe at least two heads?

Let A1 be the event that you observe at least one heads, And A2 be the event that you observe at least two heads.Then A1=S−

{TTT}, and P(A1)=7/8;

A2={HHT,HTH,THH,HHH}, and P(A2)=4/8.
Thus, we can write
P(A2|A1)P(A2|A1) =P(A2∩A1)P(A1)
=P(A2)P(A1)
=4/8.8/7=4/7

Special probability distributions

There are some standard random variables, and have their probability distribution functionsTwo of the most widely used discrete probability distributions are the **binomial** and **Poisson**.

The Binomial distribution

The binomial probability mass function (equation 6) provides the probability that *x* successeswill occur in *n* trials of a binomial experiment.

A binomial experiment has four properties:

(1) It consists of a sequence of *n* identical trials;

(2) Two outcomes, success or failure, are possible on each trial;

(3) The probability of success on any trial, denoted *p*, does not change from trial to trial; and

(4) The trials are independent.

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

For instance, suppose that it is known that 10 percent of the owners of two-year oldautomobiles have had problems with their automobile's electrical system.

To compute the probability of finding exactly 2 owners that have had electrical system problems out of a group of 10 owners, the binomial probability mass function can be used bysetting

$n = 10$, $x = 2$, and $p = 0.1$ in the above equation,So

the probability in this case is **0.1937**.

The Poisson distribution

The Poisson probability distribution is often used as a model of the number of arrivals at a facility within a given period of time. For instance, a random variable might be defined as thenumber of telephone calls coming into an airline reservation system during a period of 15 minutes. If the mean number of arrivals during a 15-minute interval is known, the Poisson probability mass function given by the following equation and can be used to compute the probability of x arrivals.

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

where

$x = 0, 1, 2, 3, \ldots$

λ = mean number of occurrences in the interval

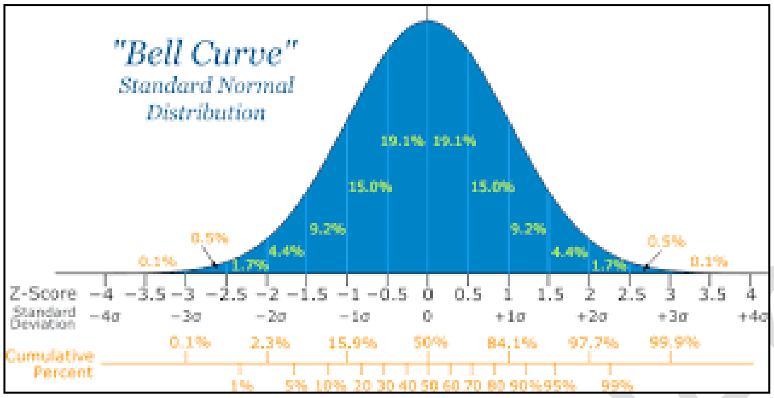
e = Euler's constant ≈ 2.71828

For example, suppose that the mean number of calls arriving in a 15-minute period is 10.To compute the probability that 5 calls come in within the next 15 minutes,

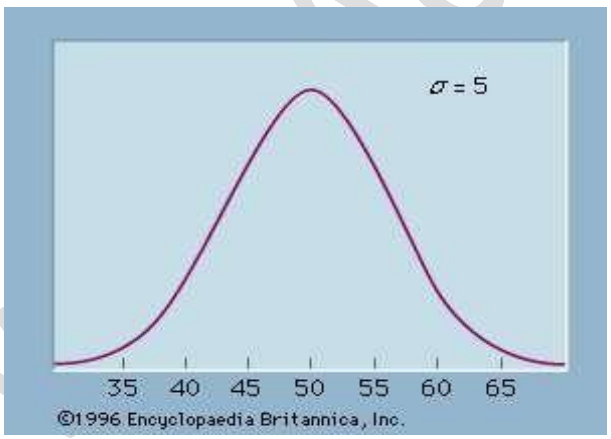
$\lambda = 10$ and $x = 5$ are substituted in equationWe

get the following probability of **0.0378**.

The Normal distribution



The most widely used **continuous probability distribution** in statistics is the normal probability distribution. The graph corresponding to a normal probability density functionwith a mean of μ = 50 and a standard deviation of σ = 5 is shown in Figure .



Like all normal distribution graphs, it is a bell-shaped curve. Probabilities for the normal probability distribution can be computed using statistical tables for the standard normal probability distribution, which is a normal probability distribution with a mean of zero and a standard deviation of one.

A simple mathematical formula is used to convert any value from a normal probability distribution with mean μ and a standard deviation σ into a corresponding value for a standard normal distribution. The tables for the standard normal distribution are then usedto compute the appropriate probabilities.

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

$$\begin{aligned} \mu &= \text{ Mean} \\ \sigma &= \text{ Standard Deviation} \\ \pi &\approx 3.14159\cdots \\ e &\approx 2.71828\cdots \end{aligned}$$

The above function is used to calculate the area under the bell curve using the concept of definite integration of it with proper limits, which gives us the probability of the Normal Distribution.

Normal Distribution is also called Gaussian Distribution, based on the mathematician who contributed to it.

Hypothesis Test

A hypothesis is an educated guess about something in the world around you. It should betestable, either by experiment or observation. For example:

- A new medicine you think might work.
- A way of teaching you think might be better.
- A possible location of new species.
- A fairer way to administer standardized tests.

It can really be *anything at all* as long as you can put it to the test.

Hypothesis testing in statistics is a way for you to test the results of a survey or experiment to see if you have meaningful results. You're basically testing whether your results are valid by figuring out the odds that your results have happened by chance. If your results may have happened by chance, the experiment won't be repeatable and so has little use.

Null Hypothesis

The null hypothesis is a statement that you want to test. In general, the null hypothesis is that things are the same as each other, or the same as a theoretical expectation.

Alternate Hypothesis

The alternative hypothesis is that things are different from each other, or different from a theoretical expectation.
Most used hypothesis testing method is Pearson's method or Chi square Method

Chi-Square ($\chi 2$) Statistic Definition
A chi-square (χ^2) statistic is a test that measures how expectations compare to actual observed data (or model results). The data used in calculating a chi-square statistic must be random, raw, mutually exclusive, drawn from independent variables, and drawn from a large enough sample.
For example, the results of tossing a coin 100 times meet these criteria.

A p- value is used in hypothesis testing to help you support or reject the null hypothesis. Thep value is the evidence against a null hypothesis.

The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.
P values are expressed as decimals although it may be easier to understand what they are ifyou convert them to a percentage. For example, a p value of 0.0254 is 2.54%.

Therefore, the smaller the p-value, the more important ("significant") your results.
When you run a hypothesis test, you compare the p value from your test to the alpha level you selected when you ran the test. Alpha levels can also be written as percentages.
The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk ofconcluding that a difference exists when there is no actual difference.

The aim is to prove that our H_o is TrueIn
Chi square Test

$\chi 2$ Value $>$ p-value REJECT H_0
$\chi 2$ Value $<$ p-value ACCEPT H_0

The 2 Values that is chi squared and p value are calculated Seperately, we will undertsnausing an example
In chi square test we always take the Significance level to be **5% or 0.05.**
In a test containing MCQ, the following data was found

	Option A	Option B	Option C	Option D
Expected Value	25%	25%	25%	25%
Observed Value	20%	20%	25%	35%

The row with expected values give the percentage of the marks that were expectedThe second row gives the values that were observed for each option in a test.

So for the above dataset we calculate the whether to accept or reject the hypothesis based on the person's test , so before we move on we have to define the Null hypothesis and the Alternate hypothesis:-

H_0 = All the options have equal chance of coming as an answer. H_1 = the options don't have equal chance of coming as an answer.The formula used to calculate the Chi square value is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Using this formula we get the following the value of $\chi 2$ which is

$$X^2 = 6$$

The other important thing to consider is the Degree of freedom which is calculated using the following formula

DoF= (No. of row-1) (No. of Column-1)

So, DoF= (2-1)(4-1) =3

Now, to find the p-value corresponding to the alpha value and degree of freedom we refer to the table used as default.

Check the row corresponding to the dof and column corresponding to the 0.05 alpha levelfrom

Degreeof Freedom

	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21

which we get p- value = 7.815

So, $\chi 2$ Value < p- value

Therefore we cannot reject the Null hypothesis and say that the options were equallydistributed, D option didn't appeared more often.