Linear Algebra

Video:1

Magnitude = Scaler i.e value
Direction = Velocity
Example = 40m/s south
Magnitude is 40 & direction is south.

More examples: 4i+9j = 4m/s in x-direction & 9m/s in y-direction

Types of Vectors

1) Zero Vector	Vector with zero magnitude
2) Unit Vector	Vector whose magnitude is 1 unit
3) Cointial Vector	Two or more vectors with same initial point
4) Collinear Vector	Two or more vectors lying on the same or parallel lines.
5) Equal Vectors	Two or more vectors with same magnitude and direction.
6) Negative Vectors	Vector with same magnitude but opposite direction as that of the given vector.

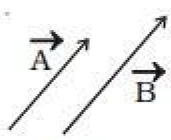


Fig.
Like vectors

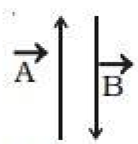


Fig.
Opposite vectors

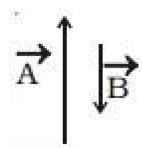


Fig.
Unlike Vectors

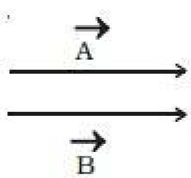


Fig.
Equal vectors

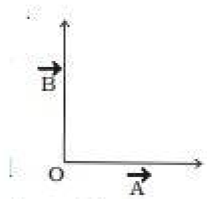
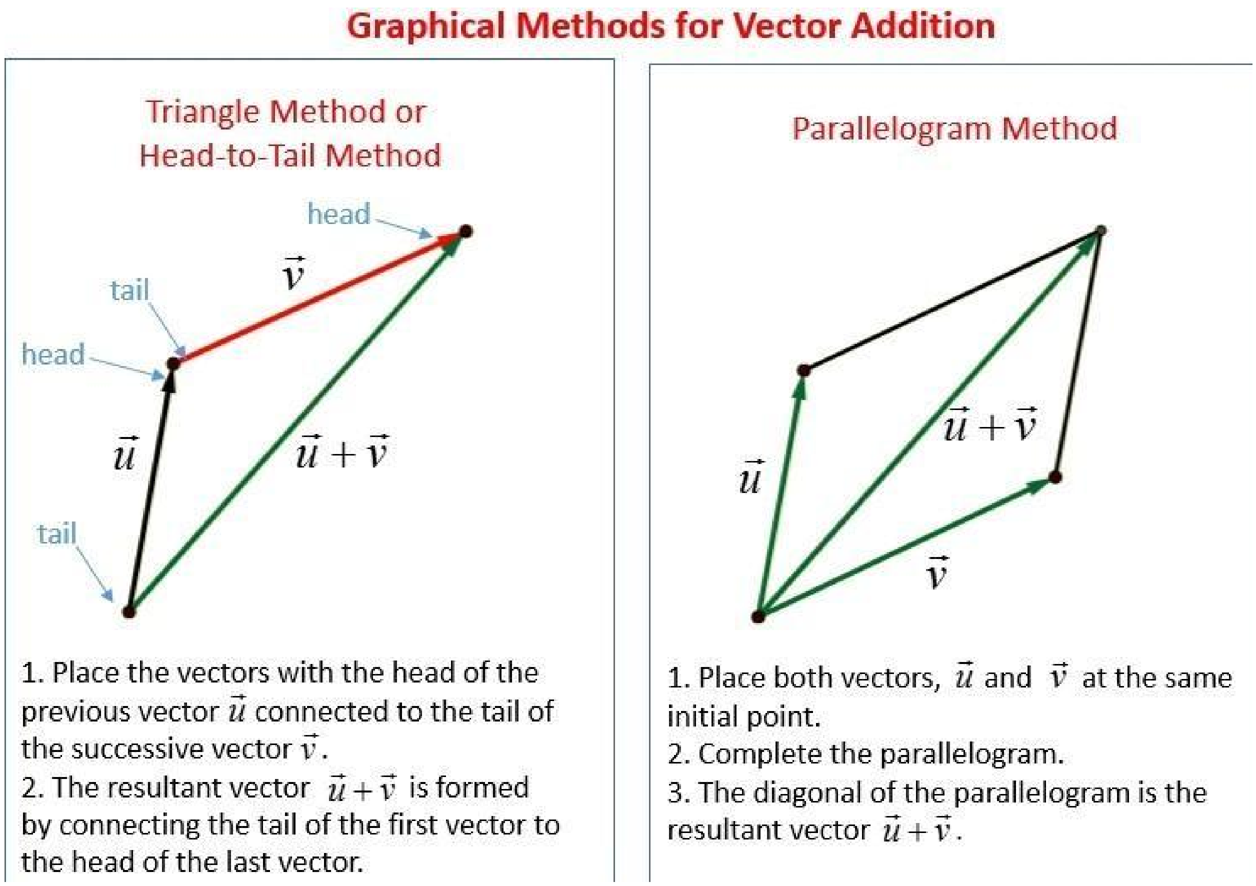


Fig
Co-initial vectors

Video:2

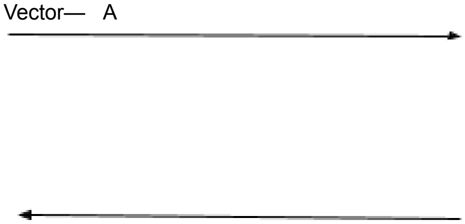
Vector Addition:



Vector Subtraction:

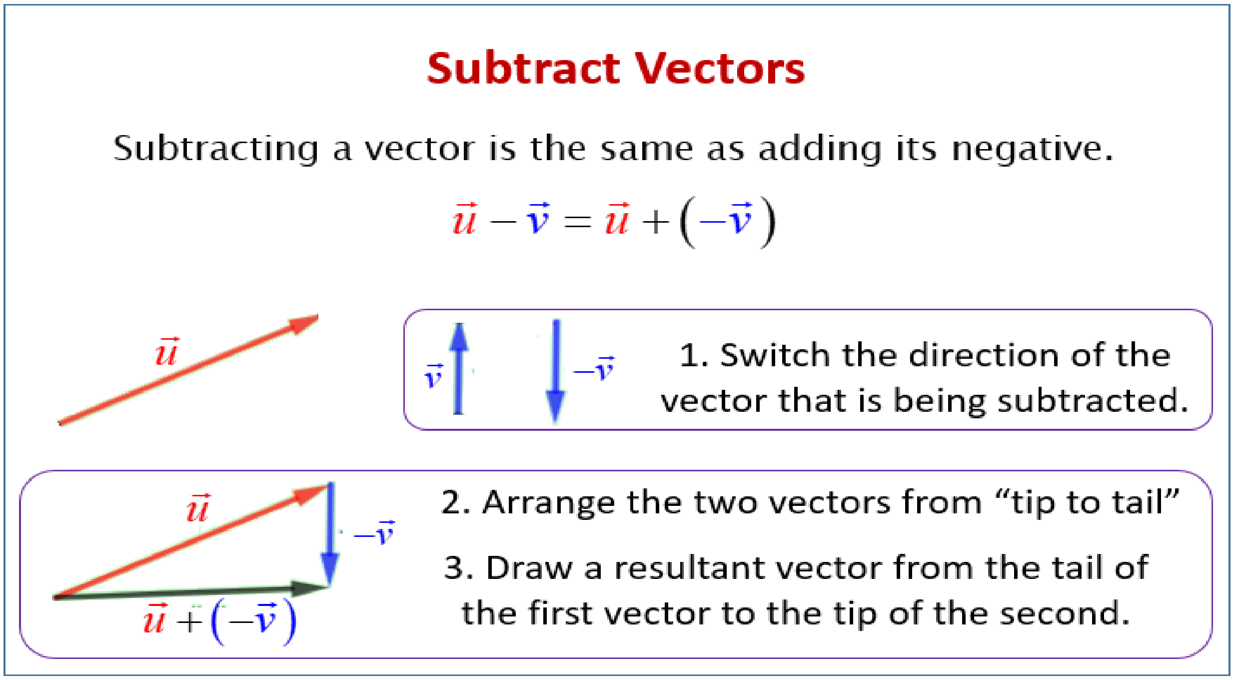
Note-

Vector— A



Vector— –A

When we want negative of a vector we just need to reverse the vector.



Video:3

Properties:

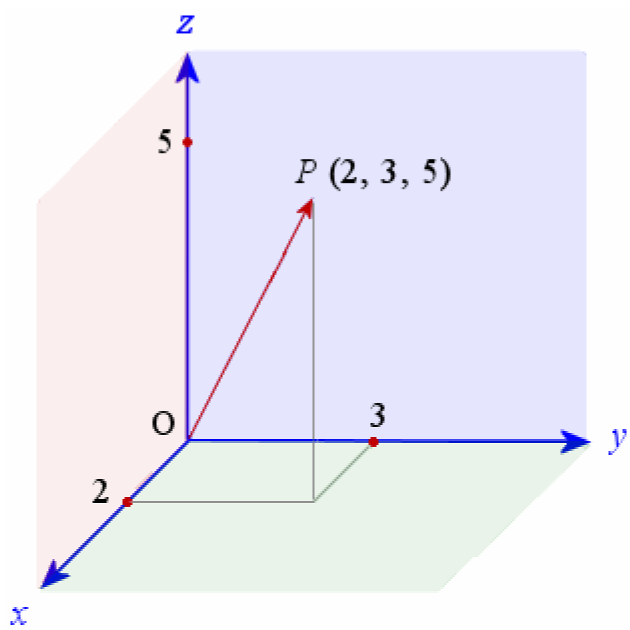
1. Commutative: u+v = v+u (where u and v are vectors)
2. Associative: u+(v+w) = (u+v)+w (where u,v,w are vectors)
3. Identity: u+0 = u
4. Additive Inverse: u+(-u) = 0

Note:

2-D vector is like 2i+3j i.e 2 units in x-direction and 3 units in y-direction

3-D vector is like 2i+3j+5k i.e 2 units in x-direction,3 units in y-direction and 5 units in z direction respectively.

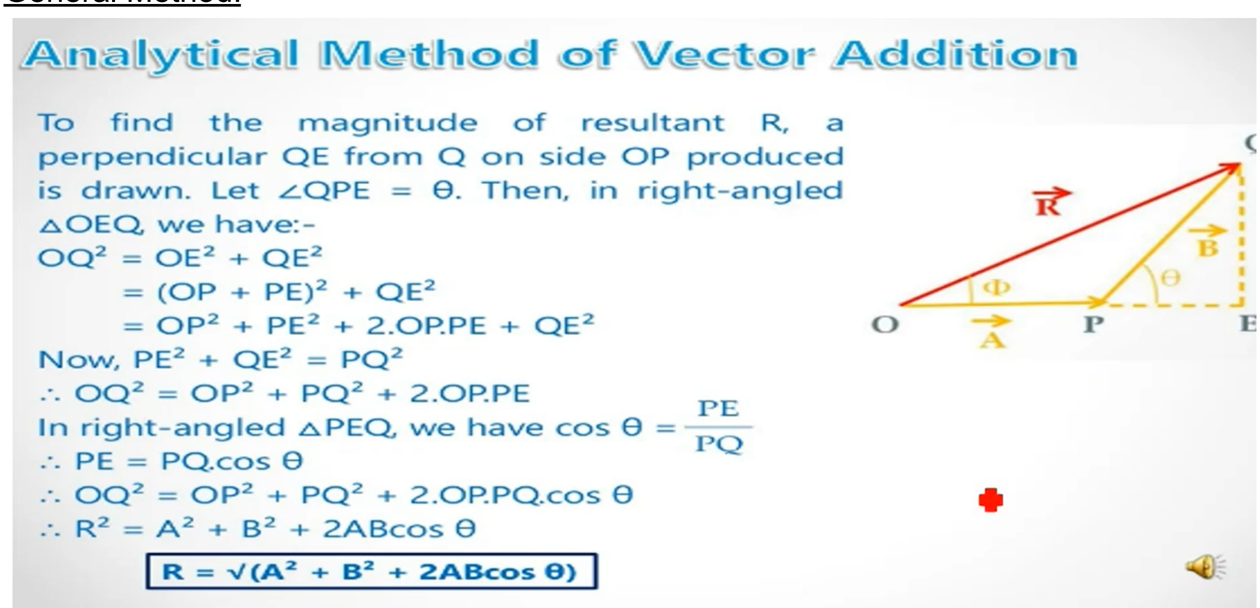
More example:
Suppose a vector has magnitude 5 then it can be 3i+4j or it can also be 4i+3j.



Video:4

Vector Addition:

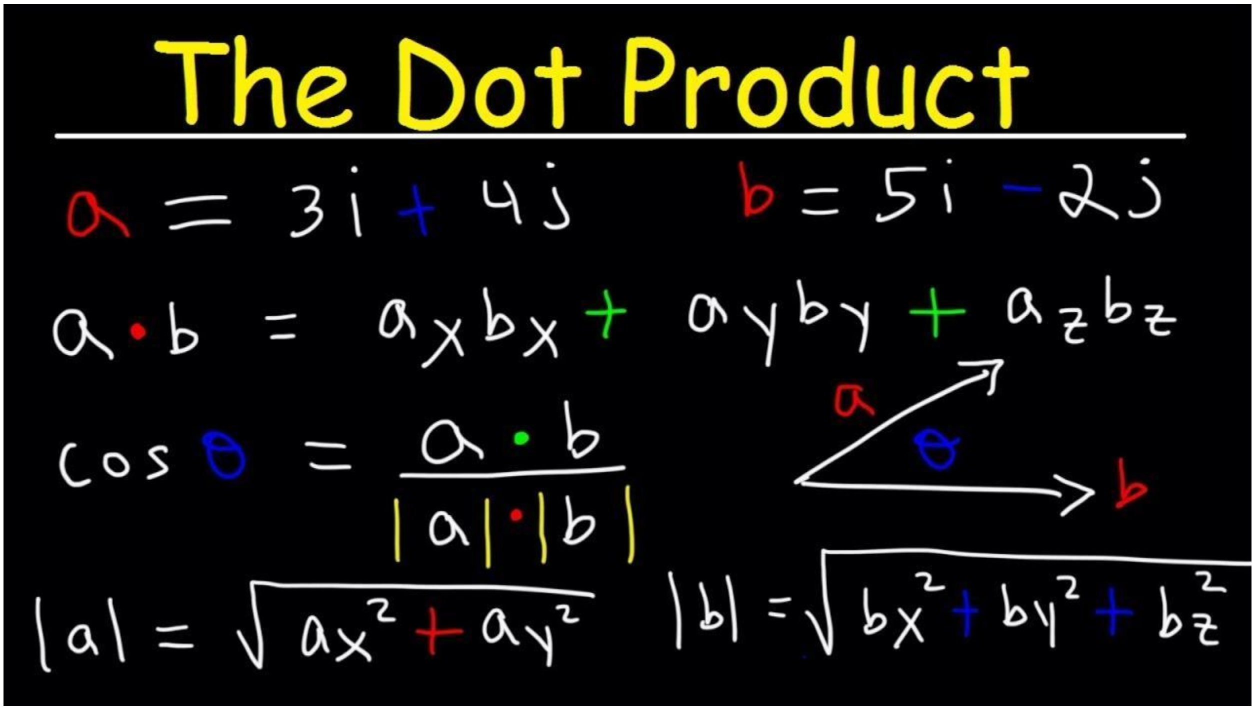
General Method:



Note: When angle between 2 vectors is 90 degrees formula is square root of a square + b square.

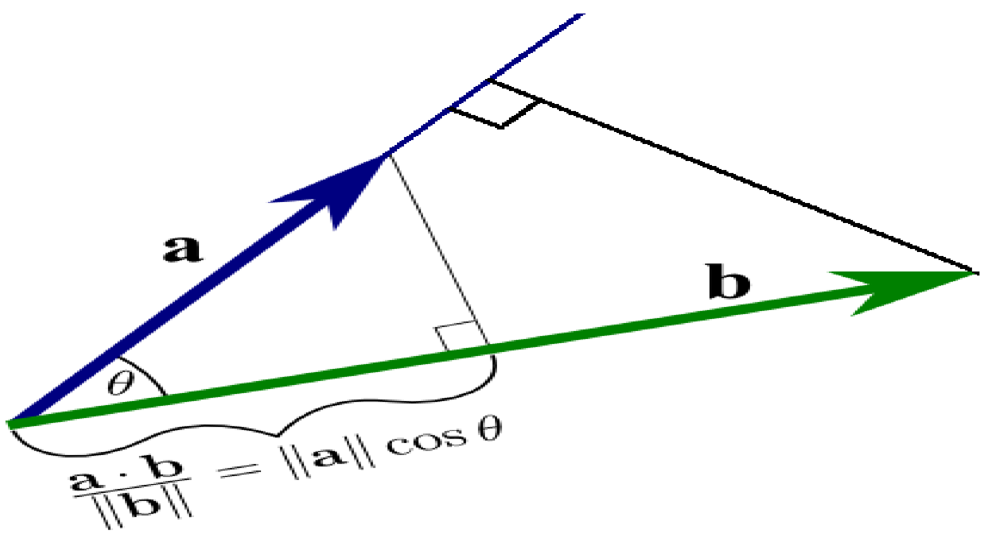
Video:5

Dot Product:



Note:

To find the projection of any vector on another use this:



Video 6&7: Python functions of dot product.

```
In [1]:  import numpy as np
         from warnings import filterwarnings
         filterwarnings('ignore')
```

```
In [2]:  var_1, var_2 = 34, 45
         dot_product_1 = np.dot(var_1, var_2)
         dot_product_1

Out[2]:  1530
```

```
In [3]:  a = np.array([[2, 3, 4], [-1, 3, 2], [9, 4, 8]])
         b = np.array([[4, -1, 2], [34, 9, 1], [2, 0, 9]])
         dot_product_2 = np.dot(a, b)
         dot_product_2

Out[3]:  array([[118,  25,  43],
                [102,  28,  19],
                [188,  27,  94]])
```

Video:8

L1 Regularization -- Lasso Regression

L2 Regularization – Ridge Regression

There are many possible norms on $\mathbb{R}^n$.

Examples: consider a vector $\mathbf{x}$ in $\mathbb{R}^n$ with elements $\mathbf{x} = (x_1, x_2, \ldots x_n)$. Then:

1. The Euclidean distance, or l_2 norm:
   $||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
2. The l_∞ or *sup* norm:
   $||x||_\infty = \max_{1 \leq i \leq n} |x_i|.$
3. The l_1 norm:
   $$||x||_1 = \sum_{i=1}^{n} |x_i|$$

Note: Refer to the python functions of L1 and L2 Norms.

Video:9

When 2 vectors are orthogonal it means that the angle between the vector's is 90 degree's.
(Apply this concept for changing the co-ordinate system)

Video:10

Linear Independent Vectors

Linear Dependence and Independence

❑ Definition:

The set of vectors $S = \{ X_1, X_2, \cdots, X_k \}$ in R^n is linearly dependent if there exist scalars $c_1, c_2, \cdots, c_k \in R$, not all zeros, such that

$$c_1 X_1 + c_2 X_2 + \cdots + c_k X_k = O.$$

Otherwise the vectors $X_1, X_2, \cdots, X_k$ are linearly independent.

That is, $X_1, X_2, \cdots, X_k$ are linearly independent whenever $c_1 X_1 + c_2 X_2 + \cdots + c_k X_k = O$ we must have $c_1 = c_2 = \cdots = c_k = 0$.

e.g. The vectors $X_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ are linearly independent

since $c_1 X_1 + c_2 X_2 + c_3 X_3 = O \Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \Rightarrow c_1 = c_2 = c_3 = 0.$

Video:11

Basis Vector

If you can write every vector in a given space as a Linear combination of some vectors and these vectors are independent of each other then we can call them as BASIS vector for the space.
(Note: Basis vectors are not unique)

Video:12

Matrices Introduction

Matrices - Introduction

Matrix algebra has at least two advantages:

•Reduces complicated systems of equations to simple expressions

•Adaptable to systematic method of mathematical treatment and well suited to computers

Definition:

A matrix is a set or group of numbers arranged in a square or rectangular array enclosed by two brackets

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \qquad \begin{bmatrix} 4 & 2 \\ -3 & 0 \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$3 \cdot 0 + 1 \cdot 2 + 0 \cdot 0 = 2$$

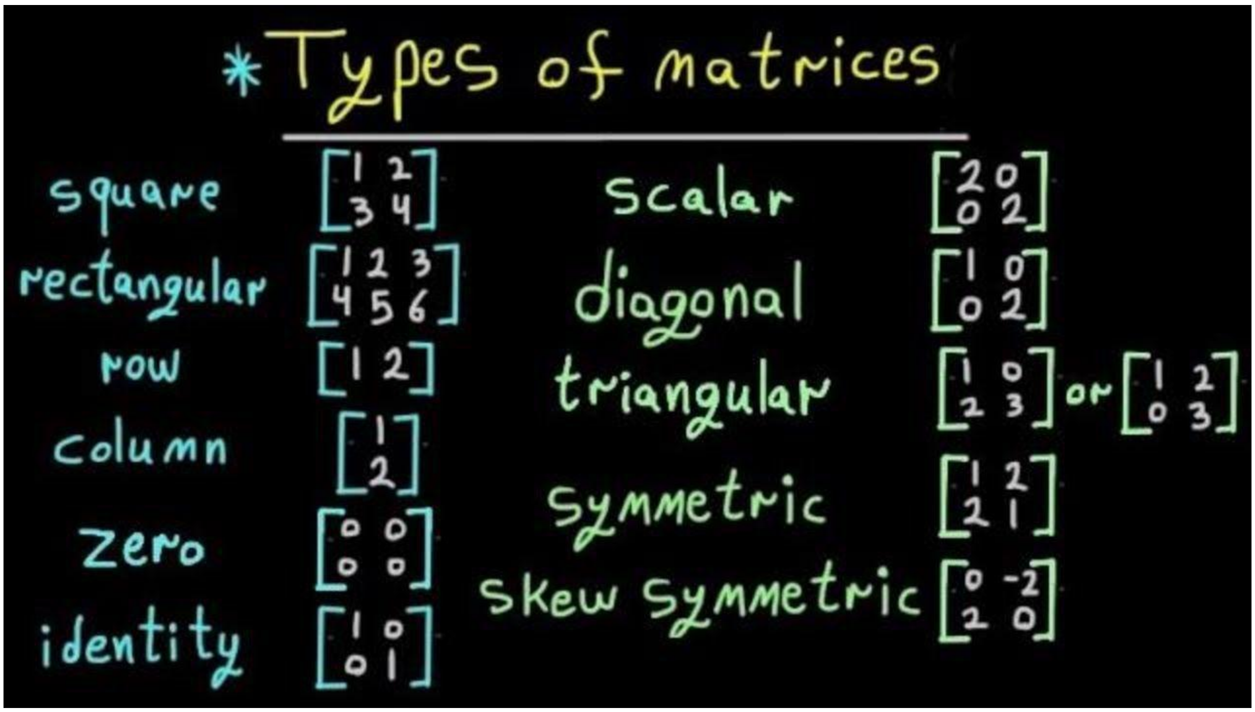
$$\begin{bmatrix} 3 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 4 \\ 2 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & \end{bmatrix}$$

$$2 \times 3 \qquad 3 \times 2 \qquad 2 \times 2$$

Note: Similarly you can find the other outputs.

<u>Video:13</u>

Type of Matrices:



Note: Matrices Multiplication

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}}_{2 \times 3} \times \underbrace{\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}}_{3 \times 1} = \underbrace{\begin{bmatrix} (1)(7)+(2)(8)+(3)(9) \\ (4)(7)+(5)(8)+(6)(9) \end{bmatrix}}_{2 \times 1} = \underbrace{\begin{bmatrix} 7+16+27 \\ 28+40+54 \end{bmatrix}}_{2 \times 1} = \underbrace{\begin{bmatrix} 50 \\ 122 \end{bmatrix}}_{2 \times 1}$$

columns on 1st = rows on 2nd

The number of rows in the 1st matrix and the number of columns in the 2nd matrix, make the dimensions of the final matrix

Properties of Matrices Multiplication:

THEOREM 2.1 Properties of Matrix Addition and Scalar Multiplication

If A, B, and C are $m \times n$ matrices, and c and d are scalars, then the properties below are true.

1. $A + B = B + A$ — Commutative property of addition

2. $A + (B + C) = (A + B) + C$ — Associative property of addition

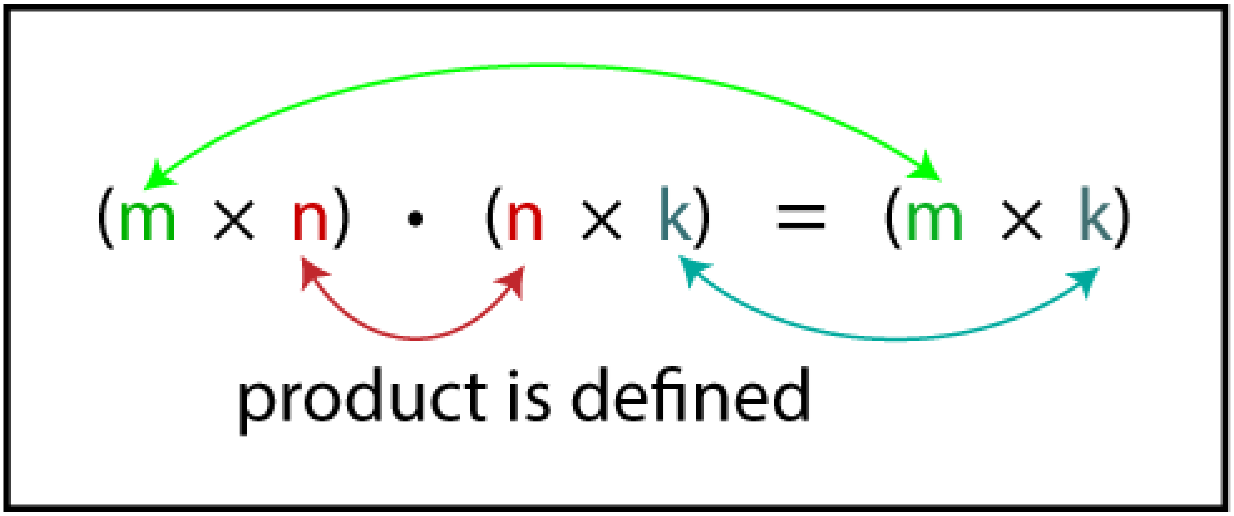
3. $(cd)A = c(dA)$ — Associative property of multiplication

4. $1A = A$ — Multiplicative identity

5. $c(A + B) = cA + cB$ — Distributive property

6. $(c + d)A = cA + dA$ — Distributive property

Note: Trick to find size of Matrix



Video:14

Rank of Matrix:

The maximum number of linearly independent columns (or rows) of a matrix is called the rank of a matrix. The rank of a matrix cannot exceed the number of its rows or columns.

$$A = \begin{bmatrix} 4 & 2 & 1 & 3 \\ 6 & 3 & 4 & 7 \\ 2 & 1 & 0 & 1 \end{bmatrix}, \text{ Find rank ?}$$

(1).

$C_3 \leftrightarrow C_1$

$$\begin{bmatrix} 1 & 2 & 4 & 3 \\ 4 & 3 & 6 & 7 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

(2).

$R_2 = R_2 - 4R_1$

$$\begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -5 & -10 & -5 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

$4 - 4 = 0$

$3 - 8 = -5$

$6 - 16 = -10$

$7 - 12 = -5$

(3).

$R_3 = R_3 + \frac{R_2}{5}$

$$\begin{bmatrix} 1 & 2 & 4 & 3 \\ 0 & -5 & -10 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$1 - 1$

$2 - 2$

$1 - 1$

$\therefore \rho(A) = 2$

Note: Please refer to some more examples to find rank of a matrix.

Video: 15

Null space for matrix:

The null space of any matrix A consists of all the vectors B such that AB = 0 and B is not zero. It can also be thought as the solution obtained from AB = 0 where A is known matrix of size m x n and B is matrix to be found of size n x k.

EXAMPLE 3 Find a spanning set for the null space of the matrix

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

SOLUTION The first step is to find the general solution of $A\mathbf{x} = \mathbf{0}$ in terms of free variables. Row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ to *reduced* echelon form in order to write the basic variables in terms of the free variables:

$$\begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 \qquad - \; x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free. Next, decompose the vector giving the general solution into a linear combination of vectors where *the weights are the free variables*. That is,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \underset{\substack{\uparrow \\ \mathbf{u}}}{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}} + x_4 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}} + x_5 \underset{\substack{\uparrow \\ \mathbf{w}}}{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \tag{3}$$

Every linear combination of **u**, **v**, and **w** is an element of Nul A and vice versa. Thus $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a spanning set for Nul A. ■

<u>Video: 16 &17</u>

Linear Equation

System of Linear Equation

$$2.0x + 4.0y + 6.0z = 18$$
$$4.0x + 5.0y + 6.0z = 24$$
$$3.0x + 1y - 2.0z = 4$$

Matrix representation

$$A = \begin{bmatrix} 2.0 & 4.0 & 6.0 \\ 4.0 & 5.0 & 6.0 \\ 3.0 & 1.0 & -2.0 \end{bmatrix} \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad b = \begin{bmatrix} 18.0 \\ 24.0 \\ 4.0 \end{bmatrix}$$

Note: Refer to some more examples

$$A = \begin{bmatrix} 2 & 3 & 3 \\ 1 & -2 & 1 \\ 3 & -1 & -2 \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, B = \begin{bmatrix} 5 \\ -4 \\ 3 \end{bmatrix}.$$

We find $|A| = \begin{vmatrix} 2 & 3 & 3 \\ 1 & -2 & 1 \\ 3 & -1 & -2 \end{vmatrix} = 2(4+1) - 3(-2-3) + 3(-1+6) = 10 + 15 + 15 = 40 \neq 0$.

So, A^{-1} exists and

$$A^{-1} = \frac{1}{|A|}(\text{adj } A) = \frac{1}{40}\begin{bmatrix} +(4+1) & -(-2-3) & +(-1+6) \\ -(-6+3) & +(-4-9) & -(-2-9) \\ +(3+6) & -(2-3) & +(-4-3) \end{bmatrix}^T = \frac{1}{40}\begin{bmatrix} 5 & 3 & 9 \\ 5 & -13 & 1 \\ 5 & 11 & -7 \end{bmatrix}$$

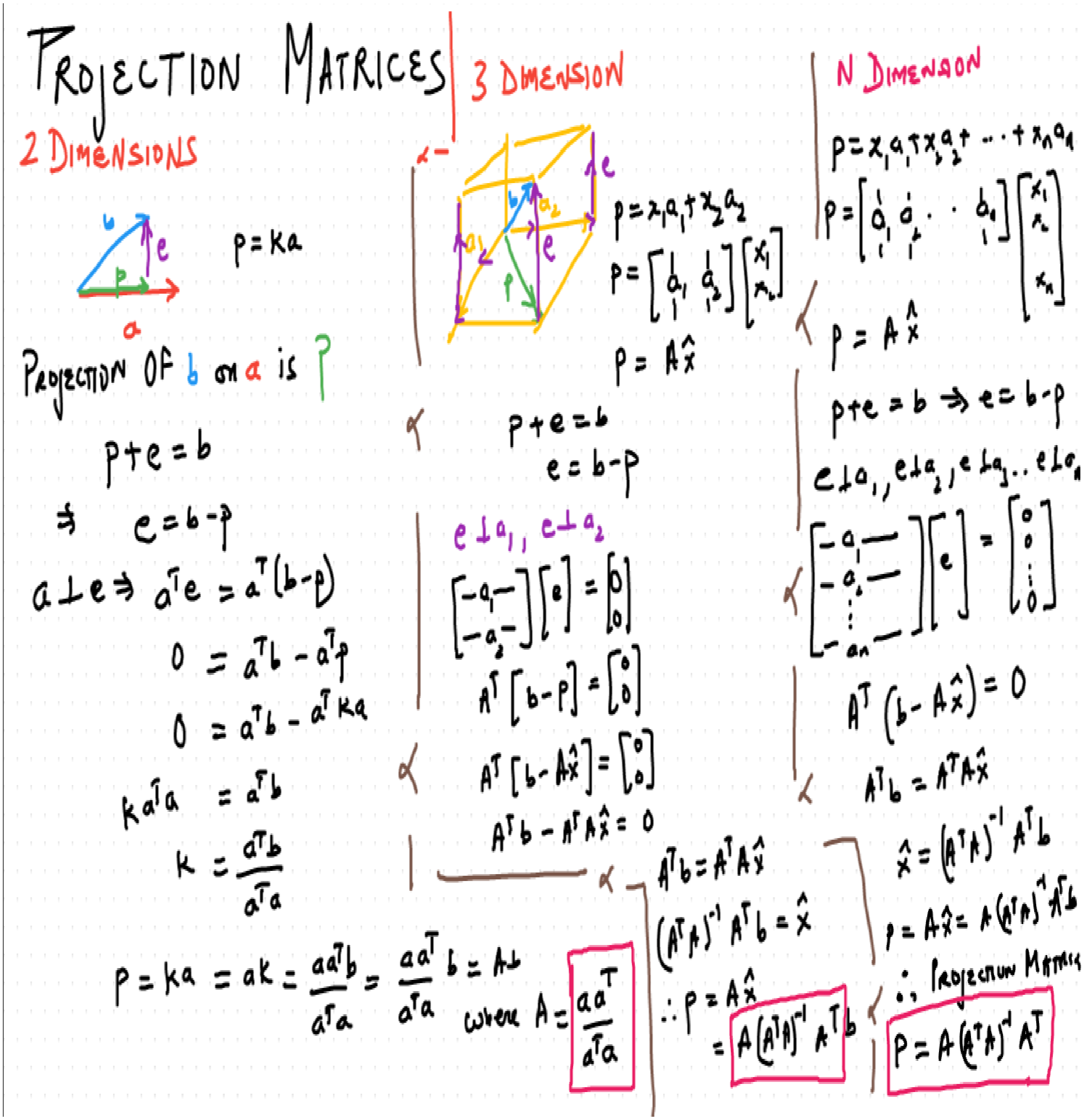
Then, applying $X = A^{-1}B$, we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{40}\begin{bmatrix} 5 & 3 & 9 \\ 5 & -13 & 1 \\ 5 & 11 & -7 \end{bmatrix}\begin{bmatrix} 5 \\ -4 \\ 3 \end{bmatrix} = \frac{1}{40}\begin{bmatrix} 25-12+27 \\ 25+52+3 \\ 25-44-21 \end{bmatrix} = \frac{1}{40}\begin{bmatrix} 40 \\ 80 \\ -40 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

Video: 20

Vector Line Equation and Projection

General Equations

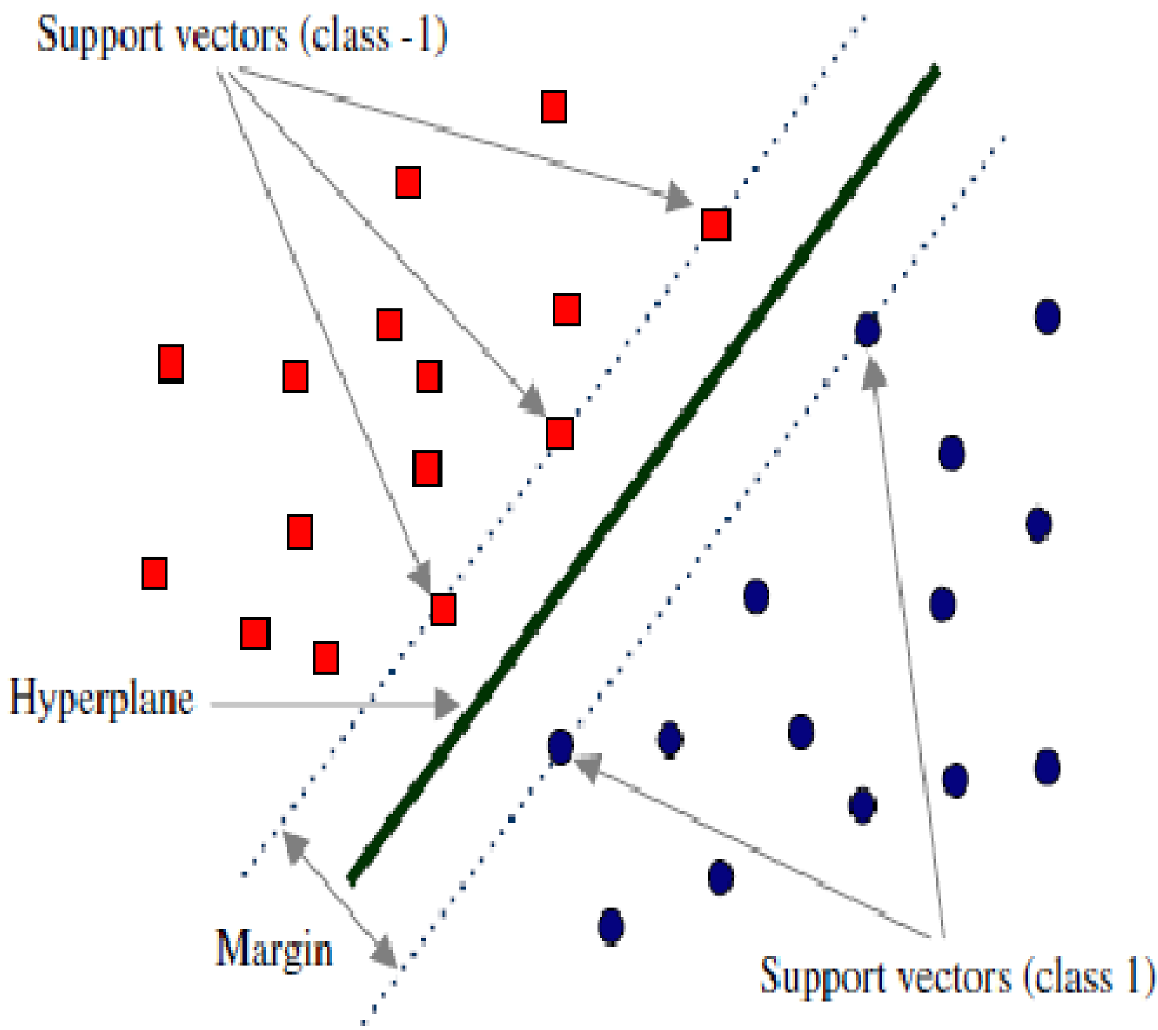


Video: 21

Hyper-Planes

What is hyperplane used for?

Hyperplanes are often used in classification algorithms such as support vector machines (SVMs) and linear regression to separate data points belonging to different classes. They are also used in clustering algorithms to identify clusters of data points in the input space.



Video: 22

Eigen Values and Eigen Vectors

General Form:

If

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

then the characteristic equation is

$$\left| A - \lambda.I = \left| \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| \right| = 0$$

$$\left| \begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix} \right| = \lambda^2 + 3\lambda + 2 = 0$$

and the two eigenvalues are

$$\lambda_1 = -1, \lambda_2 = -2$$

Eigenvalues are the special set of scalar values that is associated with the set of linear equations most probably in the matrix equations. The eigenvectors are also termed as characteristic roots. It is a non-zero vector that can be changed at most by its scalar factor after the application of linear transformations.

Eigen Vectors:

The eigenvector is a vector that is associated with a set of linear equations. The eigenvector of a matrix is also known as a latent vector, proper vector, or characteristic vector. These are defined in the reference of a square matrix.

Applications in Machine Learning:

They are used to reduce dimension space. The technique of Eigenvectors and Eigenvalues are used to compress the data. As mentioned above, many algorithms such as PCA rely on eigenvalues and eigenvectors to reduce the dimensions.

Video: 23

Properties of Eigen values and Eigen vectors

6.4 Properties of Eigenvalues and Eigenvectors

Definition: The trace of a matrix A, designated by tr(A), is the sum of the elements on the main diagonal.

Property 1: The sum of the eigenvalues of a matrix equals the trace of the matrix.

Property 2: A matrix is singular if and only if it has a zero eigenvalue.

Property 3: The eigenvalues of an upper (or lower) triangular matrix are the elements on the main diagonal.

Property 4: If λ is an eigenvalue of **A** and **A** is invertible, then $1/\lambda$ is an eigenvalue of matrix A^{-1}.

6.4 Properties of Eigenvalues and Eigenvectors

Property 5: If λ is an eigenvalue of **A** then $k\lambda$ is an eigenvalue of **kA** where **k** is any arbitrary scalar.

Property 6: If λ is an eigenvalue of **A** then λ^k is an eigenvalue of A^k for any positive integer **k**.

Property 8: If λ is an eigenvalue of **A** then λ is an eigenvalue of A^T.

Property 9: The product of the eigenvalues (counting multiplicity) of a matrix equals the determinant of the matrix.

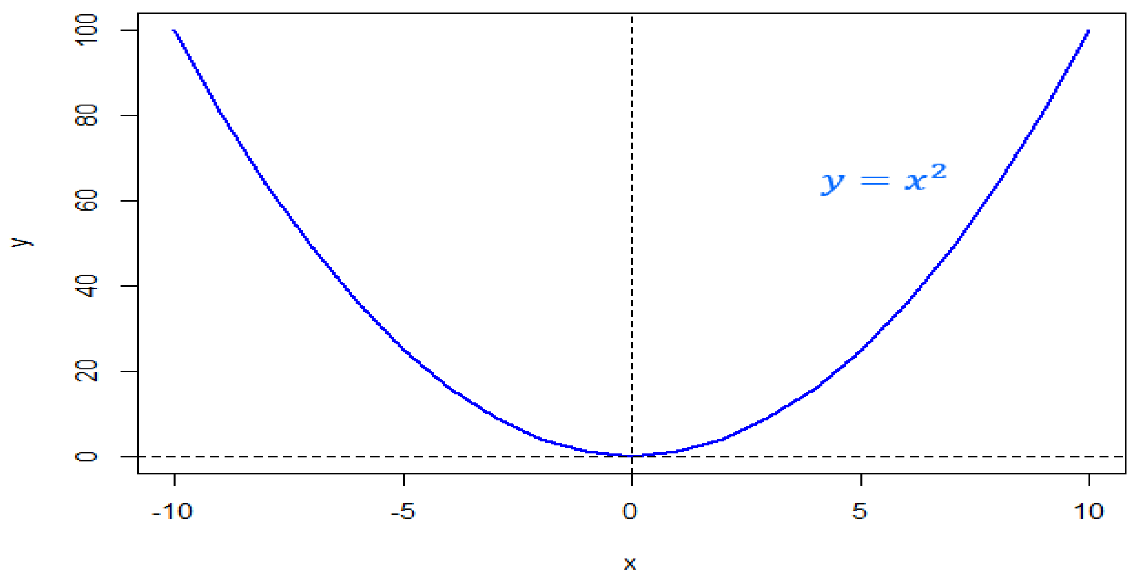
CALCULUS

<u>Video: 1 to 3:</u>

Functions

General Equation: $y = F(x)$; Where $F(x)$ = Independent Variable

y = Dependent Variable



General form to find the derivative:

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

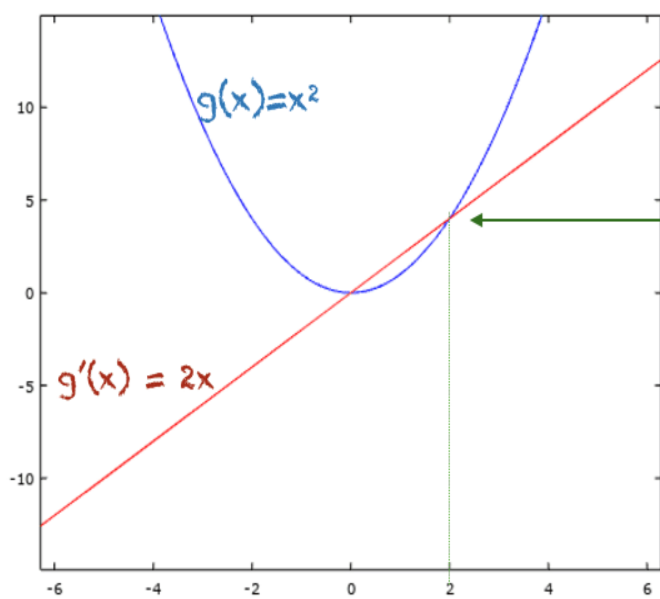
$y' = 2x$

Derivative represents the slope of the particular equation

y’’ = 2 (double derivative)
Advanced Derivative Concept:

$$
\begin{aligned}
g(x) &= x^2 \\
g(x + \Delta x) &= (x + \Delta x)^2 \\
&= x^2 + 2x\Delta x + (\Delta x)^2 \\
g'(x) &= \lim_{\Delta x \to 0} \frac{g(x + \Delta x) - g(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{2x + \Delta x}{\Delta x} \\
&= 2x
\end{aligned}
$$

Graphical Illustration



Example: The rate of change of g(x) at x=2 is the value of g'(x) at x=2, i.e., g'(2) = 4

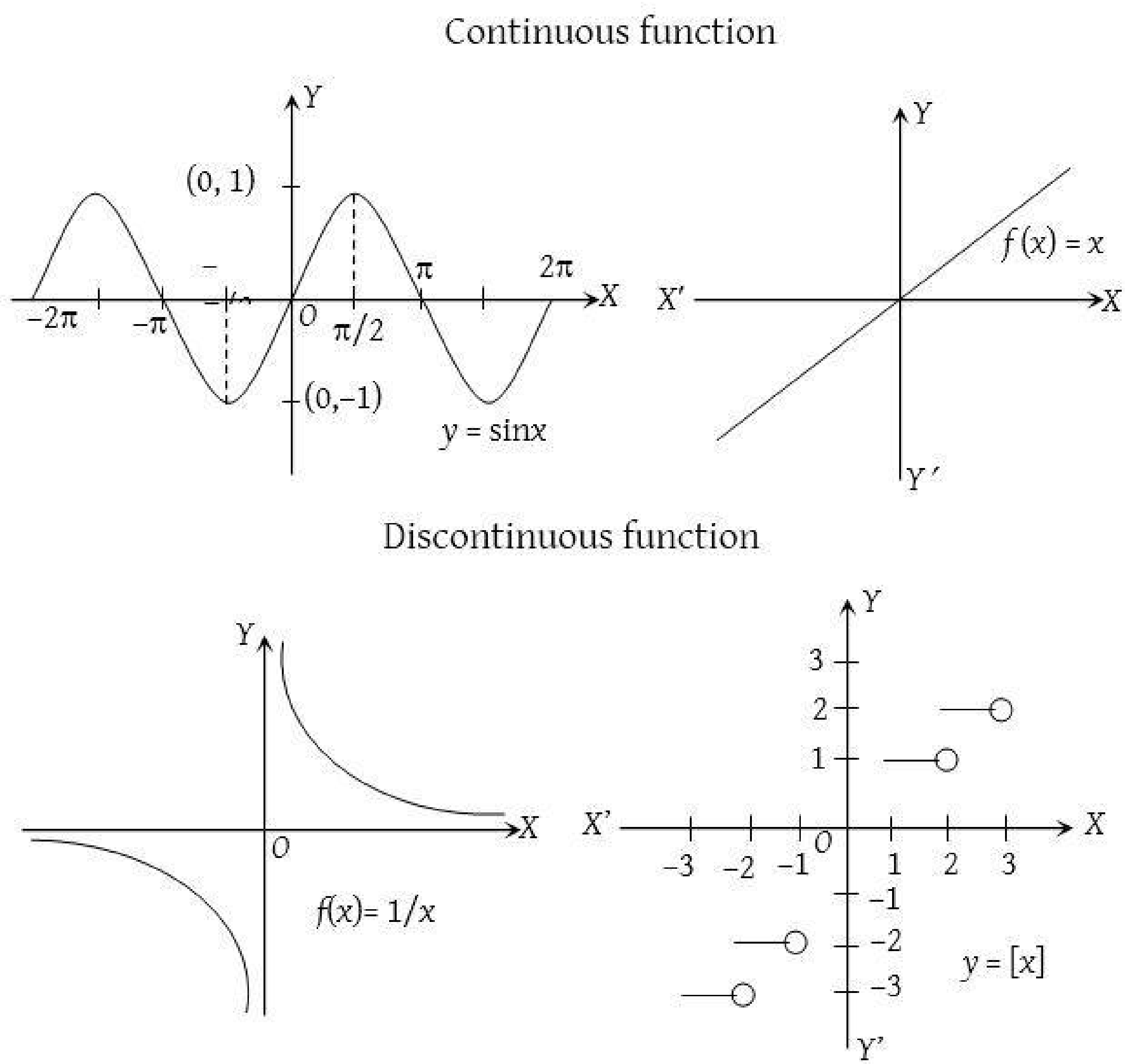
The rate of change is positive when x>0, negative when x<0 and 0 when x=0

Note: In General we are finding these derivatives to get the minimum and maximum values according to machine learning (as most of the ML Algorithms use the concept of gradient descent which is to find the point where sum of errors is minimum.

<u>Video: 4</u>

Continuous function

Continuous functions are functions that have no restrictions throughout their domain or a given interval. Their graphs won't contain any asymptotes or signs of discontinuities as well.
Example:



Video: 5

Integration:

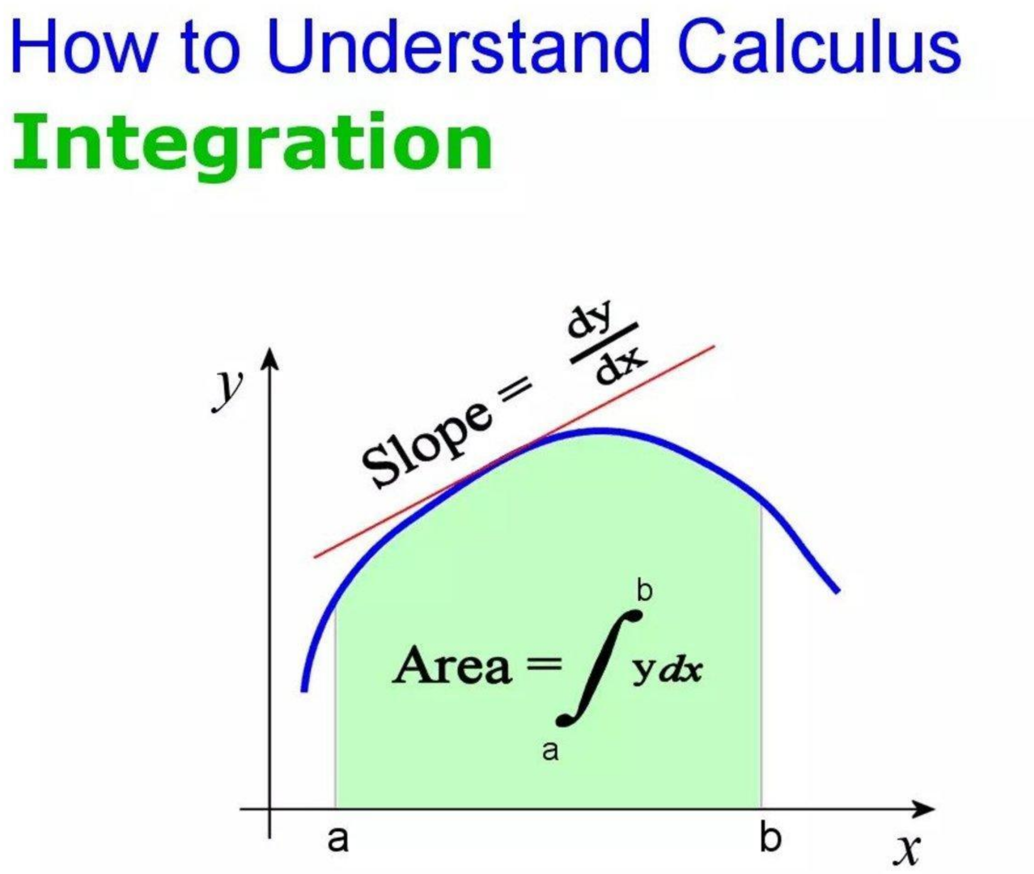
Integration is the inverse of differentiation. In other words, if you reverse the process of differentiation, you are just doing integration. The following example shows it: $y = x^2 \Rightarrow dy/dx = 2x$. So, $\int (dy/dx)\, dx = \int 2x\, dx = x^2$. $\int$ and dx go hand in hand and indicate the integration of the function with respective to x.

Example:

$$\int x^n \, dx = \frac{1}{n+1} x^{n+1} + c$$

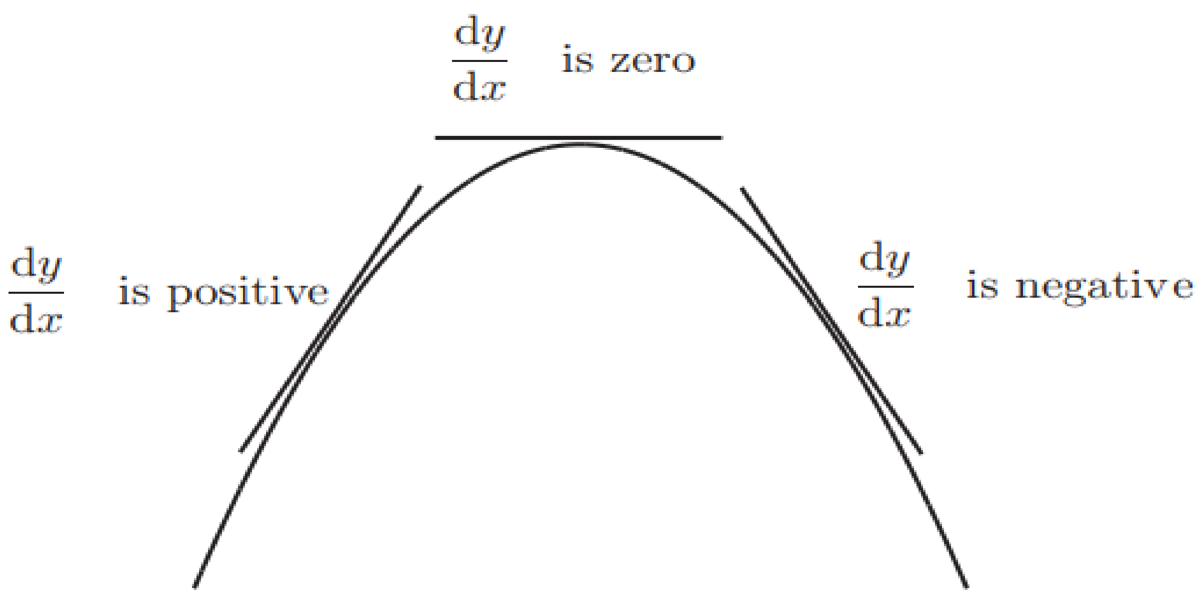
Example:

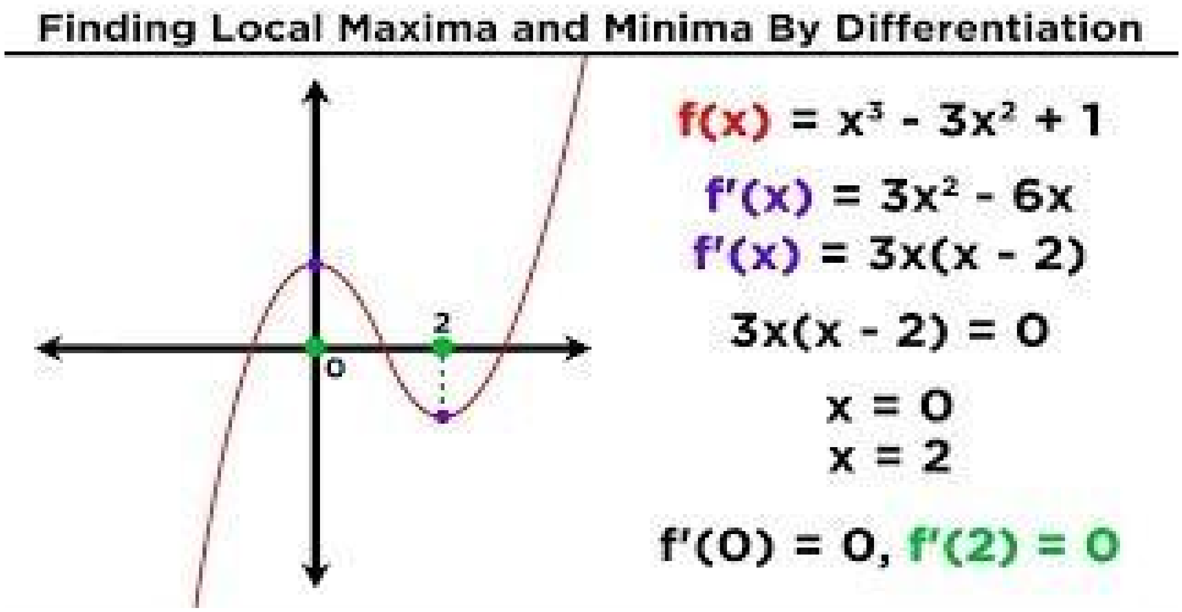
$$\int x^5 \, dx = \frac{1}{6} x^6 + c$$



Video: 6

Maxima and Minima:





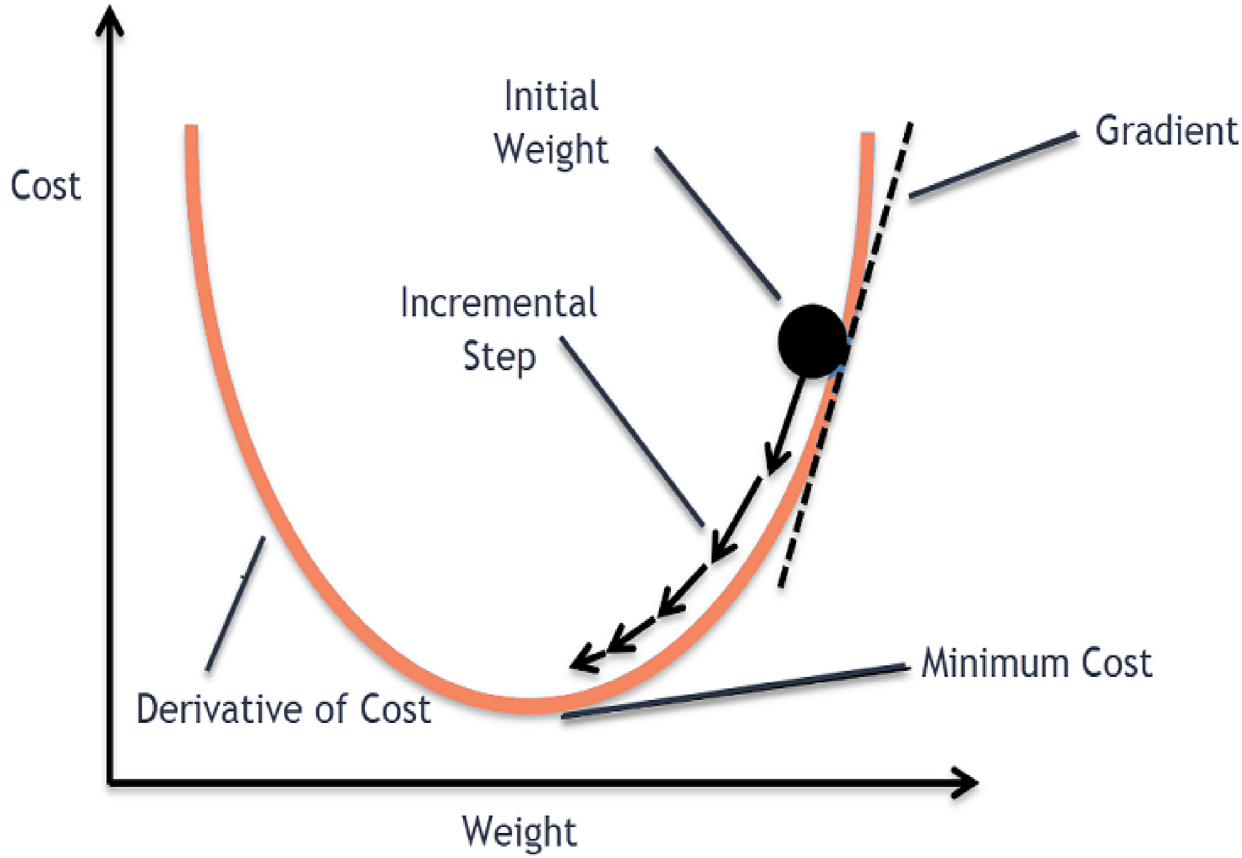
Video: 7

Gradient Descent:

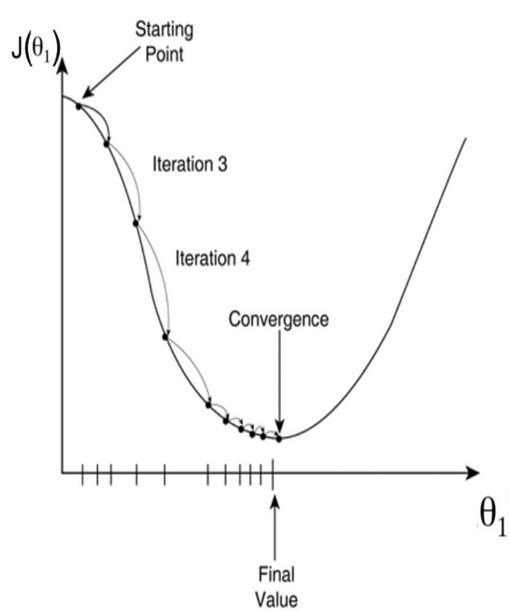
Gradient Descent is an optimization algorithm for finding a local minimum of a differentiable function. Gradient descent in machine learning is simply used to find the values of a function's parameters (coefficients) that minimize a cost function as far as possible.
(Used mostly in Linear regression and Logistic regression and Deep Learning to minimize the loss function and cost functions)

WHY GRADIENT DESCENT?
Gradient Descent is an algorithm that solves optimization problems using first-order iterations. Since it is designed to find the local minimum of a differential function, gradient descent is widely used in machine learning models to find the best parameters that minimize the model's cost function.

Application:



In Gradient Descent to need to move step by step(small steps) in order to reach the minima point.



Cost Function – "One Half Mean Squared Error":

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Objective:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Derivatives:

$$\frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x^{(i)}$$

Example of a simple gradient descent:

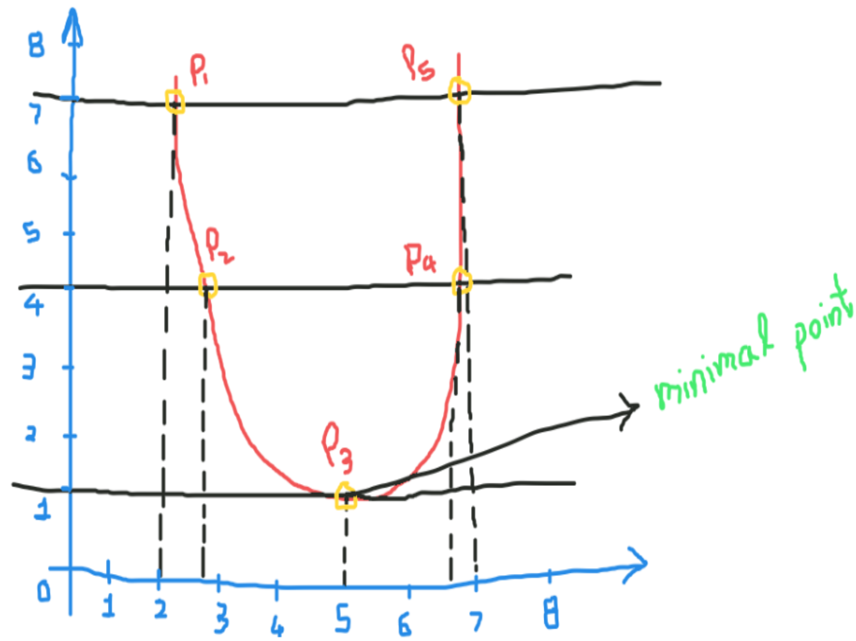
$P_1 = (2, 7)$

$P_2 = (2.8, 4)$

✓ $P_3 = (5, 1)$

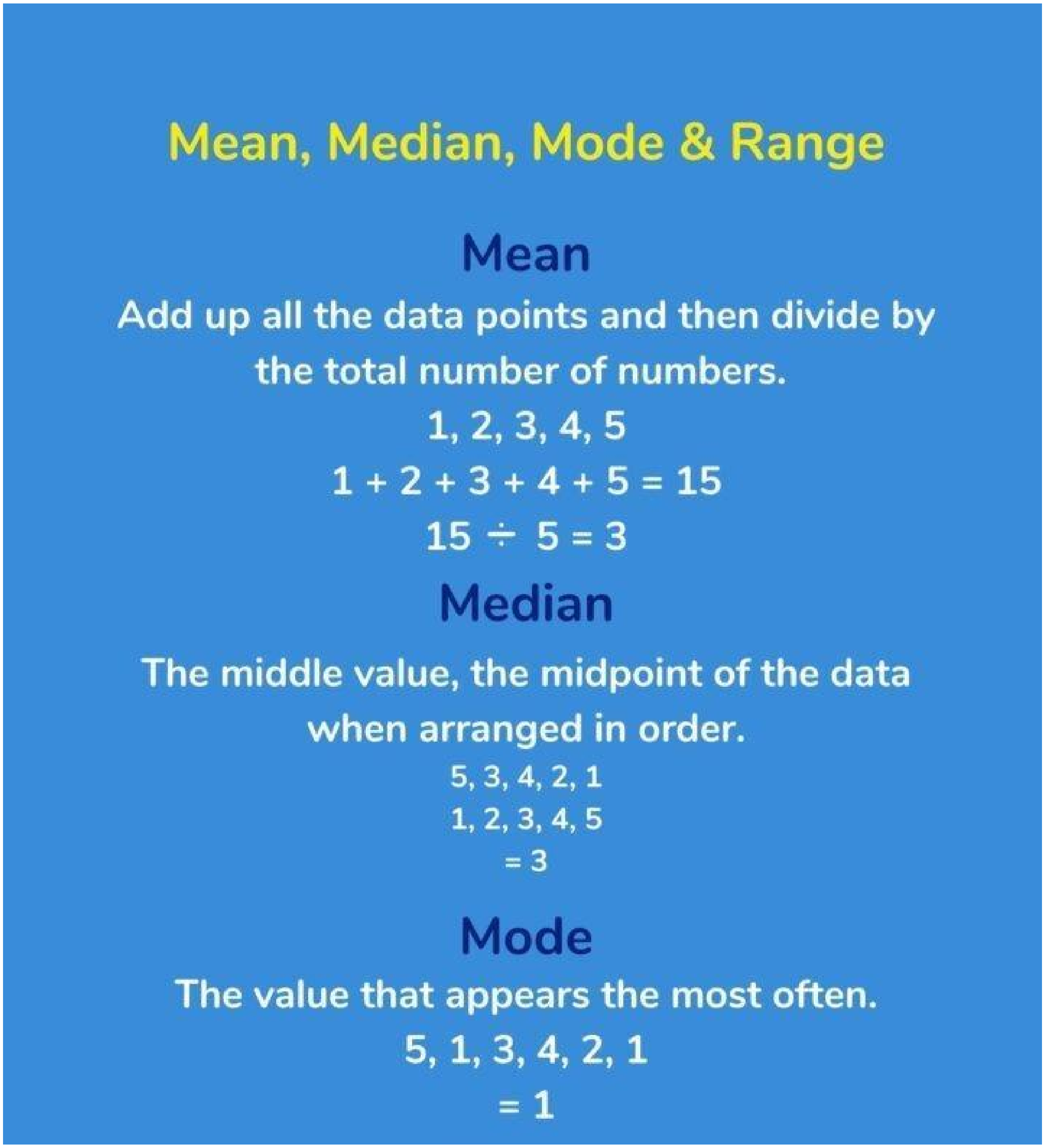
$P_4 = (6.8, 4)$

$P_5 = (7, 7)$



Statistics and probability

Mean, Median, Mode:



Population and sample mean:

A sample is a subset from the population.

Population Mean	Sample Mean
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$ N = number of items in the population	$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$ n = number of items in the sample

Population Variance & Sample variance:

Measure of dispersion

Variance

Sample variance

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

S^2=sample variance

x_i =value of i th element

$\bar{x}$ = sample mean

n=sample size

Population variance

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

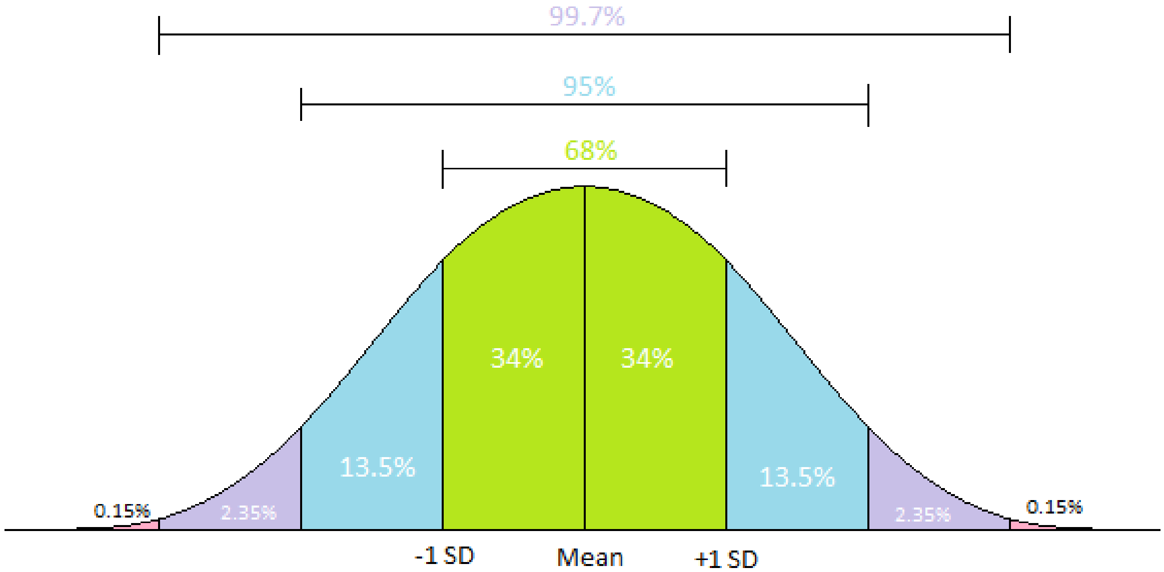
σ^2=population variance

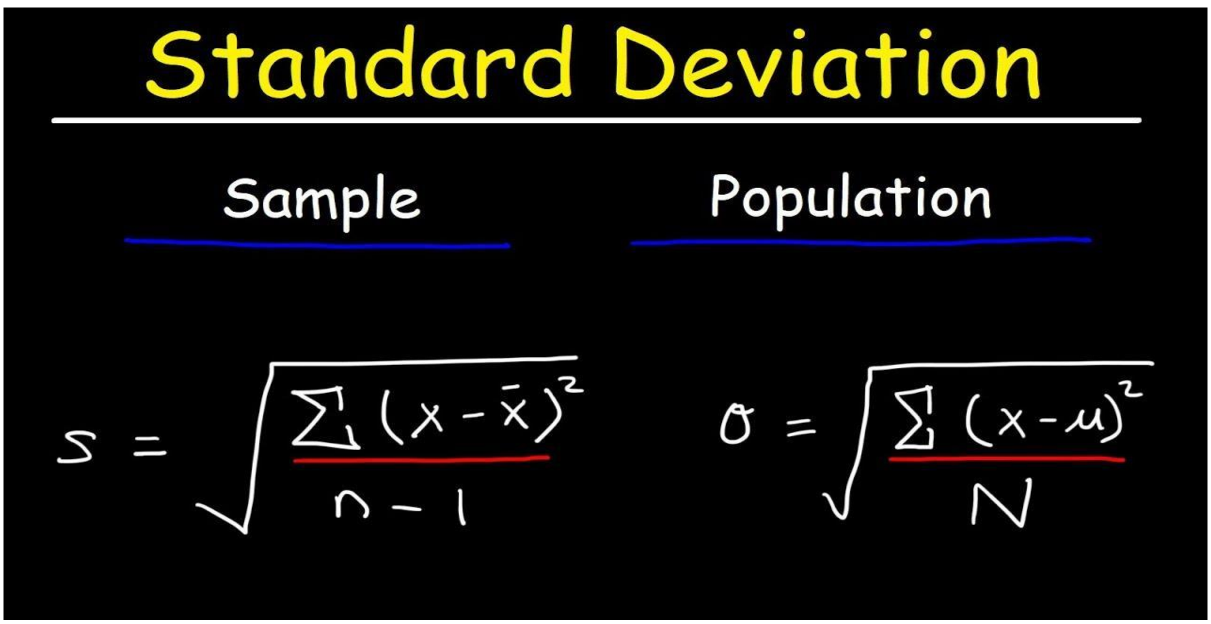
x_i =value of i th element

μ=population mean

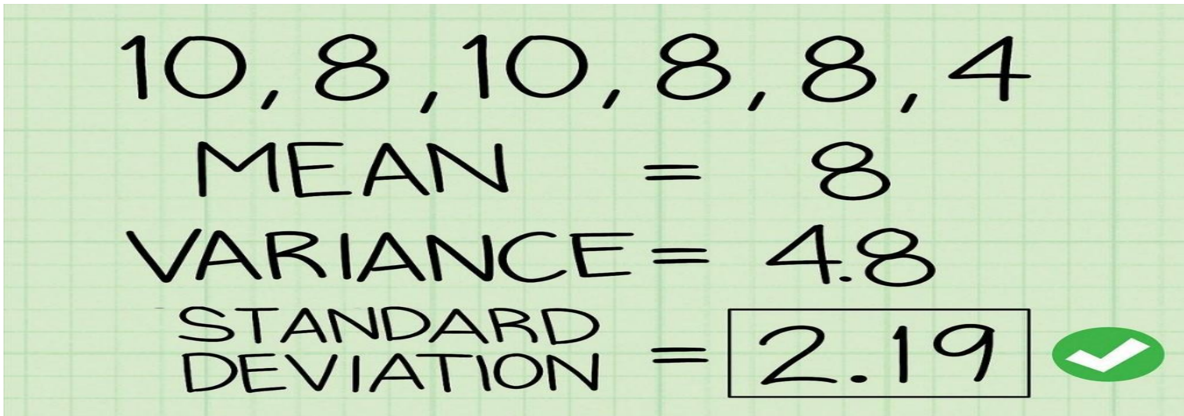
N=population size

<u>Standard Deviation:</u>



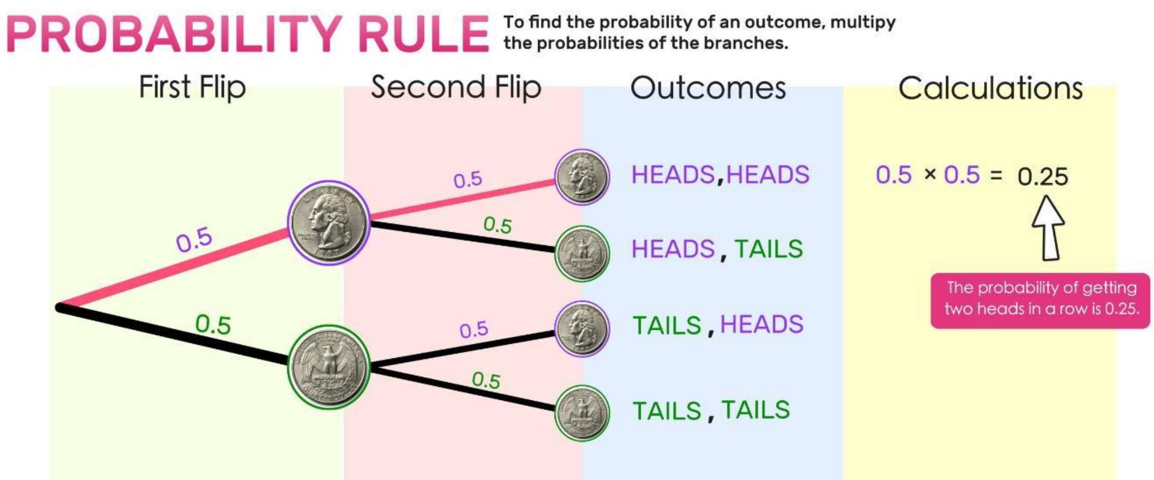


Example:



Probability:

Probability is the likelihood or chance of an event occurring. For example, the probability of flipping a coin and it being heads is ½, because there is 1 way of getting a head and the total number of possible outcomes is 2 (a head or tail). We write P(heads) = ½ .



Type of Events:

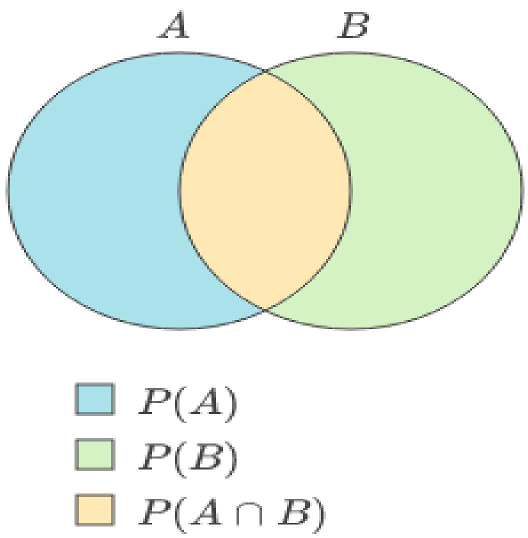
Independent Events	**Dependent Events**
The outcome of one event does not affect the outcome of the other.	The outcome of one event affects the outcome of the other.
If A and B are independent events then the probability of both occurring is	If A and B are dependent events then the probability of both occurring is
$P(A \text{ and } B) = P(A) \times P(B)$	$P(A \text{ and } B) = P(A) \times P(B \mid A)$
	Probability of B given A

Example:

Independent Events	Dependent Events
The first marble is replaced. X: blue first Y: yellow second	The first marble is not replaced. X: blue first Y: yellow second
$P(X) \times P(X/Y) = P(X \text{ and } Y)$	$P(X) \times P(X/Y) = P(X \text{ and } Y)$
$\frac{6}{15} \times \frac{9}{15} = \frac{54}{225} \approx 24\%$	$\frac{6}{15} \times \frac{9}{14} = \frac{54}{210} \approx 26\%$

<u>Conditional Probability:</u>

Conditional probability is defined as the likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome. Conditional probability is calculated by multiplying the probability of the preceding event by the updated probability of the succeeding, or conditional, event.



Conditional Probability Formula

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

Probability that A occurs given that B has already occured

Cheat sheet:

Probability Rules Cheat Sheet

complement rule

$P(A) = 1 - P(A')$

multiplication rules (joint probability)

dependent $P(A \cap B) = P(A) * P(B|A)$

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

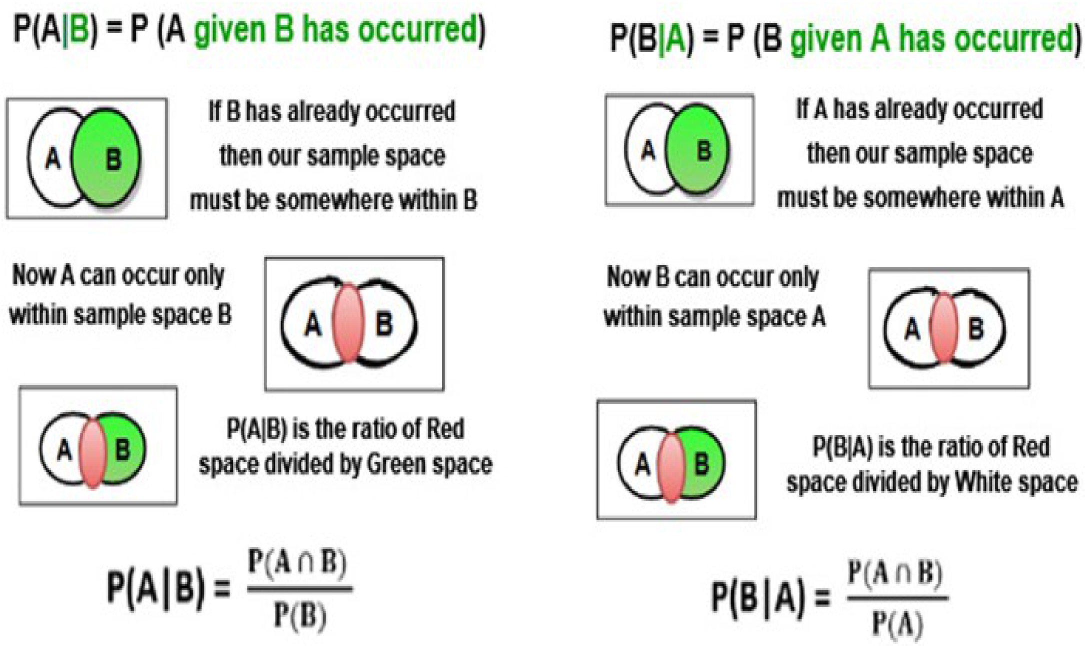
mutually exclusive $P(A \cap B) = 0$

addition rules (union of events)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

mutually exclusive $P(A \cup B) = P(A) + P(B)$

Note:


P(A|B) = P (A given B has occurred)
If B has already occurred
then our sample space
must be somewhere within B
A
B
Now A can occur only
within sample space B
A
B
A
B
P(A|B) is the ratio of Red
space divided by Green space
$P(A|B) = \frac{P(A \cap B)}{P(B)}$
P(B|A) = P (B given A has occurred)
If A has already occurred
then our sample space
must be somewhere within A
A
B
Now B can occur only
within sample space A
A
B
A
B
P(B|A) is the ratio of Red
space divided by White space
$P(B|A) = \frac{P(A \cap B)}{P(A)}$

Example:

Example #1: A bag contains 6 blue and 9 yellow marbles. For each condition below, find the probability of randomly selecting a blue marble on the first draw and a yellow marble on the second draw.

Condition #1: The first marble is replaced for the second draw.

$$P(\text{blue}) = \frac{6}{6+9} = \frac{6}{15}$$

$$P(\text{yellow}) = \frac{9}{6+9} = \frac{9}{15}$$

$$P(\text{blue and yellow}) = \frac{6}{15} \times \frac{9}{15} = \frac{54}{225} \approx 24\%$$

Condition #2: The first marble is not replaced for the second draw.

$$P(\text{blue}) = \frac{6}{9+6} = \frac{6}{15}$$

$$P(\text{yellow}) = \frac{9}{8+6} = \frac{9}{14}$$ -total number of marbles is 14 on second draw

$$P(\text{blue and yellow}) = \frac{6}{15} \times \frac{9}{14} = \frac{54}{210} \approx 26\%$$

Marginal Probability:

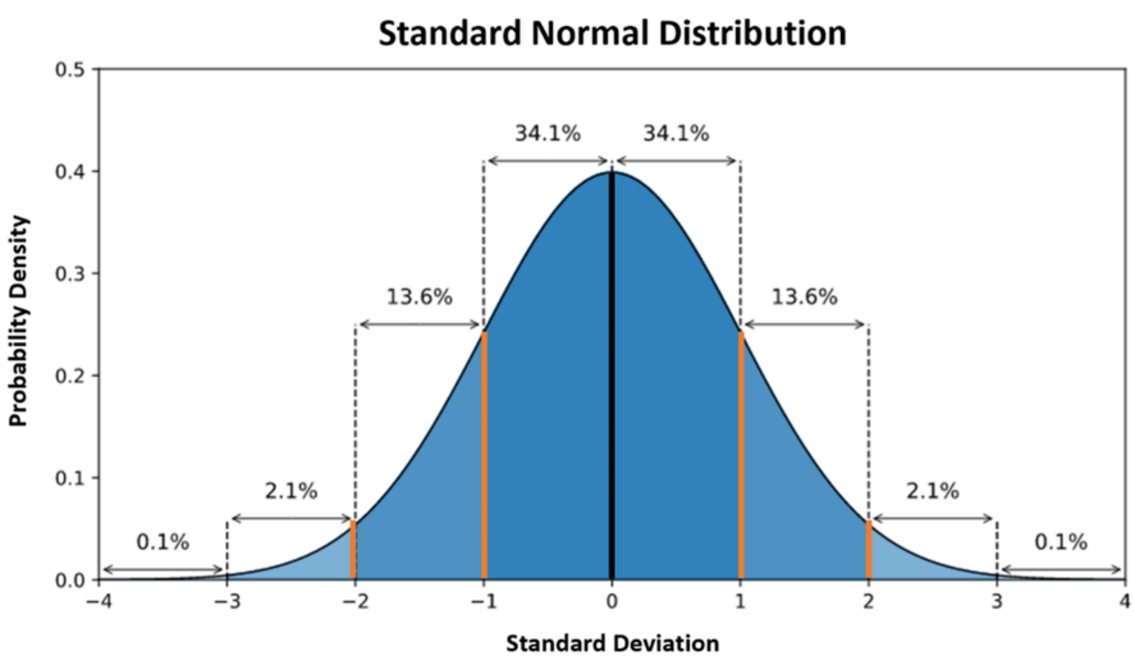
Marginal probability: the probability of an event occurring (p(A)), it may be thought of as an unconditional probability. It is not conditioned on another event. Example: the probability that a card drawn is red (p(red) = 0.5). Another example: the probability that a card drawn is a 4 (p(four)=1/13).

Marginal probabilities

	Pass	Fail	Total	
Males	46	56	102	P(male) = 0.51
Females	68	30	98	P(female) = 0.49
Total	114	86	200	
	P(passed) = 0.57	P(failed = 0.43)		

Normal Distribution:

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".



Mean, Median and Mode lie on the central black line.

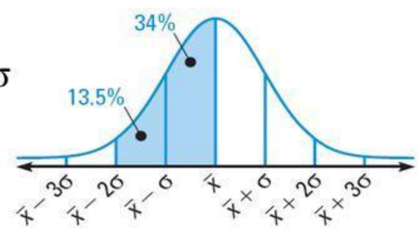
Example:

EXAMPLE 1 **Find a normal probability**

A normal distribution has mean $\overline{x}$ and standard deviation σ. For a randomly selected x-value from the distribution, find $P(\overline{x} - 2\sigma \le x \le \overline{x})$.

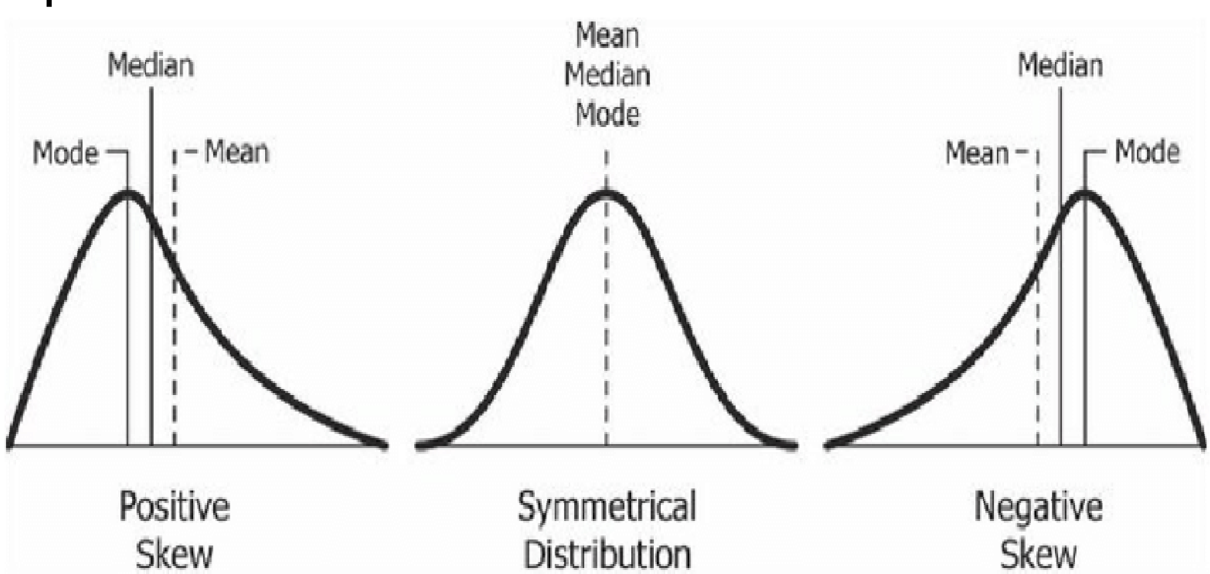
SOLUTION

The probability that a randomly selected x-value lies between $\overline{x} - 2\sigma$ and $\overline{x}$ is the shaded area under the normal curve shown.



$$P(\overline{x} - 2\sigma \le x \le \overline{x}) = 0.135 + 0.34 = 0.475$$

Important Note:



Binomial Distribution:

Binomial distribution is calculated by multiplying the probability of success raised to the power of the number of successes and the probability of failure raised to the power of the difference between the number of successes and the number of trials.

Binomial Distribution Formula

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

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

Poisson Distribution:

A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space. The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events.

Poisson Distribution Formula

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

where

x = 0, 1, 2, 3, …

λ = mean number of occurrences in the interval

e = Euler's constant $\approx$ 2.71828

Central Limit Theoram:

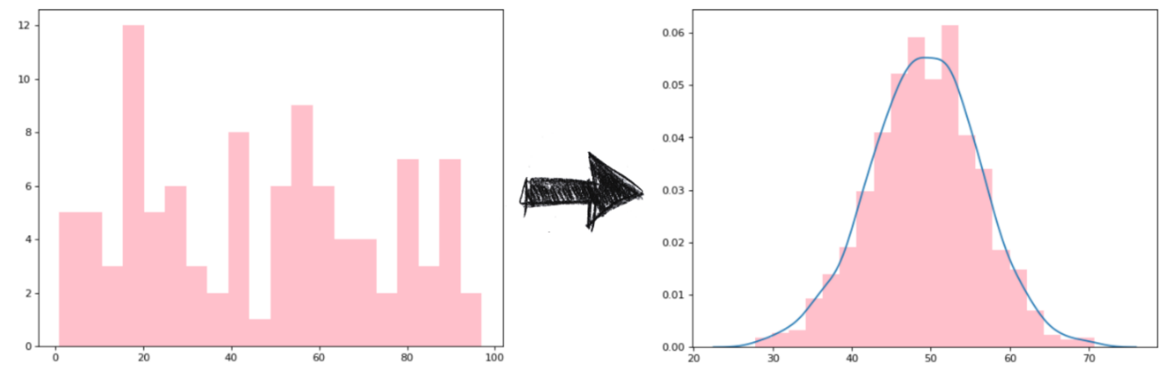
The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

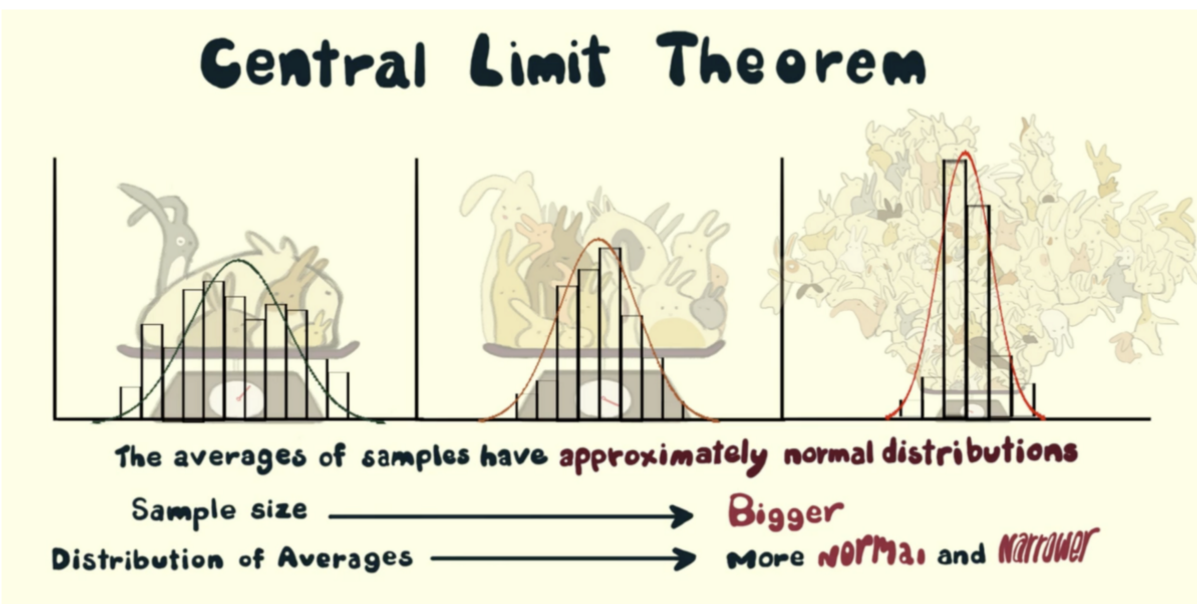
The three rules of the central limit theorem are as follows:

- The data should be sampled randomly.
- The samples should be independent of each other.
- The sample size should be sufficiently large but not exceed 10% of the population.

CENTRAL LIMIT THEOREM

WHAT IS IT AND WHY IS IT USEFUL?





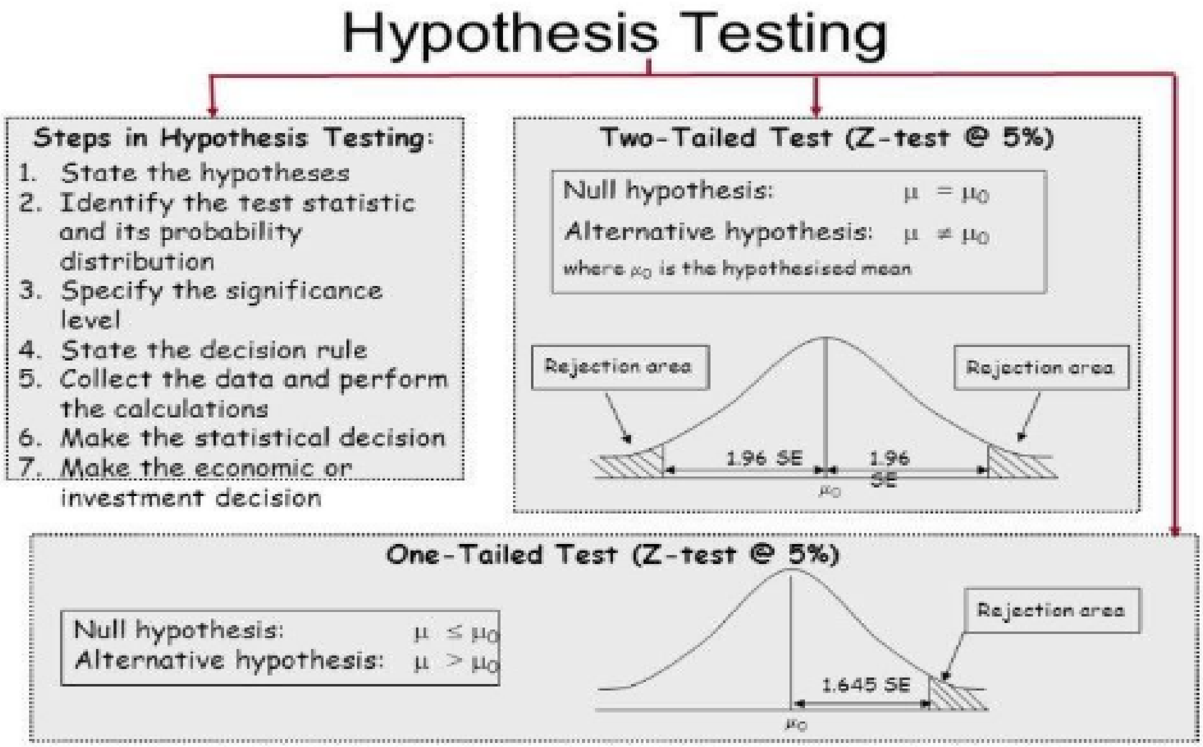
Hypothesis:

Hypothesis testing is a statistical interpretation that examines a sample to determine whether the results stand true for the population. The test allows two explanations for the data—the null hypothesis or the alternative hypothesis. If the sample mean matches the population mean, the null hypothesis is proven true.

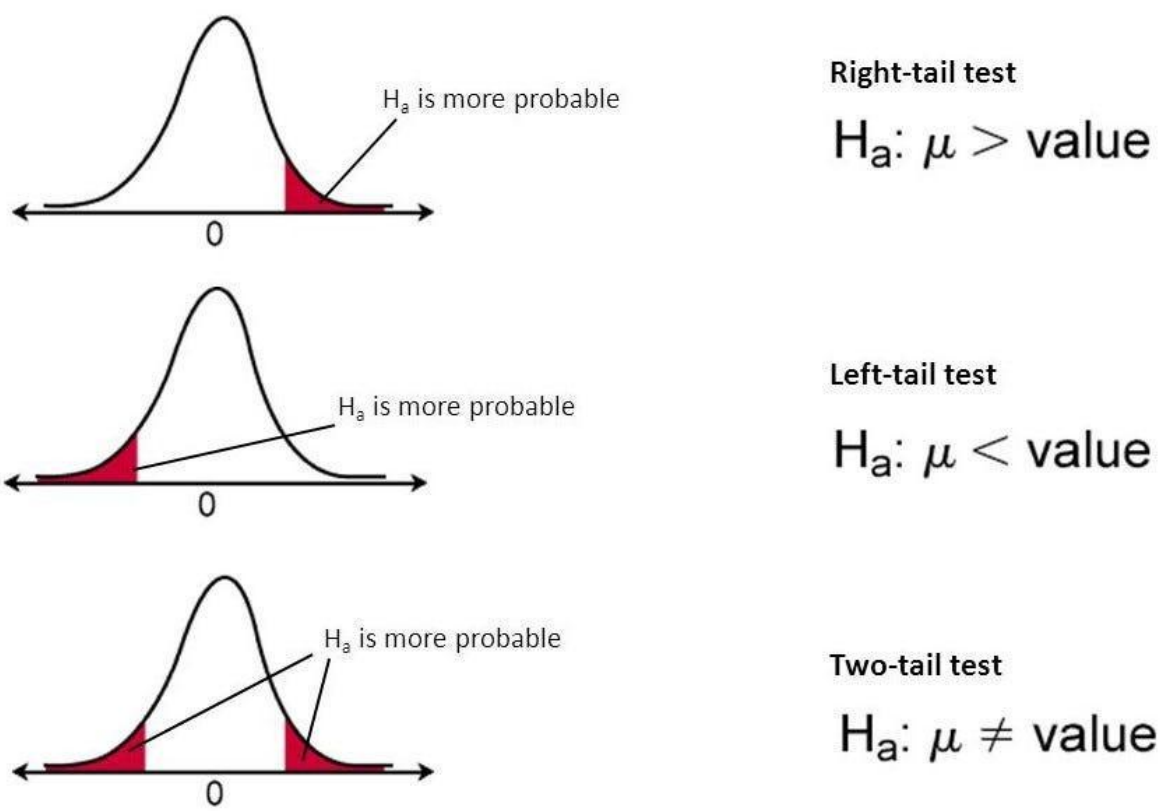
Hypothesis Testing

[hī-ˈpä-thə-səs ˈte-stiŋ]

An act in statistics whereby an analyst tests an assumption regarding a population parameter.


Hypothesis Testing
Steps in Hypothesis Testing:
1. State the hypotheses
2. Identify the test statistic and its probability distribution
3. Specify the significance level
4. State the decision rule
5. Collect the data and perform the calculations
6. Make the statistical decision
7. Make the economic or investment decision
Two-Tailed Test (Z-test @ 5%)
Null hypothesis: $\mu = \mu_0$
Alternative hypothesis: $\mu \neq \mu_0$
where μ_0 is the hypothesised mean
Rejection area
Rejection area
1.96 SE
1.96 SE
μ_0
One-Tailed Test (Z-test @ 5%)
Null hypothesis: $\mu \leq \mu_0$
Alternative hypothesis: $\mu > \mu_0$
Rejection area
1.645 SE
μ_0

Types:


H_a is more probable
0
Right-tail test
H_a: $\mu > $ value
H_a is more probable
0
Left-tail test
H_a: $\mu <$ value
H_a is more probable
0
Two-tail test
H_a: $\mu \neq$ value

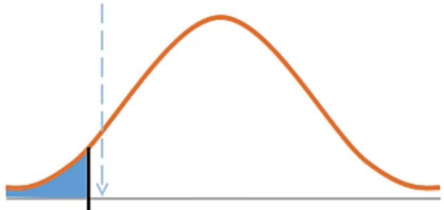
Application of hypothesis testing:

Example: A random sample of 27 observations from a large population has a mean of 22 and a standard deviation of 4.8. Can we conclude at α = 0.01 that the population mean is significantly below 24?

$n = 27 \quad \bar{x} = 22 \quad s = 4.8 \quad \alpha = 0.01$

$$H_0: \mu \geq 24$$
$$H_1: \mu < 24$$

$\alpha = 0.01 \qquad df = 26$



$t = -2.479$

Reject H_0 if $t < -2.479$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{22 - 24}{4.6/\sqrt{27}} = -2.165$$

Since t = -2.165 is not less than -2.479,

Fail to Reject H_0

There is not enough evidence that the population mean is less than 24.

Degree of freedom:

Degrees of freedom and hypothesis testing. The degrees of freedom of a test statistic determines the critical value of the hypothesis test. The critical value is calculated from the null distribution and is a cut-off value to decide whether to reject the null hypothesis.

Application:

Solution: Find the values of n, $\bar{x}$ and σ:

$n = \underline{45}$, $\bar{x} = \underline{14.68}$ and $\sigma = \underline{4.2}$

Step 1: State the null hypothesis (H_0) and the alternative hypothesis (H_a). The null and alternative hypotheses are:

$H_0: \ \mu \ = \ \underline{18\ mg}$
$H_a: \ \mu \ \leq \ \underline{18\ mg}$ ⇒ A left/~~two/right~~ tailed test

Step 2: Decide on the siginificance level, (α) :

Degree of significance, α% = <u>1%</u>

Level of significance, α = <u>0.01</u>

Step 3: Compute the value of the test statistics (Z) :

$$\mu_{\bar{x}} = \mu = \underline{18}$$

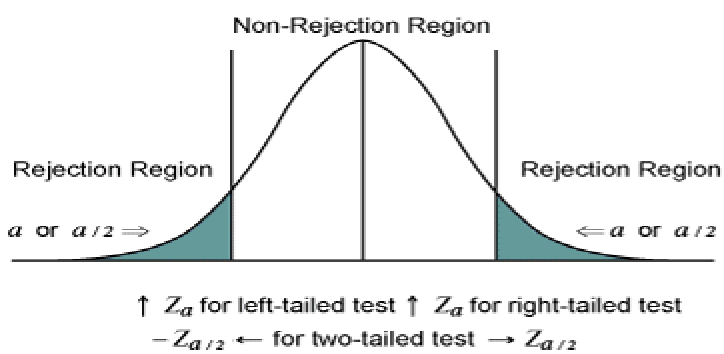
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.2}{\sqrt{45}} \Rightarrow$$

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma/\sqrt{n}} = \frac{(14.68 - 18)}{(4.2/\sqrt{45})} = \underline{-5.30}$$

Step 4: Type of the test: **Check one only**

1. Left-tailed test: [x] Critical value* ⇒ $Z_a = \underline{-2.33}$
2. Two-tailed test: [ ] Critical values* ⇒ $\pm Z_{a/2} = \pm$ ____
3. Right-tailed test: [ ] Critical value* ⇒ $Z_a =$ ____

*Use normal table to find the critical value(s), Z or $\pm Z_{a/2}$:



Step 5: Compare the values of test statistics, Z, and critical value(s), Z_a or $\pm Z_{a/2}$:

Check one only — Check one only

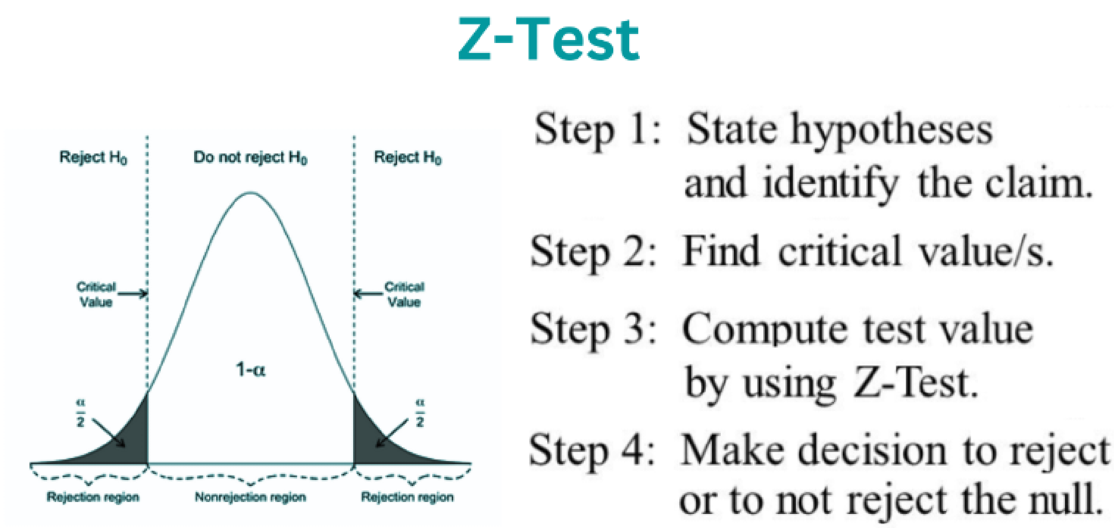
1. $Z \geq Z_a \Rightarrow$ [ ]
2. $-Z_{a/2} \leq Z \leq Z_{a/2}$ [ ] Do not reject H_0 [ ] Otherwise reject H_0 [x]
3. $Z \leq Z_a \Rightarrow$ [x]

Step 6: Interpret the results of the hypothesis test:

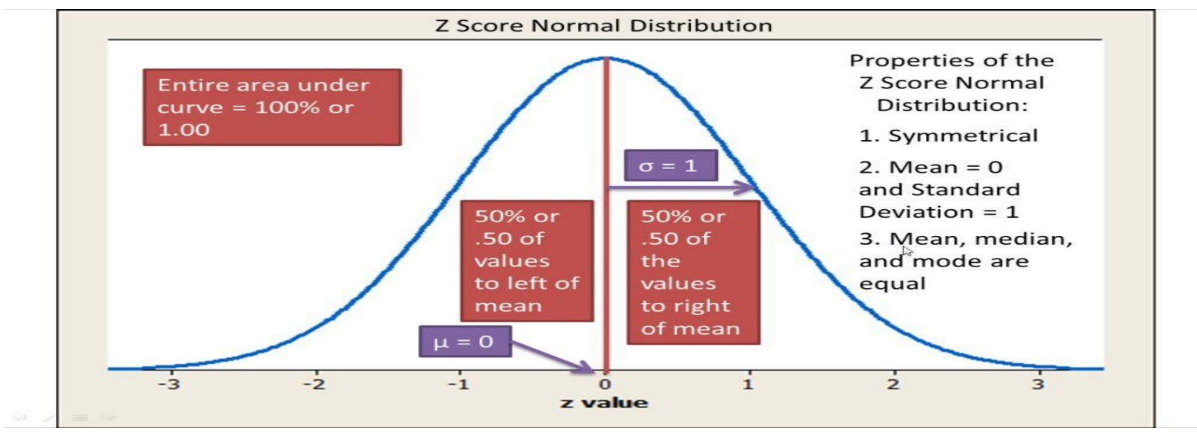
At the <u>1 percent</u> significance level, the data provides/~~does not provide~~ sufficient evidence to conclude that <u>adult females under the age of 51 are, on average, getting less than the RDA of 18 mg of iron.</u>

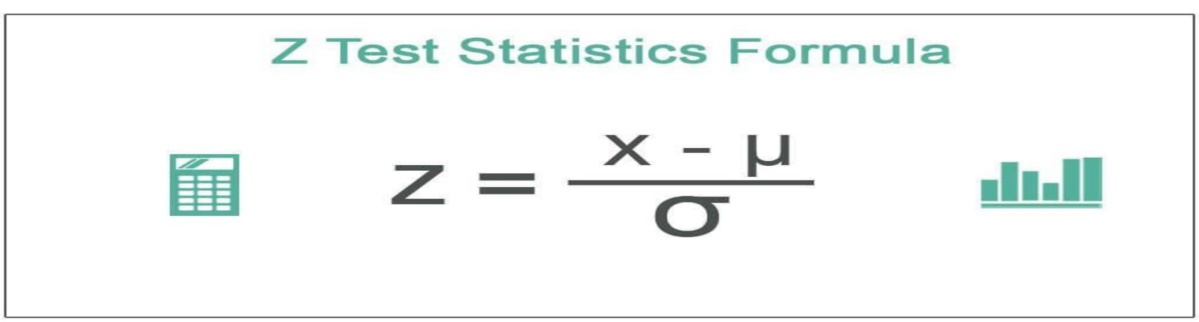
<u>Z-Test:</u>

A z-test is a statistical test to determine whether two population means are different when the variances are known and the sample size is large. A z-test is a hypothesis test in which the z-statistic follows a normal distribution. A z-statistic, or z-score, is a number representing the result from the z-test.



Application:





T-Test:

A t test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

$$\boxed{t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

where,

'x' bar is the mean of the sample,

μ is the assumed mean,

σ is the standard deviation

and n is the number of observations

	Z-test	T-test
population variance	*is known*	*is unknown*
distribution	*normal distribution*	*student's t-distribution*
degrees of freedom	*don't need*	*are needed*
calculated with	*standard error*	*estimated standard error*
proportion testing	*when $np > 10$ and $n(1 - p) > 10$*	*is not used for this*

Chi-Square Test:

A chi-square test is a statistical test used to compare observed results with expected results. The purpose of this test is to determine if a difference between observed data and expected data is due to chance, or if it is due to a relationship between the variables you are studying.

The Formula for Chi Square Is

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

where:

c = degrees of freedom

O = observed value(s)

E = expected value(s)

Application:

chi-square tests

Conditions for inference

1. Random
2. Large counts ✓
3. Independent

$$\boxed{\chi}^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= \frac{(60-65)^2}{65} + \frac{(125-117)^2}{117} + \frac{(80-78)^2}{78} + \frac{(65-60)^2}{60} + \frac{(100-108)^2}{108} + \frac{(70-72)^2}{72}$$

$$= \frac{25}{65} + \frac{64}{117} + \frac{4}{78} + \frac{25}{60} + \frac{64}{108} + \frac{4}{72}$$

$$\chi^2 \approx 2.05$$

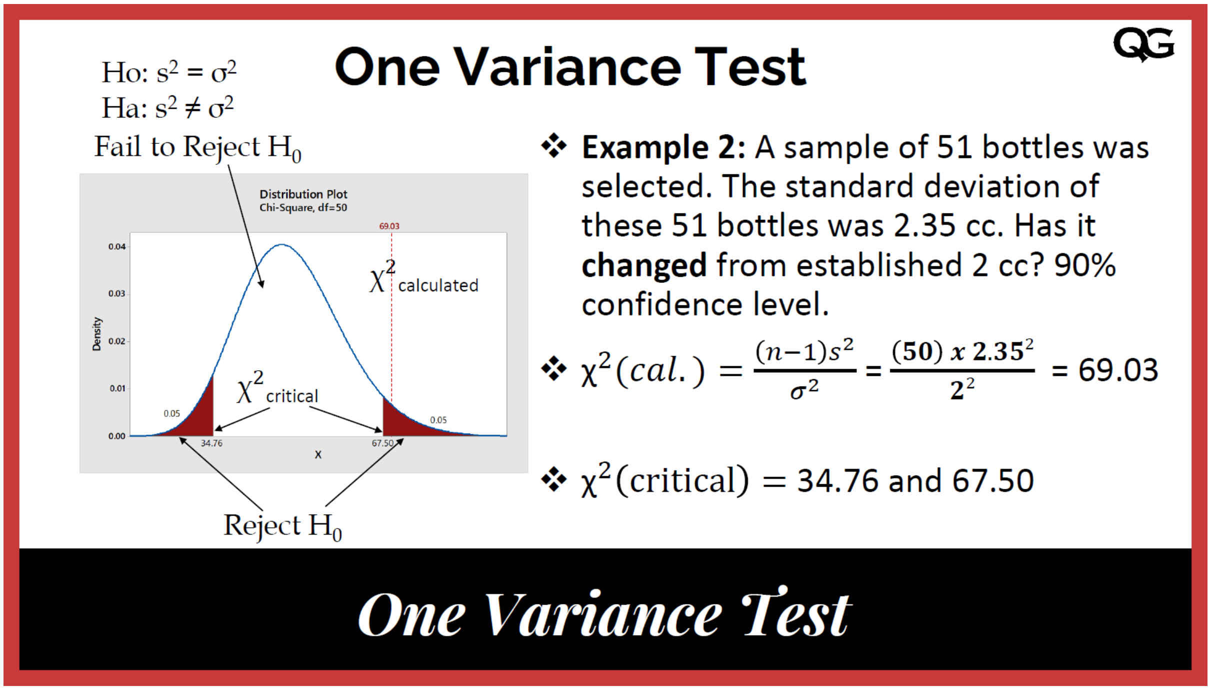
Homogeneity

	MLB	NFL	NBA	Total
30s	60^{65}	125^{117}	80^{78}	$260^{0.52}$
40s	65^{60}	100^{108}	70^{72}	$240^{0.48}$
Total	0.25 125	0.45 225	0.30 150	500

H_0: Age doesnt affect sport

H_a: Age does affect sport

Chi-Square practical:



Working of chi-square test:

Working:

Plant Phenotype	Observed Number (O)	*Expected Ratio*	Expected Number (E)	$O - E$	$\frac{(O-E)^2}{E}$
Purple / long	296	*9*	427 × (9/16) = 240	56	13.067
Purple / round	19	*3*	427 × (3/16) = 80	– 61	46.513
Red / long	27	*3*	427 × (3/16) = 80	– 53	35.113
Red / Round	85	*1*	427 × (1/16) = 27	58	124.593
Total	427			χ^2	**219.29**

Analysis:

df	p values for Chi-Square (χ^2) distribution							
	0.90	**0.75**	**0.50**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**
3	0.584	1.212	2.366	4.110	6.251	7.815	9.348	11.345
	genes unlinked					*potentially linked*		

Conclusion:

H_1: There **is** a significant difference between observed and expected frequencies (genes may be linked)