Explanation of Stats terms with real world examples:

1. Scalar:

Scalar is a value that has only one component which is magnitude. For example: Speed is a scalar value because it only measures how fast you are going.

Speed = 9 m/s,

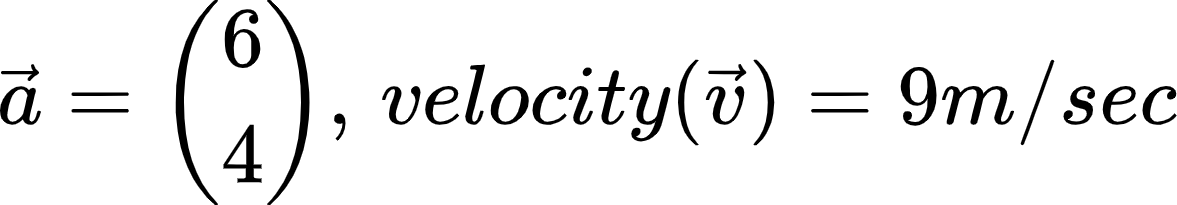
Some variables like a=8,b=10(having constant value) these all are example of scalar

1. Vector:

Vector is a value having 2 components i.e magnitude and direction. For example:

Velocity is a vector value because it measures how fast you are going(magnitude) and at which way you are going(direction)

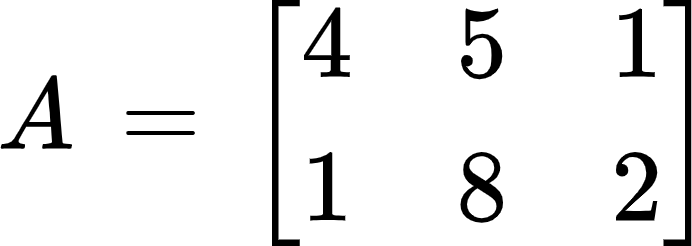
Vector\_a = [3,4]

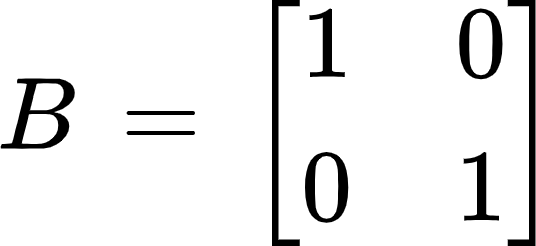


1. Matrices:

An m × n (read 'm by n') matrix is an arrangement of numbers (or algebraic expressions ) in m rows and n columns . Each number in a given matrix is called an element or entry .A zero matrix has all its elements equal to zero.

Examples:

 has a shape of (2,3)

 has space of (2,2) and is also called identity matrix

1. Tensor:

In simpler terms a tensor is a dimensional data structure**.** Vectors are one-dimensional data structures and matrices are two-dimensional data structures.

Examples:

Tensor\_a =[[[[1,2],[2,3]]]]

**2. Matrix Multiplication:**

Let's suppose we have two matrices A,B.

A = [[1,2,3],[2,3,4]] #shape(2,3)

B = [[1,2],[2,3],[3,4]] #shape(3,2)

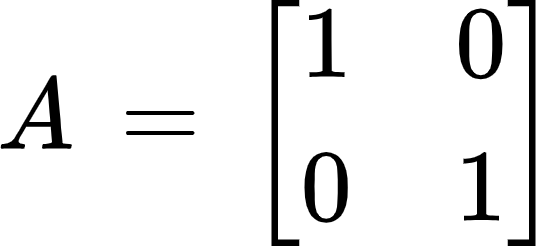
When we do matrix multiplication we need to at first see if the shape[1] of A matches with shape[0] of matrix B.Here in this case it matches so the resultant matrix is of shape(2,2)

Resultant\_matrix = [[14,20],[20,29]]

**3. Identity Matrix and Inverse Matrix:**

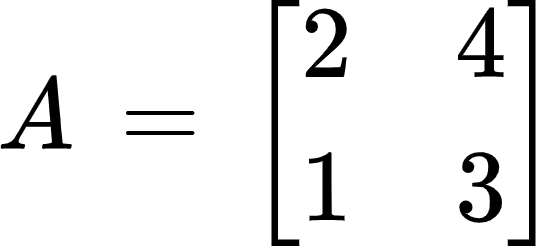
Identity Matrix of size n is a square matrix n\*n which has all its diagonal element one and non diagonal elements as zero.

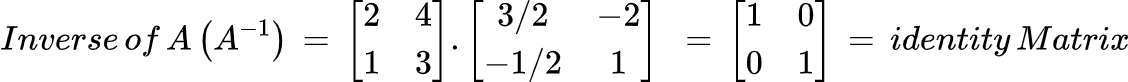
A is an identity matrix of size n=2.



Inverse matrix of matrix A is defined as that matrix when multiplied with A gives us an identity matrix.

For example:





**4. Linear Dependence and Span:**

The span of a set of vectors is the set of all linear combinations of the vectors. A set of vectors is linearly independent if the only solution to

*c1v1 + c2v2+ ....... + ckvk = 0 is ci = 0 for all i*.

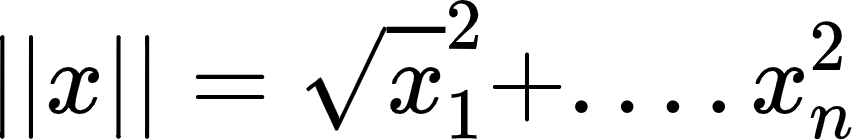
Given a set of vectors, you can determine if they are linearly independent by writing the vectors as the columns of the matrix A, and solving Ax = 0 where x is a constant. If there are any non-zero solutions, then the vectors are linearly dependent. If the only solution is x = 0, then they are linearly independent.

**5. Norm:**

Given a vector space X over a subfield F of the complex numbers C a norm on X is a real-valued function p: X→R with the following properties, where |s| denotes the usual absolute value of a scalar s.

Example :

||x|| = |x| is an example of absolute norm

 is an example of eulidean norm

**Part -2**

1. **Random Variable:**

A random variable is a variable whose value is unknown to the function or task i.e, the value is depends upon the outcome of experiment

For example, while throwing a dice, the variable value depends upon the outcome.

Mostly random variables are used for regression analysis to determine statistical relationship between each other.

There are 2 types of random variable:

1 — Continuous random variable

2 — Discrete random variable

**Continuous random variable:-**

A variable which has the values between the range/interval and takes an infinite number of possible ways is called Continuous random variable .

Examples: A average height of 100 peoples, measurement of rainfall

**Discrete Random Variable:-**

A variable which takes a countable number of distinct values is called Discrete Random Variable.

Examples: number of students present in class

1. **Probability Mass Function and Probability Density Function:**

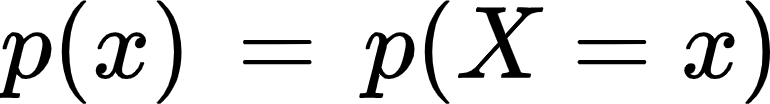
Both PMF and PDF are a distribution function, the difference is between random variables: PDF is relevant for continuous random variables while PMF is relevant for discrete random variables.

**Probability Mass Function:**

The Probability Mass function depends on the values of any real number. It does not go to the value of X which equals zero and in case of x, the value of PMF is positive.

The PMF plays an important role in defining a discrete probability distribution and produces distinct outcomes.

The formula of PMF is



i.e the probability of (x)= the probability (X=one specific x)

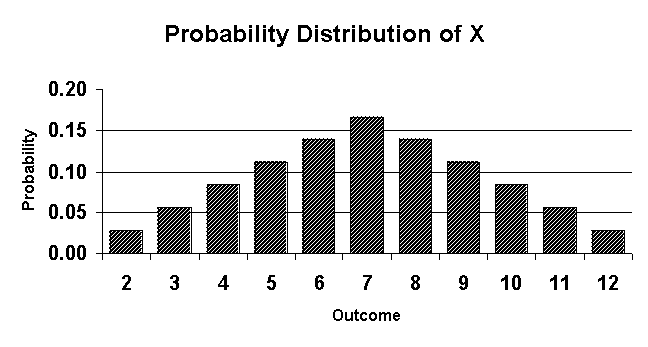
Examples are Number of students in class, Numbers on a dice etc.

Numbers on a dice is a discrete random variable so the probability distribution is PMF. Suppose we have 2 dices and Let X represent the sum of two dice.

Then the probability distribution of *X* is as follows:

| *X* | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P(*X*) |  |  |  |  |  |  |  |  |  |  |  |

When we graph the above table the distribution looks like



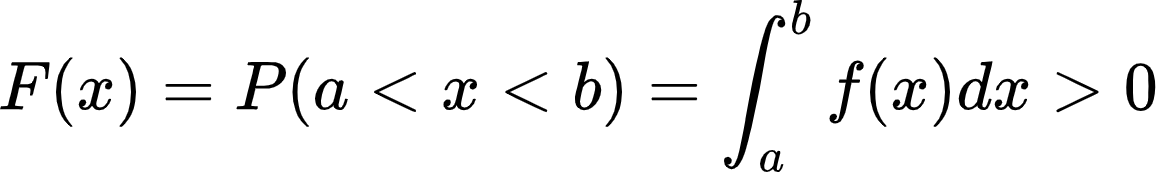
**Probability Density Function:**

The Probability Density Function (PDF) depicts probability functions in terms of continuous random variable values presenting in between a clear range of values.

The PDF is essentially a variable density over a given range. It is positive/non-negative at any given point in the graph and the whole of PDF is always equal to one.

To calculate the probability of X resting in an interval (a, b) along with for P(a< X< b) which can take place using a PDF.

The Probability distribution function formula is defined as,



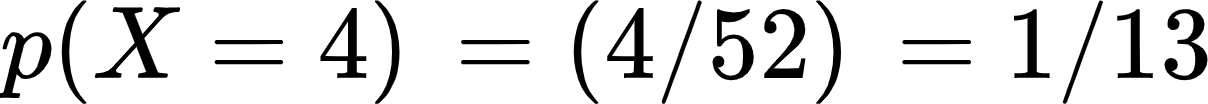
Examples:Calculating the Temperature, rainfall and overall weather distribution, Time computer takes to process input and give output etc.

1. **Marginal probability:**

If the probability of an event to occur without any condition is p(A) then this probability is called marginal probability. It is not conditioned on any other event.

For example:

Let us consider a deck of 52 playing cards, the probability of getting a 4 number card is



Where X is the event of getting 4.

This is the example of marginal probability as it does not have any condition

1. **Joint Probability and Conditional Probability:**

**Joint Probability:**

Joint probability is the probability where two events are occurring simultaneously p(A and B). The probability of event A and event B occurring. It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written p(A ∩ B).

Example:

Let us consider a deck of 52 playing cards, the probability that a card is a six numbered and is black is

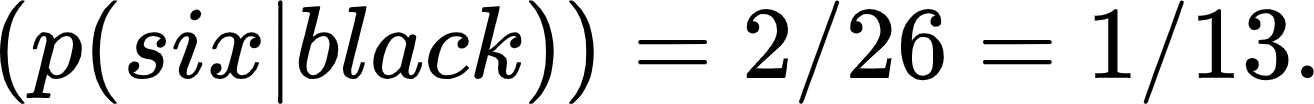
p(six and black) = 2/52=1/26.

(There are two black six numbered cards in a deck of 52, the 6 of jack and the 6 of spade).

**Conditional Probability:**

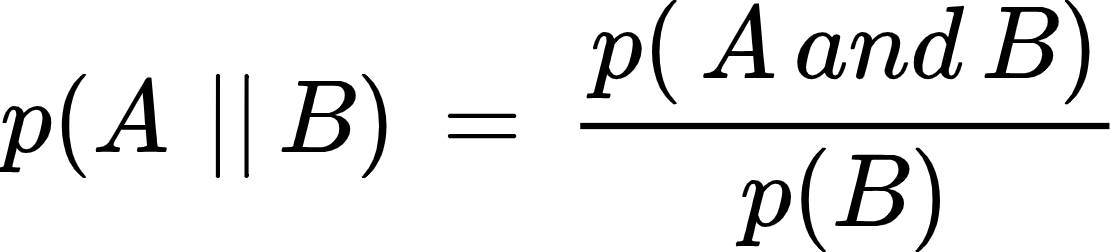
p(A|B) is the probability of event A occurring, given that event B occurs. Example:

Given we have a deck of 52 cards and we drew a black card, what’s the probability that it’s a six



So out of the 26 black cards (given a blackcard), there are two sixes so 2/26=1/13.

There is a relation between joint probability and conditional probability. The conditional probability of A given B is equal to the joint probability of A and B divided by the marginal of B.



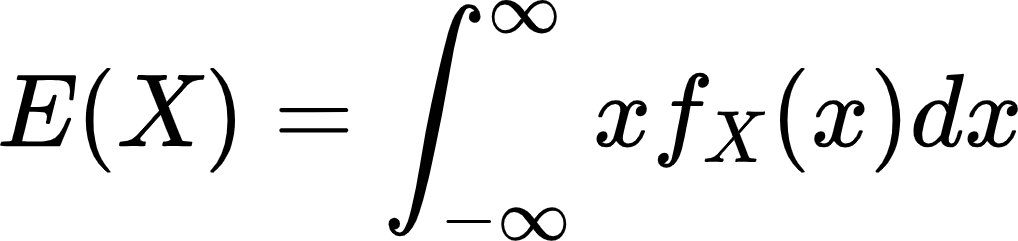
1. **Expectation, Variance and Covariance:**

**Expectation:**

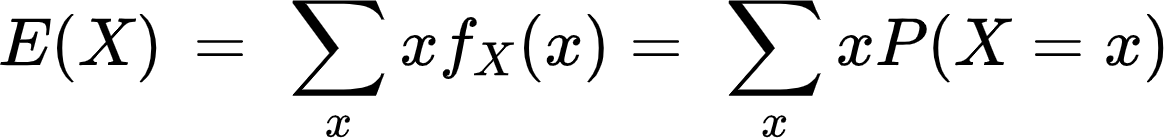
In probability, the average value of some random variable X is called the expectation.It uses the notation E with square brackets around the name of the variable.

for example: E[X]

The expectation is defined differently for continuous and discrete random variables. Let X be a continuous random variable with p.d.f. fX(x). The expected value of X is



Let X be a discrete random variable with probability function fX(x).The expected value of X is



Example:

If you take a 20 question multiple-choice test with A,B,C,D as the answers, and you guess all “A”, then you can expect to get 25% right (5 out of 20). The math behind this kind of expected value is:

The probability (P) of getting a question right if you guess: .25

The number of questions on the test (X): 20

E(X) = P x n = .25 x 20 = 5

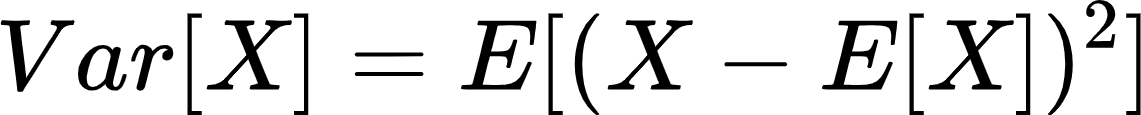
**Variance:**

the variance of some random variable X is a measure of how much values in the distribution vary on average with respect to the mean.Variance is a measure of how spread out the data is.

Example: Let us consider we have 5 students in out classroom and their heights in cm are 160, 165,170,175,180

mean/Expected value = 170

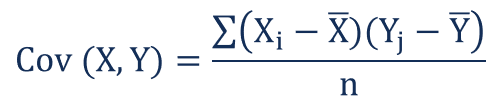
Using Variance formula:



Variance\_value(Var(X)) = (100+25+0+25+100)/5 = 50

**Covariance:**

Covariance measures the total variation of two random variables from their expected values.For example, the covariance between two random variables X and Y can be calculated using the following formula(for population):



Example: Let us consider we have dataset with variables

x: 2.1, 2.5, 3.6, 4.0 (mean = 3.1)

y: 8, 10, 12, 14 (mean = 11)

Substitute the values into the formula and solve:

Cov(X,Y) = ΣE((X-μ)(Y-ν)) / n-1

= (2.1-3.1)(8-11)+(2.5-3.1)(10-11)+(3.6-3.1)(12-11)+(4.0-3.1)(14-11) /(4-1)

= (-1)(-3) + (-0.6)(-1)+(.5)(1)+(0.9)(3) / 3

= 3 + 0.6 + .5 + 2.7 / 3

= 6.8/3

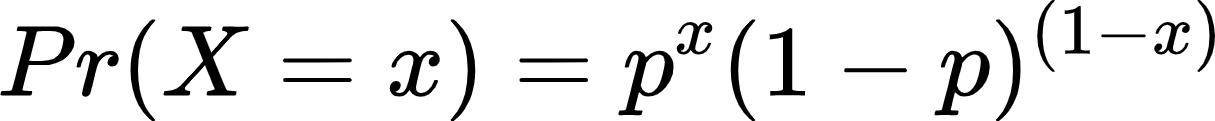
= 2.267

1. **Distributions:**

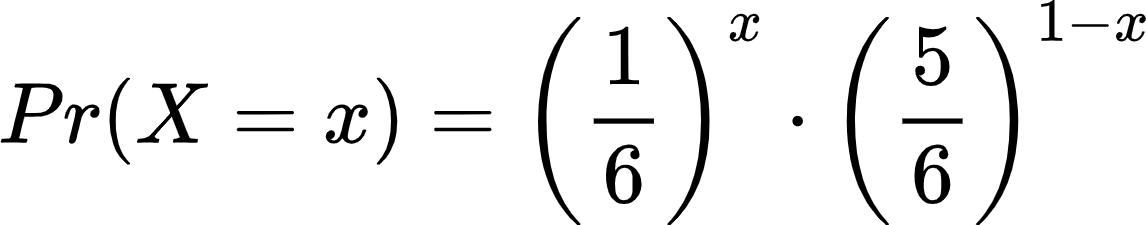
**Bernoulli Distribution:**

Bernoulli trials describe the probability of success or failure in experiments that have only two possible outcomes.A Bernoulli distribution is simply a probability distribution that describes the probability of obtaining a success for a Bernoulli trial.

Example: Suppose we are interested in two dice rolling a 7, then rolling two dice one time would be a Bernoulli trial since there are two outcomes of interest: we roll a 7 (success) or we don't roll a 7 (failure).If X is a Bernoulli variable with X=1 denoting a success and X=0 denoting a failure, then the Bernoulli distribution is



X=1 if we roll a 7 and X=0 if we don't roll a 7. Since the probability of rolling a 7 is Pr(roll a 7)=6/36=1/6 , X has the Bernoulli distribution



**Multinoulli distribution:**

This distribution is also called categorical distribution, since it can be used to model events with K possible outcomes. Bernoulli distribution can be seen as a specific case of Multinoulli, where the number of possible outcomes K is 2.

In machine learning, the multinoulli can be used to model the expected class of one sample into a set of K classes. For instance, one may want to predict to which species in the set a flower belongs based on its attribute. Then species K follow a multinoulli distribution.

Consider the p(x=k) the probability that the sample x belongs to class k. Here x could be the attributes of a flower in the example above, or one side of a die in the roll of it. If the set of classes is K belongs 1,2,3,.....K , then the probability of each outcome can be written as:

p(x=1) = p1

p(x=2) = p2

……

So sum of all probabilities equals 1.

In flower classification problem,

p(x=1) = 0.1

p(x=2) = 0.3

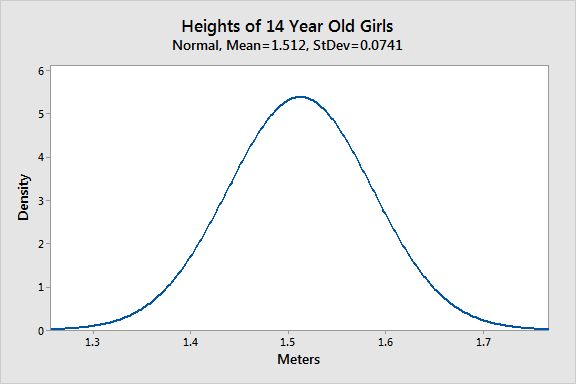
p(x=3) = 0.6

Sum of all probabilities equals =1 and one would say that the sample is most probable from class 3

**Normal Distribution:**

The normal distribution is a probability function that describes how the values of a variable are distributed.It is also known as the Gaussian distribution and the bell curve.

Height data are normally distributed. The distribution in this example fits real data that I collected from 14-year-old girls during a study.

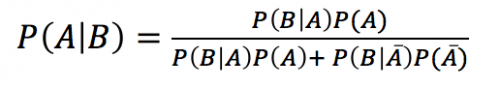


Most girls are close to the average (1.512 meters). Small differences between an individual’s height and the mean occur more frequently than substantial deviations from the mean. The standard deviation is 0.0741m, which indicates the typical distance that individual girls tend to fall from mean height.

The distribution is symmetric. The number of girls shorter than average equals the number of girls taller than average. In both tails of the distribution, extremely short girls occur as infrequently as extremely tall girls.

1. **Bayes Rule:**

For two events, A and B, Bayes’ theorem lets us to go from p(B|A) to p(A|B) if we know the marginal probabilities of the outcomes of A and the probability of B, given the outcomes of A. Here is the equation for Bayes’ theorem for two events with two possible outcome (A and not A).



Let’s assume we know that 1% of women over the age of 40 have breast cancer.

[p(cancer)=0.01]

Let’s assume that 90% of women who have breast cancer will test

positive for breast cancer in a mammogram.

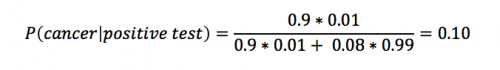
[p(positive test|cancer)=0.9]

Eight percent of women that do NOT have cancer will also test positive.

[p(positive test|no cancer)=0.08]

What is the probability that a woman has cancer if she tests positive [p(cancer|positive test)]?

We will call p(cancer) = P(A), and the P(positive test) = P(B). We want to know P(A|B)–the probability of having cancer if you have a positive test.



Using Bayes’ theorem, we calculate that the likelihood that a woman has breast cancer, given a positive test, equals approximately 0.10. This makes intuitive sense as (1) this result is greater than 1% (the percent of breast cancer in the general public).

.