Mathematics for AI

Inferential Statistics

Inferential statistics are methods for quantifying properties of a population from a small Sample:

You take data from a sample and make a prediction about the whole population.

For example, you can stand in a shop and ask a sample of 100 people if they like chocolate.

From your research, using inferential statistics, you could predict that 91% of all shoppers like chocolate.

Descriptive Statistics

Descriptive Statistics summarizes (describes) observations from a set of data.

Since we register every newborn baby, we can tell that 51 out of 100 are boys.

From these collected numbers, we can predict a 51% chance that a new baby will be a boy.

It is a mystery that the ratio is not 50%, like basic biology would predict. We only know that we have had this tilted sex ratio since the 17th century.

Descriptive Statistics Measurements

Descriptive statistics are broken down into different measures:

Tendency (Measures of the Center)

- The Mean (the average value)value
- The Median (the midpoint value)
- The Mode (the most common value)

Spread (Measures of Variability)

- Min and Max
- Standard Deviation
- Variance
- Skewness
- Kurtosis

Measures of Central Tendency

It represents the whole set of data by a single value. It gives us the location of the central points. There are three main measures of central tendency:

- Mean
- Mode
- Median

Mean

It is the sum of observations divided by the total number of observations. It is also defined as average which is the sum divided by count.

```
lst = [23, 78, 1, 90, 36, 82]
```

```
mean = np.mean(lst)
mean
```

```
51.666666666666664
```

```python
mean = np.mean(df['Price'])
mean
```

```
582209.6295285609
```

```python
mn = df['Price'].mean()
mn
```

```
582209.6295285609
```

```python
df.mean()
```

```
ID                      250.500000
Square_Feet             174.640428
Num_Bedrooms              2.958000
Num_Bathrooms             1.976000
Num_Floors                1.964000
Year_Built             1957.604000
Has_Garden                0.536000
Has_Pool                  0.492000
Garage_Size              30.174000
Location_Score            5.164410
Distance_to_Center       10.469641
Price                582209.629529
dtype: float64
```

Mode

It is the value that has the highest frequency in the given data set. The data set may have no mode if the frequency of all data points is the same. Also, we can have more than one mode if we encounter two or more data points having the same frequency.

```python
mode = df['Year_Built'].mode()
mode
```

```
0    1920
1    1959
Name: Year_Built, dtype: int64
```

```python
df['Num_Bedrooms'].mode()
```

```
0    1
Name: Num_Bedrooms, dtype: int64
```

```python
df2['forehead_width_cm'].mode()
```

```
0    12.0
Name: forehead_width_cm, dtype: float64
```

Median

It is the middle value of the data set. It splits the data into two halves. If the number of elements in the data set is odd then the center element is the median and if it is even then the median would be the average of two central elements.

```python
df['Year_Built'].median()
```

```
1959.0
```

```python
np.median(lst)
```

```
57.0
```

Measure of Variability

Measures of variability are also termed measures of dispersion as it helps to gain insights about the dispersion or the spread of the observations at hand. Some of the measures which are used to calculate the measures of dispersion in the observations of the variables are as follows:

- Range
- Variance
- Standard Deviation
- Distribution
- Skewness
- Kurtosis

Range

The range describes the difference between the largest and smallest data point in our data set. The bigger the range, the more the spread of data and vice versa.

```
print('Range of Year :', max(df['Year_Built']) - min(df['Year_Built']))
```

```
Range of Year : 122
```

```
print('Range of Year :', df['Year_Built'].max() - df['Year_Built'].min())
```

```
Range of Year : 122
```

```
print('Range of Bedrooms :', df['Num_Bedrooms'].max() - df['Num_Bedrooms'].min())
```

```
Range of Bedrooms : 4
```

Variance

It is defined as an average squared deviation from the mean. It is calculated by finding the difference between every data point and the average which is also known as the mean, squaring them, adding all of them, and then dividing by the number of data points present in our data set.

```
h = df['Price'].var()          # pandas dataframe calculates sample variance by default meaning in the denominator (N - 1) is used.
print('Sample variance :', h)  # sample variance is used for large dataset
                               # To calculate population variance --> df['Price'].var(ddof = 0)
```

```
Sample variance : 14950781986.482094
```

```
p = df['Price'].var(ddof = 0)
print('Population variance :', p)
```

```
Population variance : 14920880422.509129
```

```
np.var(df['Price'])  # numpy calculates population variance by default meaning in the denominator (N) is used.
                     # To calculate sample variance --> np.var(df['Price'], ddof = 1)
```

```
14920880422.509129
```

```
np.var(df['Price'], ddof = 1)
```

```
14950781986.482094
```

Standard Deviation

It is defined as the square root of the variance. It is calculated by finding the Mean, then subtracting each number from the Mean which is also known as the average, and squaring the result. Adding all the values and then dividing by the no of terms followed by the square root.

Standard Deviation is a measure of how spread out numbers are.

The symbol is σ (Greek letter sigma).

```
print('Population :', np.std(lst))
print('Sample     :', np.std(lst, ddof = 1))
```

```
Population : 33.45976024354561
Sample     : 36.653330908208964
```

```
print('Sample Standard Deviation     :', df['Price'].std())          # sample Standard Deviation (Default)
print('Population Standard Deviation :', df['Price'].std(ddof = 0))  # Population Standard Deviation
```

```
Sample Standard Deviation     : 122273.39034508732
Population Standard Deviation : 122151.05575683384
```

```
print('Population Standard Deviation :', np.std(df['Price']))            # Population Standard Deviation (Default)
print('Sample Standard Deviation     :', np.std(df['Price'], ddof = 1))  # sample Standard Deviation
```

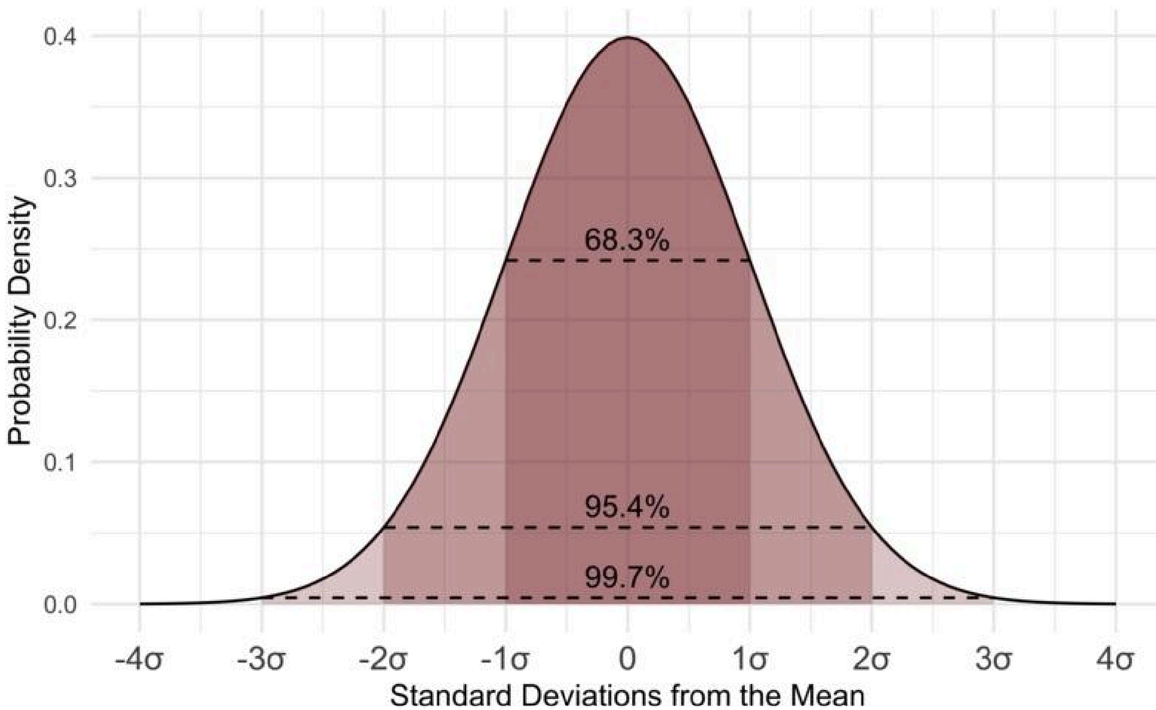
```
Population Standard Deviation : 122151.05575683384
Sample Standard Deviation     : 122273.39034508732
```

Distributions

Normal Distribution

The Normal Distribution Curve is a bell-shaped curve.

Each band of the curve has a width of 1 Standard Deviation.



Normal Distribution Facts

Normal distribution is Symmetric. The peak always divides the distribution in half.

Normal distribution is a Probability distribution.

A lot of observations follow the normal distribution:

- Your IQ
- Your Weight
- Your Height
- Your Salary
- Your Blood Pressure

Binomial Distribution

Type : Discrete distribution.

Characteristics : Models the number of successes in a fixed number of independent trials, each with the same probability of success.

Example : Flipping a coin multiple times, number of defective items in a batch.

Poisson Distribution

Type : Discrete distribution.

Characteristics : Describes the number of events occurring in a fixed interval of time or space, with events happening independently of each other.

Example : Number of emails received per hour, number of accidents at a crossroads.

Exponential Distribution

Type : Continuous distribution.

Characteristics : Models the time between consecutive events in a Poisson process.

Example : Time until a radioactive particle decays, time between arrivals of buses.

Uniform Distribution

Type : Continuous distribution.

Characteristics : All outcomes are equally likely within a given range.

Example : Rolling a fair die, random number generation within a specific interval.

Student’s Distribution

Type : Continuous distribution.

Characteristics : Similar to the normal distribution but with heavier tails. Used when sample sizes are small and population standard deviation is unknown.

Example : Estimating population parameters from a small sample.

The Margin of Error

Statisticians will always try to predict everything with 100% accuracy.

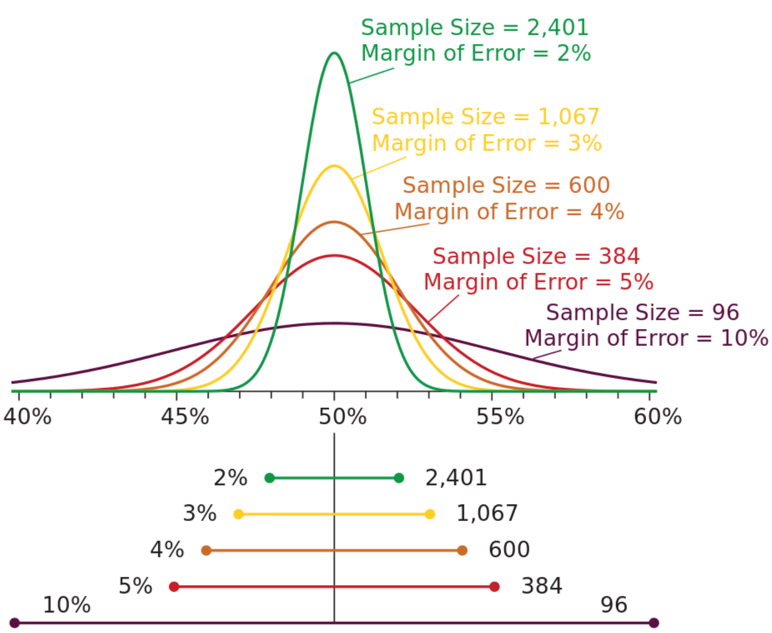
But, there will always be some uncertainty.

The Margin of Error is the number that quantifies this statistical uncertainty.

Different margins define different ranges for where we believe the correct answers can be found.

The acceptable margin is a matter of judgment, and relative to how important the answer is.

The more samples we collect, the lower the margin of error is.



How to Interpret Margin of Error

Suppose 55% of a sampled population say they plan to vote "Yes".

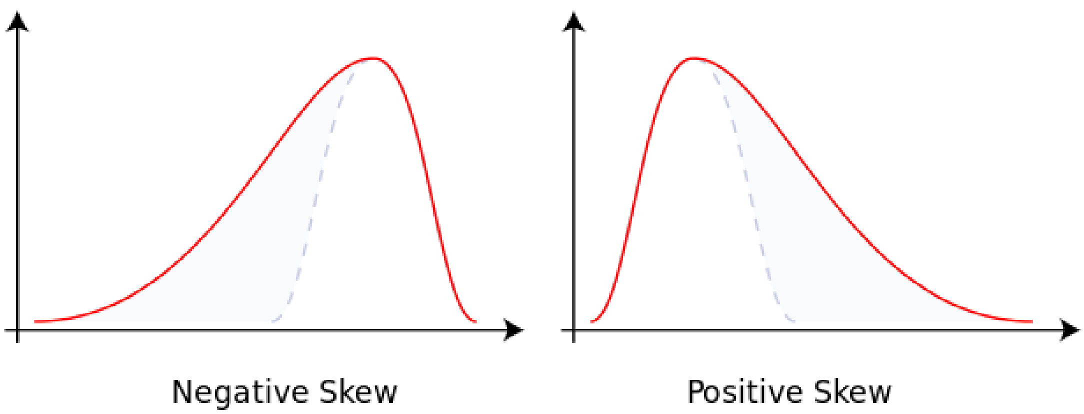
When projecting this to a whole population, you add/subtract the margin of error to give a range of possible results.

With a margin of error of 3%, you are confident that between 52% and 58% will vote "Yes".

With a margin of error of 10%, you are confident that between 45% and 65% will vote "Yes".

Skewness

Skewness is a distortion (an asymmetry) from the bell curve (normal distribution).



Types:

- Positive Skew: Tail on the right side is longer.
- Negative Skew: Tail on the left side is longer.

```
df['Price'].skew()
```

```
0.20860884190692716
```

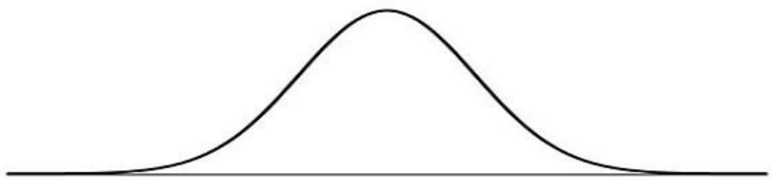
```
df['Year_Built'].skew()
```

```
0.05817114466569298
```

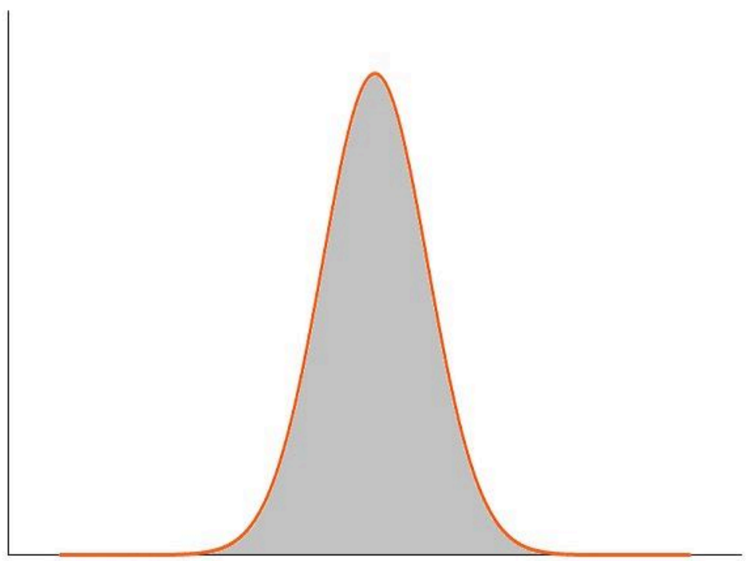
Kurtosis

Kurtosis is also a distortion from the normal distribution (bell curve).

While skewness describes unexpected values in one tail, kurtosis describes unexpected values in both tails.



Negative kurtosis (lower than normal distribution)



Positive kurtosis (higher than normal distribution)

```
df['Price'].kurt()
```

```
-0.15354674728688567
```

```
df['Year_Built'].kurt()
```

```
-1.2388754892468503
```

Probabilities

Probability is defined as the chance of happening or occurrences of an event. Generally, the possibility of analyzing the occurrence of any event with respect to previous data is called probability. For example, if a fair coin is tossed, what is the chance that it lands on the head? These types of questions are answered under probability.

Probability for Data Science

Probability plays a crucial role in data science, and it is used in many different ways to understand and analyze data, make predictions, and model complex systems. We will explore some of the ways in which probability is used in data science and the concepts and tools that data scientists use to work with probability with examples.

- **Statistical Inference :** Probability is used in statistical inference to make predictions about future events based on past data. For example, a data scientist might use a normal distribution to model the height of adult males in a population. By understanding the parameters of this distribution, the data scientist can make predictions about the height of an individual man based on a sample of heights from the population.

- **Machine Learning :** Probability is also used in machine learning to estimate the likelihood of certain outcomes and make predictions. For example, a data scientist might use maximum likelihood estimation (MLE) to train a machine learning model to recognize images of dogs. The data scientist would use a set of labelled images of dogs and non-dogs to train the model, and the MLE algorithm would estimate the probability of an image being a dog based on its features.

- **Bayesian Statistics :** Bayesian statistics is a branch of statistics that uses probability to update our beliefs about the world based on new data. For example, a data scientist might use a Bayesian model to predict the success rate of a new product. The data scientist would start with a prior belief about the success rate, and as new data becomes available, the model would update the success rate estimate using Bayes theorem.

- **Monte Carlo Methods :** Monte Carlo methods are a class of computational algorithms that use random sampling to solve mathematical problems. For example, a data scientist might use a Monte Carlo method to estimate the probability of an event occurring in a complex system. The data scientist would simulate the system many times with different random inputs, and the probability of the event occurring would be estimated based on the proportion of simulations in which the event occurred.

- **Bayes Theorem :** Bayes theorem is a mathematical formula that describes how a belief in an event should change based on new evidence. For example, a data scientist might use Bayes theorem to improve the accuracy of a spam filter. The data scientist would start with a prior belief about the probability of an email being spam, and as new data becomes available, the filter would update its belief about the email using Bayes theorem.

Dependent Events

Dependent events are those events that are affected by the outcomes of events that had already occurred previously. i.e. Two or more events that depend on one another are known as dependent events. If one event is by chance changed, then another is likely to differ. Thus, If whether one event occurs does affect the probability that the other event will occur, then the two events are said to be dependent. For example:

1. Let's say three cards are to be drawn from a pack of cards. Then the probability of getting a king is highest when the first card is drawn, while the probability of getting a king would be less when the second card is drawn. In the draw of the third card, this probability would be dependent upon the outcomes of the previous two cards. We can say that after drawing one card, there will be fewer cards available in the deck, therefore the probabilities tend to change.
2. A card is chosen at random from a standard deck of 52 playing cards. Without replacing it, a second card is chosen. What is the probability that the first card chosen is a king and the second card chosen is a queen?

Probabilities :

- P (king on the first pick) = 4 /52
- P (queen on 2nd pick given king on 1st pick) = 4 /51
- P (king and queen) = (4/52 × 4/51) = 16/2652 = 4 /663

It involved two compounds, dependent events. The probability of choosing a queen on the second pick given that a king was chosen on the first pick is called a conditional probability. When the occurrence of one event affects the occurrence of another subsequent event, the two events are dependent events. The concept of dependent events gives rise to the concept of conditional probability.

Conditional Probability

Conditional probability is a concept in probability theory that deals with the probability of an event occurring given that another event has already occurred. It is represented by the notation $P(A|B)$, where A and B are the events in question. The conditional probability of an event A given that event B has already occurred is the probability that event A will happen, assuming that event B has already happened.

If the probability of events A and B are $P(A)$ and $P(B)$ respectively then the conditional probability of A such that B has already occurred is $P(A|B)$. Given, $P(B)>0$,

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

$P(B) = 0$ means B is an impossible event. In $P(A \cap B)$ the intersection denotes a compound probability.

Independent events

Independent events are those events whose occurrence is not dependent on any other event. If the probability of occurrence of an event A is not affected by the occurrence of another event B, then A and B are said to be independent events.

Examples :

- Tossing a coin

  Here, Sample Space S = {H, T}, and both H and T are independent events.

- Rolling a die

  Sample Space S = {1, 2, 3, 4, 5, 6}, all of these events are independent too.

Mutually Exclusive Events

In an experiment, if the occurrence of an occasion precludes or rules out the happening of all the opposite events in the same experiment. Two events are said to be mutually exclusive if they can't occur at an equivalent time or simultaneously. In other words, mutually exclusive events are called disjoint events. If two events are considered disjoint events, then the probability of both events occurring at an equivalent time is going to be zero.

Let's understand with examples:

- When a coin is tossed either head or tail will appear. Head and tail cannot appear simultaneously. Therefore, the occurrence of a head or a tail is two mutually exclusive events.
- In throwing a die all the 6 faces numbered 1 to 6 are mutually exclusive since if any one of these faces comes, the possibility of others in the same trial is ruled out.

Joint Probability

Joint probability is a concept in probability theory that deals with the probability of two or more events occurring simultaneously. It is represented by the notation P(A and B), where A and B are the events in question. The joint probability of two events tells us the likelihood of both events happening together.

For example, consider a scenario where we are flipping a coin and rolling a dice. The probability of getting heads on a coin flip is 0.5 and the probability of rolling a 3 on a dice is 1/6. If we want to know the probability of getting heads on the coin flip and rolling a 3 on the dice, we would use the joint probability formula, which is P(heads and 3) = P(heads) * P(3). In this case, the joint probability would be 0.5 * (1/6) = 1/12.

Joint probability is also used to calculate the probability of mutually exclusive events. For example, if we have two events A and B, and it is impossible for both events to occur at the same time, the joint probability would be 0. For example, the probability of rolling a 2 on a dice and getting a 7 on a dice is 0 as it is not possible to get both numbers at the same time.

Bayes' Theorem

Bayes theorem is also known as the Bayes Rule or Bayes Law. It is used to determine the conditional probability of event A when event B has already happened. The general statement of Bayes' theorem is "The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B." i.e.

$$P(A|B) = P(B|A)*P(A) / P(B)$$

Where,

P(A) and P(B) are the probabilities of events A and B

P(A|B) is the probability of event A when event B happens

P(B|A) is the probability of event B when A happens

Bayes Theorem Statement

Bayes' Theorem for n set of events is defined as,

Let $E_1, E_2, \ldots, E_n$ be a set of events associated with the sample space S, in which all the events $E_1, E_2, \ldots, E_n$ have a non-zero probability of occurrence. All the events $E_1, E_2, \ldots,$ E form a partition of S. Let A be an event from space S for which we have to find probability, then according to Bayes' theorem,

$$P(E_i|A) = P(E_i)*P(A|E_i) / \sum P(E_k)P(A|E_k)$$

for k = 1, 2, 3, …., n

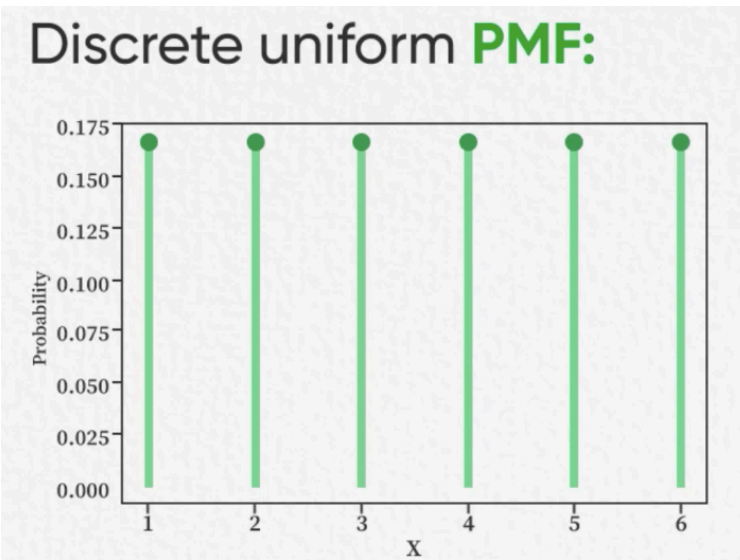
Distributions

Discrete Uniform Distribution

The Discrete Uniform Distribution is the probability distribution where probability of each possible outcome is the same, and there is a finite number of outcomes.

It is used in Data Science to model situations where all the outcomes are equally likely.

A Simple example of the Discrete Uniform Distribution is throwing a simple dice, the Discrete Uniform Probability Mass Function would be



Binomial Distribution

Bernoulli Trial :-

A Bernoulli trial is a random experiment with only two possible outcomes: "Success" and "Failure". These trials are named after Jacob Bernoulli, a Swiss mathematician. The outcomes are typically denoted as 1 for success and 0 for failure.

Example -

1. Flipping a coin - If we Flip a coin there are only two outcomes heads and tails, say we assign head to be success then upon flipping the coin if we get heads then it is success, if it is a tail then it is a failure.
2. Getting a 6 on a Dice - If we roll a die and get the outcome as 6, then it is success. Anything else would be considered a failure.

The binomial distribution is a discrete probability distribution that describes the number of successes in a fixed number of independent and identically distributed Bernoulli trials. Each trial has only two possible outcomes: success (usually denoted as 1) and failure (usually denoted as 0).

Example

1. Flipping a coin 10 times.
2. Getting a 6 on a dice upon 10 rolls of dice

What are the applications of Binomial Distribution in Data Science ?

Some applications of Binomial Distributions are

- A/B Testing :- If you are running an A/B test on a website, you can use the binomial distribution to model the number of visitors who convert to customers. By choosing one as success and other as failure.
- Risk Management :- The Binomial Distribution can be used in risk assessment to model the probability of a certain number of successes or failures in a fixed number of trials.
- Medical Trial :- A Drug Trial is conducted to determine the efficacy of a new drug. The drug is given to a fixed number of patients, and the number of patients who respond positively can be modeled using binomial distribution.

Poisson Distribution

The Poisson distribution is a probability distribution that describes the number of events that occur within a fixed interval of time or space. It is named after the French mathematician Siméon Denis Poisson, who introduced it in the early 19th century.

Let's Take examples to understand the concept better.

- If someone eats twice a day, What is the Probability he will eat thrice a day
- In a class of 10 students, on an average 4 students ask doubts in each class, what is the likelihood that on a particular day 7 students ask for doubt.

We use Poisson Distribution in situations where the probability of an event occurring is small, but the number of possible events is very large.

What are the used cases of Poisson Distribution ?

Some Used Cases of Poisson Distribution are -

1. Web Traffic Analysis
2. Call Center Analysis
3. Heath Analysis