Summary

Probability and Probability Distributions

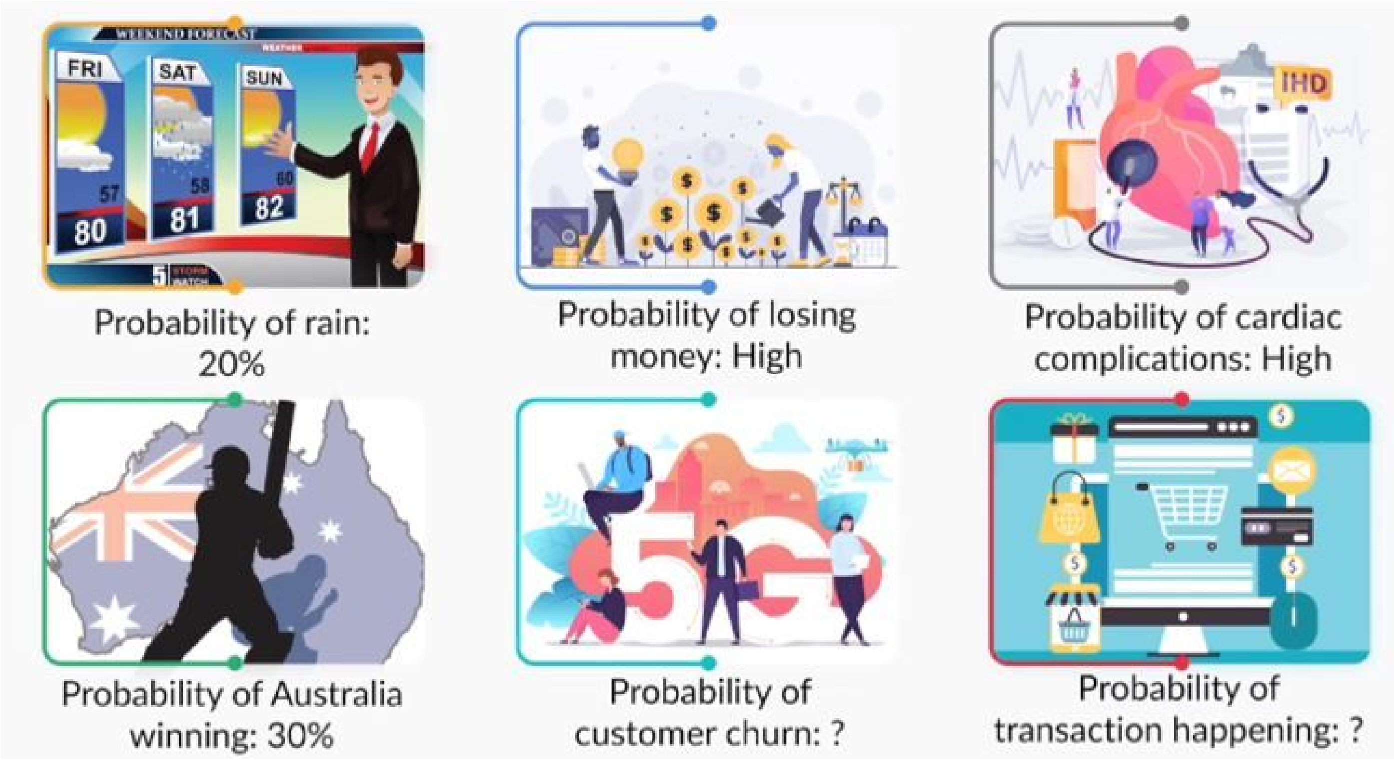
Inferential statistics involves picking a sample out of the population data set and then trying to understand the population based on the sample. Naturally, the samples might not be representative of the entire population. Hence, the inference made from any sample might not accurately describe the population. However, it is possible to assign a probability value to the accuracy of any inference made about the population from the sample. This session focuses on establishing the theoretical frameworks necessary to explore such topics in greater detail.

In this session ,you learnt how to approach statistically uncertain events and focused on the following topics:

1. Probability
2. Random variables
3. Probability distributions
4. Binomial variable
5. Normal distribution
6. Standard normal distribution and z-scores

Introduction to Probability

Probabilities are used to convey information about the likelihood of an event. These events can occur in our day-to-day lives as described by the examples presented in the image given below.



Probability is a measure of uncertainty that helps you understand the chance that a certain event can happen out of all the possible outcomes. For example, it might or might not rain on any given day. Hence, there are two possible outcomes–rain and no rain–-and each of these two outcomes has a certain probability associated with it. Similarly, when you toss a coin, the two possible outcomes are heads and tails, and each of these two outcomes has a certain probability associated with it.

Calculation of Probability

- Probability of an event = Number of favourable outcomes / Total number of equally likely outcomes

It is important to remember that the formula is applicable only when **all the possible outcomes are equally likely**.

For example, on any given day, it might or might not rain. Hence, you have two outcomes. However, based on this information, you **cannot** conclude that the probability of rain on any given day is ½, as the events are not equally likely to happen. On the other hand, in the case of a toss of a fair coin, both heads and tails are equally likely, and hence, the probability of getting heads (or tails) is ½.

Sometimes, we do not know if all the outcomes are equally likely. In that case, we need other methods of arriving at a probability value for an event. The different methods for arriving at a probability value of an event are:

1. The **classical/theoretical approach** for calculating probability is the one in which certain assumptions are made (for instance, all possible outcomes are equally likely), and the probability

values are then calculated using a formula. One does not need to perform any experiments or gather any data when following this approach to arrive at the probability of an event. Hence, if you assume that a coin is fair, then you can conclude that the probability of getting heads (or tails) is ½.
2. In the **empirical/frequentist approach**, probabilities are derived from observations. Hence, if you want to find the probability of getting heads on tossing a coin, then you would toss the coin several times and note down the outcome of each trial. Suppose you tossed the coin 10,000 times and got heads 5,052 times and tails 4,948 times. Following the frequentist approach, you would conclude that the probability of getting heads is 0.5052 and that of getting tails is 0.4948 for that particular coin.
3. Another way of arriving at the probability of an event is the **subjective approach**, which reflects an individual's understanding and judgement of how likely the outcomes are.

The basic rules of probability are:
- The value of probability of any event always lies between 0 and 1.
- If two events, A and B, are **mutually exclusive** (both the events cannot occur at the same time), then the probability that either of them occurs is equal to the sum of their individual probabilities.
    - **P(A or B) = P(A) + P(B)**; if A and B are mutually exclusive events
- If two events, A and B, are **independent** (the occurrence of one event does not affect the probability of occurrence of the other), then the probability that both of them occur is equal to the multiplication of their individual probabilities.
    - **P(A and B) = P(A)*P(B)**; if A and B are independent events

Random Variables

A random variable maps the outcome of a random process to a numerical value. Suppose you are tossing a coin, which is a random process. You can define a random variable X that takes the value 10.5 if the outcome is 'heads' and 15.7 if the outcome is 'tails'. In this case, the probability of getting heads will be denoted by P(X = 10.5) and that of getting tails will be denoted by P(X = 15.7). If you use a fair coin, then both these values will be equal to ½.

It is important to note that even if a process inherently produces numerical outcomes, such as the roll of a die, a random variable can be defined such that it alters the numerical outcomes. For example, define a random variable Y as twice the number that appears on the face of a die when you roll it. In this case, your random variable Y takes the values 2, 4, 6, 8, 10 and 12 when the face of the die shows 1, 2, 3, 4, 5 and 6, respectively.

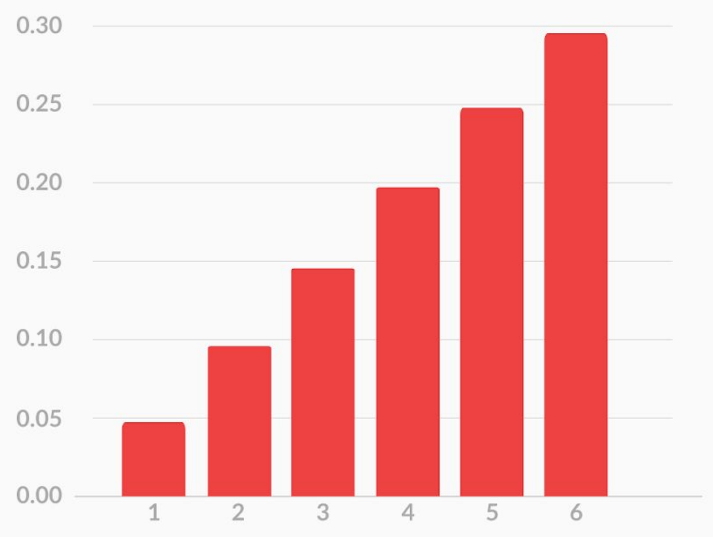
Discrete Random Variables

Probability distribution is a form of representation that indicates the probability of **all** the possible values of a random variable. A probability distribution could be any of the following:
- An equation:
    - P(x) = x/21
    - (for x = 1, 2, 3, 4, 5 and 6)

- A table:

x	P(x)
1	1/21
2	2/21
3	3/21
4	4/21
5	5/21
6	6/21

- A graph:



The **sum of the probabilities of all of the outcomes must be equal to 1**.

- If the probability of getting heads on tossing a coin is 0.7, then that of not getting heads (i.e., getting tails) must be 0.3 so that the sum of the probabilities of all the possible outcomes equals 1.

The **mean** of the random variable is also referred to as the **expected value** of the random variable.

- The expected value of a random variable X is the average value of X that you would 'expect' to obtain after performing the experiment for an infinite number of times.

The **variance** of X gives you an indication of the spread of the values from the mean that the random variable can take. The higher the variance, the higher is the number of the values that are spread out from the mean of the distribution.

The **standard deviation** is also a measure of the dispersion of the values of the random variable from its mean.

Binomial Distribution

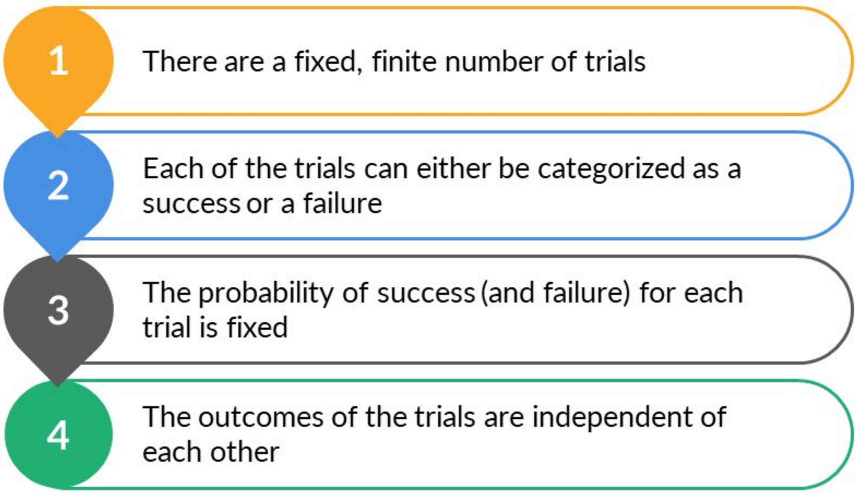
^nC_x is a formula that tells you the number of ways in which x items can be chosen out of n items, which is as follows:

- $\mathbf{^nC_x = n!/(n - x)! * n!}$

The factorial is defined only for non-negative integers, which can be given by the following formula:

- **n! = n * (n - 1) * (n - 2) * ... * 2 * 1**, where **0! = 1**

The conditions that need to be met for a random variable to be called a binomial variable are mentioned in the image given below.



The outcomes of a binomial distribution are classified as successes or failures; you can use combinatorics to compute the number of ways in which 'r' successes can occur in 'n' trials. The binomial distribution can be expressed as follows:

- $\mathbf{P(X = r) = {^nC_r} * (p)^r * (1 - p)^{n-r}}$

where

n = number of trials

r = number of successes

p = probability of success

As ‘p’ is the probability of success, ‘1-p’ is the probability of failure. Clearly, the Multiplication Rule of Probability, which requires us to assume that the outcomes of each of the trials are independent of each other, has been applied in deriving the formula for the binomial distribution.

Continuous Random Variables

The probability distribution of a continuous random variable is called a probability density function (PDF).

The properties of a probability density function (PDF) are as follows:

1. The probability of getting any single point is zero.
2. The probability of an interval is the area under the PDF curve in that interval.
3. The total area under the PDF curve is always equal to 1.

The cumulative probability of a point is the probability of any value occurring that is lower than or equal to that point. It is denoted by F(x) and can be represented in mathematical terms as follows:

- **Cumulative probability at x = F(x) = P(X ≤ x)**

As the PDF never takes a negative value, the cumulative probability function is a non-decreasing function. This means that the value of the cumulative probability function either stays the same or increases as you move to the right in the graph.

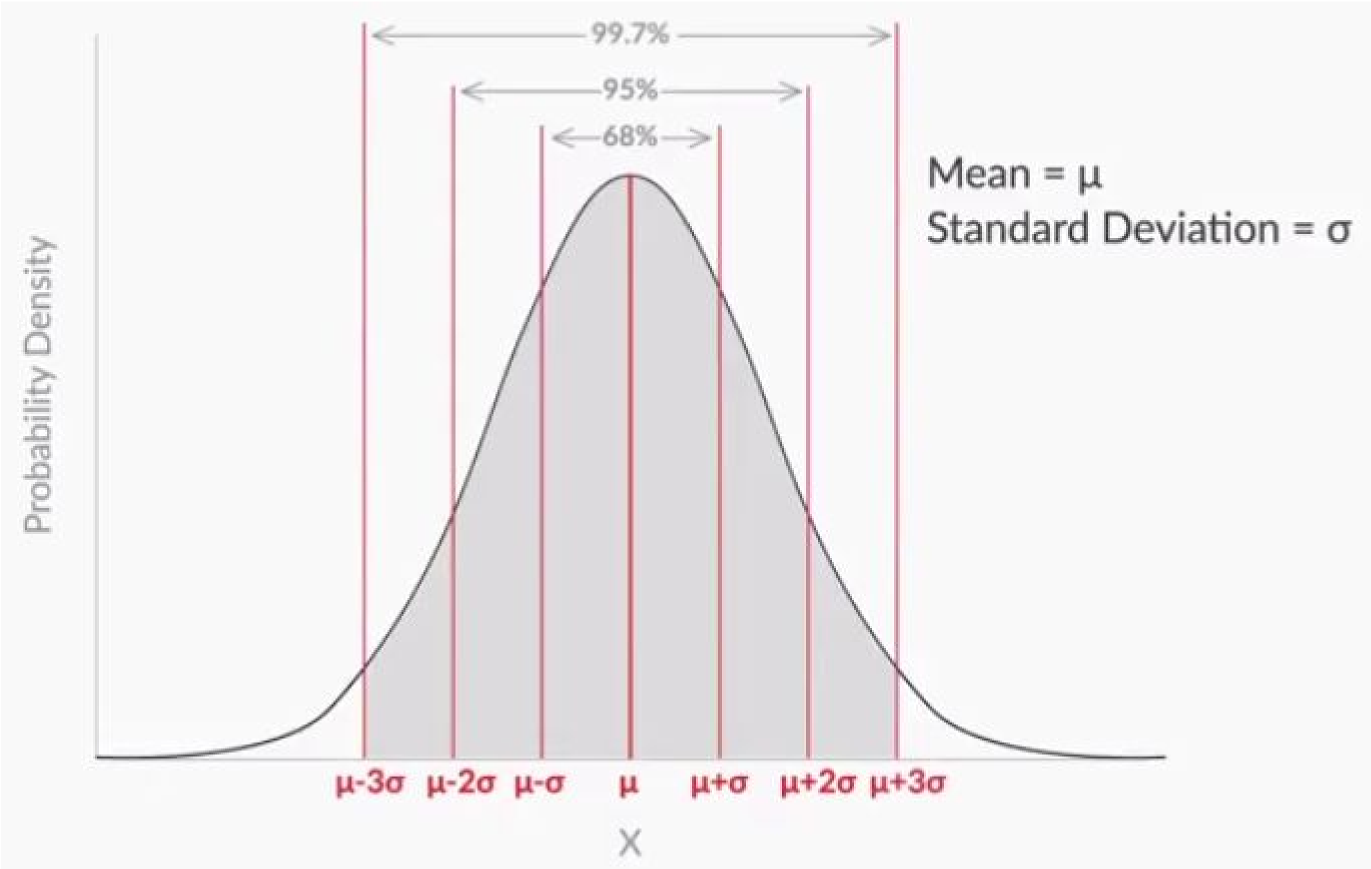
Normal Distribution

The most commonly occurring continuous probability distribution is a normal distribution. Several natural phenomena, such as the height of men/women of a certain age, blood pressure and IQ scores follow a normal distribution.

Normal distribution is symmetric with respect to its mean and extends infinitely on both sides.

In a normal distribution, the probability density is higher close to the mean and decreases exponentially as you move further away from the mean. In simple language, it means that there is a high probability that the value of the random variable is close to the mean. As you move further away from the mean, the probability of the occurrence of such values decreases.

The empirical rule is illustrated in the image given below.



The empirical rule states that there is:

1. a 68% probability of the variable lying within one standard deviation of the mean,
2. a 95% probability of the variable lying within two standard deviations of the mean, and
3. a 99.7% probability of the variable lying within three standard deviations of the mean.

In other words, in case you look at the outcome of one trial at random, there is a 68% chance that the outcome lies within one standard deviation of the mean, 95% chance that the outcome lies within two standard deviations of the mean and 99.7% chance that the outcome lies within three standard deviations of the mean.

It is important to note that the numbers specified in the empirical rule are only **approximations**.

Standard Normal Distribution

The standard normal distribution is a normal distribution with a mean of 0 and standard deviation of 1.

The formula used to convert any point in a normal distribution into its equivalent point (referred to as the **z-score**) in the standard normal distribution is as follows:

- **Z-score = $x - \mu / \sigma$**

An important point to note is that the properties of the original point on a normal distribution are retained when it is translated on to the standard normal distribution. For instance, if the cumulative probability of x is 0.813 in a normal distribution, the cumulative probability of its equivalent z-score will also be 0.813 in the standard normal distribution.

The z-score tells you how many standard deviations away your observed value is from the mean. This is extremely helpful in probability calculations and for comparing points on two different normal distributions.

The ‘NORM.DIST’ function in Excel can be used to calculate the cumulative probability of any point on a normal distribution. Similarly, the ‘NORM.S.DIST’ function can be used to calculate the cumulative probability of any point on the standard normal distribution.