Inferential Statistics

URL for reference : https://jovian.com/sharathdeepu9751/inferential-statistics-notes-and-graded-questions#C67

Module 1 : Introduction to Probability

Introduction: Inferential Statistics:-

Welcome to the module on '**Inferential Statistics**'. In the last module on **EDA**, you learnt how to explore data and derive insights from its exploration.

In this module

Exploratory data analysis helped you understand how to **discover patterns in data** using various techniques and approaches. As you learnt, EDA is one of the most important parts of the data analysis process. It is also the part on which data analysts spend most of their time.

However, sometimes, you may require a very large amount of data for your analysis, which may need too much time and resources to acquire. In such situations, you are forced to work with a **smaller sample of the data** instead of having to work with the entire data.

Situations like these arise all the time at big companies like Amazon. For example, let's say the Amazon QC department wants to know what proportion of the products in its warehouses are defective. Instead of going through all of its products (which would be a lot!), the Amazon QC team can just check a small sample of 1,000 products and then find, for this sample, the defect rate (i.e., the proportion of defective products). Then, based on this sample's defect rate, the team can 'infer' what the defect rate is for all the products in the warehouses.

This process of 'inferring' insights from sample data is called 'inferential statistics'.

Note that even after using inferential statistics, you will arrive at only an estimate of the population data from the sample data, not the exact values. This is because when you don't have the exact data, you can only make reasonable estimates about it with a limited level of certainty. Therefore, when certainty is limited, we talk in terms of probability and in the first session of this module we will explain to you the basics concepts of probability which are useful and important in inferential statistics.

In this session

In this session, you will learn the basic concepts of probability and the various rules associated with it. The broad agenda of the session covers the following:

- Permutation and combination

- Definition of probability and its properties
- Key terms related to probability
- Probability rules (Addition and Multiplication)

Permutations:-

Before we turn our focus to probability, it is important to understand some basic tools that form its essential building blocks. This session will require you to do a fair bit of calculations, so be prepared.

One of the key things that you need to master is the idea of the two major counting principles – **permutations and combinations**. Knowing these two concepts will enable you to calculate the probability for a given scenario or events that you are interested in. We will start by understanding the concept of 'permutations' in the upcoming video.

To summarise the video, a permutation is a way of arranging a select group of objects in such a way that the **order is of significance**. As shown in the example, when you arrange the top order batsmen of a cricket team, you use permutation to find all the possible orders in which they can be arranged. The following list shows some other examples where permutation is used to count the number of ways in which a particular sequence of events can occur:

- Finding all possible four-letter words that can be formed using the alphabets R, E, A and D
- Finding all possible ways in which the final league standings of the eight teams can be in an Indian Premier League (IPL) tournament
- Finding all possible ways that a group of 10 people can be seated in a row in a cinema hall, and so on

Generally speaking, if there are n 'objects' that are to be arranged among r available 'spaces', then the number of ways in which this task can be completed is **n!/(n-r)!**. If there are n 'spaces' as well, then the number of ways would be just **n!**. Here **n!** (pronounced as n factorial) is simply the product of all the numbers from n till 1 and is given by the following formula:

$$n! = n*(n-1)*(n-2)....*3*2*1.$$

For now, we will skip the proof of the permutation formula. It is advised to verify and get acquainted with it for scenarios where n is small.

You have learnt to find the number of ways in which you can arrange items in an order using permutation. In the next segment on combinations, you will learn how to find the number of ways in which to choose a particular set of items.

Combinations:-

The second important counting principle that you need to be aware of is the method of using **combinations**. In the case of counting using the method of permutations, you had considered the 'order' to be an important factor. Now, in the case of combinations, you need not take the order into account while finding the number of ways to arrange a group of objects. Let's watch the following video as Amit explains the concept.

As explained in the video, when you just have to choose some objects from a larger set and the **order is of no significance**, then the rule of counting that you use is called **combination**. In the example mentioned in the video, you had to choose three bowlers from a set of four bowlers and obviously you did not need to order them here. Some other examples of combinations are as follows.

- The number of ways in which you can pick three letters from the word 'UPGRAD'
- The number of ways a team can win three matches in a league of five matches
- The number of ways in which you can choose 13 cards from a deck of 52 cards, and so on

The formula for counting the number of ways to choose r objects out of a set of n objects is as follows:

$$^{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Now, you might be wondering when to use permutations and when to use combinations. As mentioned in the video, one way to look at it is to see if the order matters or not. If it does, then use the permutations formula, and if does not, then use the one for combinations.

Note: A helpful hint here would be to look for a **keyword** in the given scenario to know which method is needed. If the problem requires you to **order/ arrange** a group of objects, then you would most probably use the method of permutations. Else, if you are told to **pick/choose** a group of objects, then more often than not you would be using the formula for combinations.

There are some other rules of counting as well. For example, recall that in the permutation case, we had assumed that no repetition was allowed in the order and, hence, we proceeded with the given formula. Now, what do we do if repetition is allowed in the process of counting the number of ways? You will get to know those types of examples in future sessions and will also learn how to find the answer in such cases. For the time being, you need to know only these two methods, i.e., permutation and combination.

Now, answer the following questions.

Now that you have learnt the two fundamental counting rules, we will go ahead and finally learn the basic definition of probability and its associated properties.

Probability: Definition and Properties:-

Probability = predicts the likelihood of future event(i.e. it measure the chances of something happening.).

Now that you have understood the two fundamental rules of counting, we will go ahead and finally establish the formal definition of probability. Let's hear from Amit as he explains the same with the help of an example in the upcoming video.

As explained in the video, the formula for calculating the probability is as follows:

$$Probability = \frac{No\ of\ desired\ outcomes}{Total\ number\ of\ possible\ outcomes.}$$

Now, there are some additional concepts and properties of probability that you need to know in order to understand it better. Let's hear from Amit as he discusses the same in the upcoming video.

As explained in the video, probability values have the following two major properties:

- **Probability values always lie in the range of 0 to 1**. The value is 0 in the case of an impossible event (like the probability of you being in Delhi and Mumbai at the same time) and 1 in the case of a sure event (like the probability of the sun rising in the east tomorrow).
- **The probabilities of all outcomes for an experiment always sum up to 1**. For example, in a coin toss, there can be two outcomes, heads or tails. The probability of both of the outcomes is 0.5 each. Hence, the sum of the probabilities turns out to be 0.5 + 0.5 = 1.

Next, you learnt a couple of definitions that are crucial in understanding probability. They are as follows:

- **Experiment**: Essentially, any scenario for which you want to compute the probabilities is considered to be an experiment. It is of the following two types:
    - **Deterministic:** Outcome is the same every time and determined.

    - **Random:** Outcome can take many possible values. Throughout the majority of our business analytics course, we will only be discussing the random experiment.
- **Sample space:** A sample space is nothing but the list of all possible outcomes of a random experiment. It is denoted by S = {all the possible outcomes}. For example, in the coin toss example, the sample space is S = {H, T}, where H = heads and T = tails.
- **Event:** It is a subset, i.e., a part of the sample space that you want to be true for your probability experiment. For example, if in a coin toss you want heads to be the desired outcome, then the event becomes {H}. As you can see clearly, {H} is a part of {H, T}.

Now, if you observe the definitions carefully, you will see that the probability formula can be modified as follows:

$$Probability = \frac{Total\ Number\ of\ outcomes\ in\ event}{Total\ number\ of\ outcomes\ in\ sample\ space}$$

Now, the counting principles that you learnt earlier will help you compute the total number of outcomes in both the sample space and the event that you are interested in. Solve the following questions in order to drive home these concepts.

Types of Events:-

In the previous segment, you understood how probability is defined formally and also learnt some of its properties. You were also introduced to the concept of 'events', situations or scenarios for which we compute the probabilities. In this segment, we will take a look at the different types of 'events' that can be defined. Note that earlier you considered a single event for which you computed the probabilities. Now, you will look at two or more events and understand how they are related to each other.

The two main categories of events that you need to know right now are **independent** events and **disjoint** or **mutually exclusive** events. Let's learn their formal definitions.

- **Independent events:** If you have two or more events and the occurrence of one event has no bearing whatsoever on the occurrence/s of the other event/s, then all the events are said to be

independent of each other. For example, the chances of rain in Bengaluru on a particular day has no effect on the chances of rain in Mumbai 10 days later. Hence, these two events are independent of each other.
- **Disjoint or mutually exclusive events:** Now, two or more events are mutually exclusive when they do not occur at the same time, i.e., when one event occurs, the rest of the events do not occur. For example, if a student has been assigned grade C for a particular subject in an exam, he or she cannot be awarded grade B for the same subject in the same exam. So, the events in which a student gets a grade of B or C for the same subject in the same exam are mutually exclusive or disjoint.

Now, what about events that are both independent and mutually exclusive? Do such events exist? Let's hear from Amit as he explains the difference between the two in the upcoming video.

As explained in the video, two or more events cannot be independent and disjoint simultaneously. Two events can be either exclusively independent or exclusively disjoint. Here are some examples to drive home this point.
- The events 'Customer A buys the product' and 'Customer B buys the product' are independent, whereas the events 'Customer A buys the product' and 'Customer A does not buy the product' are disjoint.
- The events 'You will win Lottery A' and 'You will win Lottery B' are independent, whereas 'You will win Lottery A' and 'You will not win Lottery A' are disjoint events.

Complement Rule for Probability

Now, disjoint events have one special property that is pretty intuitive and easy to understand. For example, let’s say A and B are two disjoint events. If A = 'Event that it rains today' and B =' Event that it does not rain today' and you know the P(A) = 0.3, can you guess what P(B) might be?

You must have guessed the answer here. It is nothing but 1 - P(A) = 1 - 0.3 = 0.7. This is something known as the **complement rule for probability**. It states that if A and A' are two events which are mutually exclusive/disjoint and are complementary/in negation of each other (you can read **A'** as '**not A**'), then:

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

In the mentioned example, B is the complement of A and hence P(B) = 1 - P(A) = 1 - 0.3 = 0.7. Here are some examples where you can use this rule to find the probability of the complement of an event.
- If the probability that a customer buys a product is 0.4, then the probability that he/she does not buy the product is 0.6.
- If the probability that you win the lottery is 33%, then the probability that you do not win the lottery is 67%, and so on.

This rule is basically an extension from the basic rule of probability that you learnt previously – **the sum of probabilities for all events always add up to 1**. You can read more about the complement rule for probability here.

Now, answer the following questions to strengthen your conceptual understanding of the aforementioned topics.

Comprehension

Box8 is an online food-ordering app that operates in Bangalore, Delhi-NCR, Mumbai, Pune and Hyderabad. Next week, it is going to launch a month-long marketing campaign for its membership program – **Box8 Pass** – that gives access to a flat 25% discount on all orders and access to other exclusive discounts and benefits in the city of Mumbai.

Now, the company's lead business analyst has estimated that there is a high chance that they will gain around 500 new subscribers at the end of this campaign. She also calculated that they need at least 200 new subscribers to make sure that they recuperate the costs of the campaign and break even.

Given this information, answer the following questions:

Now that you have learnt two of the most common types of events that you generally encounter – **independent** and **disjoint** events – you will learn some rules to compute the probabilities of those events using the addition and the multiplication rules.

Rules of Probability - Addition:-

In the previous examples, we were always involved in finding the probability of a single event. For example:

- The probability of getting an ace card from a deck of 52 cards
- The probability of selecting three bowlers from a list of four, and so on

Now, what about two events occurring simultaneously? For example, what would be the **probability of selecting an ace card or a heart card** from the deck? How would this specific probability relate to the individual probabilities, i.e., the probability of getting a heart card separately and the probability of getting an ace card separately? This is something that you will get to learn using the two important rules related to probability – addition and multiplication. First, we will discuss the **addition rule** in the upcoming video.

So, to summarise the learnings from the video, when you have the individual probabilities of two events A and B, denoted by **P(A)** and **P(B)**, the addition rule states that the probability of the event that either A or B will occur is given by:

$\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$, where, **P(A∪B)** denotes the probability that either event A or B occurs.

P(A) denotes the probability that only event A occurs

P(B) denotes the probability that only event B occurs

P(A∩B) denotes the probability that both events A and B occur simultaneously.

Note: The symbols ∪ and ∩ are obtained from the world of '**set theory**' and are used to denote **union** and **intersection**, respectively. You do not need to learn about them in detail right now. All you need to learn are the meanings of the probability terms mentioned earlier. Also, we would be skipping the proof of the formula right now. Another important thing to note here is that the formula mentioned earlier works for all types of events A and B, irrespective of the fact that they are independent or disjoint.

You can also read **P(A∪B)** as **P(either event A or B occurs)** and **P(A∩B)** as **P(both events A and B occur).**

As mentioned in the video, for disjoint events A and B, **P(A∩B)** = 0 since both cannot occur simultaneously. Hence, the formula can be rewritten as **P(A∪B) = P(A) + P(B)**.

Now, answer the following questions.

In the next segment you will learn about the multiplication rule of probability.

Rules of Probability - Multiplication:-

Now that you have understood the addition rule and how it can be used to compute the probabilities of two events A and B, let's now understand the concept of the multiplication rule. This rule is applicable only on independent events about which you have already learnt before. Let's take a look at how you can compute the probabilities of two events A and B occurring simultaneously.

As discussed in the video as well as in the previous segments, when an event A is not dependent on event B and vice versa, they are known as independent events. The multiplication rule allows us to compute the probabilities of both of them occurring simultaneously, which is given as:

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

Now, this rule can be extended to multiple independent events where all you need to do is multiply the respective probabilities of all the events to get the final probability of all of them occurring simultaneously. For example, if you have four independent events A, B, C and D, then:

P(A and B and C and D) = P(A)*P(B)*P(C)*P(D).

Comparison Between Addition Rule and Multiplication Rule

Both the addition rule and the multiplication rule allow you to compute the probabilities of the occurrence of multiple events. However, there is a key difference between the two, which should help you to decide when to use which rule.

- The addition rule is generally used to find the probability of multiple events when **either of the events can occur at that particular instance**. For example, when you want to compute the probability of picking a face card or a heart card from a deck of 52 cards, a successful outcome occurs when either of the two events is true. This includes either getting a face card, a heart card, or even both a face and a heart card. This rule works for all types of events.
- The multiplication rule is used to find the probability of multiple events when all the events need to occur simultaneously. For example, in a coin toss experiment where you toss the coin three times and you need to find the probability of getting three heads at the end of the experiment, a successful outcome occurs when you get a head in the first, second and third toss as well. This rule is used for independent events only.
- Also, in the addition rule, do you remember the **P(A∩B)** that we used to compute the final value of P(A∪B)? This value is **exactly the same as the P(A and B)** that we compute in independent events using the multiplication rule. You can go back and verify it for the same example shown in the video. There we had P(Heart Card) = P(H) = 13/52, P(Face Card) = P(F) = 12/52 and P(Heart Card and Face Card) = P(H∩F) = 3/52. Now, as mentioned by the multiplication rule, you can see that P(H and F) = P(H)*P(F) = (13/52)*(12/52) = 3/52, which is the same as the value of P(H∩F).

Note: A helpful hint here to decide when to use the addition rule and when to use the multiplication rule is to observe the language of the question. If the question mentions an 'OR' to denote the relationship between the events, then you need to apply the addition rule. That is, either of the given events can occur at that time, P(Event A or Event B). Else, if an 'AND' is used to denote the relationship between the events, then the multiplication rule should be used. Here, the events need to happen simultaneously and must be independent, i.e., P(Event A and Event B).

Now, answer the following questions to strengthen your concepts in the topics taught in the last two segments.

Summary:-

Here is a summary of what you have learnt in this session.

First, you were introduced to two main counting principles through which you can calculate the probability for the outcomes that you are interested in. These counting principles are as follows:

- **Permutations:** A permutation is a way of arranging a selected group of objects in such a way that **the order is of significance**. For example, arranging some letters to form different words, arranging the top order batsmen of a team, or finding all the ways in which a group of friends can be seated in a cinema hall – all these methods use permutations. When there are 'n' objects to be arranged among 'r' spaces, the permutation value is given by the following formula:

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

- **Combinations:** When you just have to choose some objects from a larger set and **the order is of no significance**, then the rule of counting that you use is called combination. For example, if you need to find the number of ways to choose three vowels from the given list of five, or choose four bowlers from the given roster of seven – all these methods use combinations. If you want to choose 'r' objects from a larger set of 'n' objects, then the number of ways in which you can do that is given by the following formula:

$$^{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Next, you learnt about some specific terms related to probability – events, sample space and experiments. You also learnt some basic rules of probability such as follows:

- Probability values always lie between 0 and 1.
- Probability values for all outcomes of an experiment always add up to 1.

Then, you learnt about specific types of events that can be encountered while computing probabilities of more than two events. They are as follows:

- **Independent events:** If you have two or more events and the occurrence of one event has no bearing whatsoever on the occurrence/s of the other event/s, then all the events are said to be independent of each other.
- **Disjoint or mutually exclusive events:** Now, two or more events are mutually exclusive when they do not occur at the same time, i.e., when one event occurs, the rest of the events do not occur.

You also learnt about the idea of complement **A'** for any event A and the probability rule for them, i.e., **P(A) + P(A') =1.**

Next, you were introduced to two main rules of probability that help you in finding the probabilities of two or more events. They are as follows:

- **Addition rule:** When you have the individual probabilities of two events A and B, denoted by **P(A)** and **P(B)**, the addition rule states that the probability of the event that either A or B will occur is given as
  - **P(A∪B) = P(A) + P(B) - P(A∩B),**

  Where, **P(A∪B)** denotes the probability that either event A or B occurs.
  P(A) denotes the probability that only event A occurs.
  P(B) denotes the probability that only event B occurs.
  P(A∩B) denotes the probability that both events A and B occur simultaneously.
- **Multiplication rule:** When an event A is not dependent on event B and vice versa, they are known as independent events. The multiplication rule allows us to compute the probabilities of both of them occurring simultaneously, which is given as:
  - **P(A and B) = P(A)*P(B).**

Now, this rule can be extended to multiple independent events where all you need to do is multiply the respective probabilities of all the events to get the final probability of all them occurring simultaneously. For example, if you have four independent events A, B, C and D, then:

P(A and B and C and D) = P(A)*P(B)*P(C)*P(D)

Finally, you also saw a comparison between the addition rule and the multiplication rule that would help you to decide which formula to use in a given scenario. If the question mentions an '**OR**' to denote the relationship between the events, then you need to apply the addition rule. That is, either of the given events can occur at that time, **P(Event A or Event B)**. Else, if an '**AND**' is used to denote the relationship between the events, then the multiplication rule is used. Here, the events need to happen simultaneously and must be independent, **P(Event A and Event B)**.

In the next segment we will provide the practice questions, answer them based on the concepts learnt in this session.

Module 2 : Basics of Probability

Introduction: Basics of Probability:-

Welcome to the session on '**Basics of Probability**'.

In this session

In this session, you will learn a few basic concepts of **probability** through a specific example. This session broadly covers the following topics:

- Random variables
- Probability distributions
- Expected value

Random Variables:-

Welcome to the first session on inferential statistics. This will be a very interactive session, with a lot of questions that will compel you to think about a concept, helping you explore it more actively.

So, let's get started.

Recall the original question: In the long run (i.e., if it is played a lot of times), is this game profitable for the players or for the house? Or, will everybody break even in the long run?

Recall that we established a three-step process for answering this question:

1. Find all the possible combinations.
2. Find the probability of each combination.
3. Use the probabilities to estimate the profit/loss per player.

We have completed step 1, which involves finding all the possible combinations. Now, let's proceed to step 2, which involves finding the probability of each combination. What are the steps involved in finding the probability? Let's hear more from Professor Tricha on this.

So, the **random variable X** converts the outcomes of experiments to measurable values.

For example, let’s say as a data analyst at a bank, you are trying to find out which of the customers will default on their loan, i.e., stop paying their loans. Based on some data, you have been able to make the following predictions:

Customer Number	Yearly Income (in ₹)	Amount of Loan Due (in ₹)	Number of Dependants	Default Prediction (Yes/No)
1	10 lakh	75 lakh	3	Yes
2	15 lakh	50 lakh	2	No
3	20 lakh	40 lakh	1	No

Now, instead of processing the yes/no response, it will be much easier if you define a random variable X to indicate whether the customer is predicted to default or not. The values will be assigned according to the following rule:

X = 1, if the customer defaults;
X = 0, if the customer does not default.

Now, the data changes to the following:

Customer Number	Yearly Income (in ₹)	Amount of Loan Due (in ₹)	Number of Dependants	X (Random Variable)
1	10 lakh	75 lakh	3	1
2	15 lakh	50 lakh	2	0
3	20 lakh	40 lakh	1	0

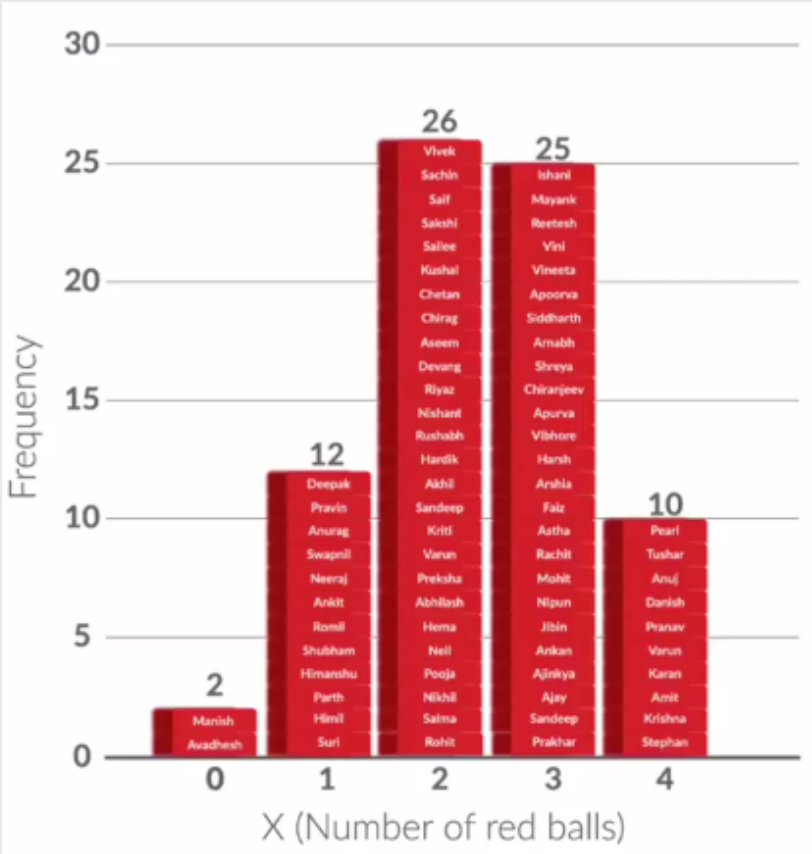
Now, in this form, the table is entirely **quantified**, i.e., converted to numbers. Now that the data is entirely in quantitative terms, it becomes possible to perform a number of different kinds of statistical analyses on it. In the next segment, you will learn about probability distributions.

Probability Distributions - I:-

Recall that in our upGrad game example, we need to find out if the game would be profitable for the players or for us (i.e., the house) in the long run. The three-step process for this is as follows:

1. Find all the possible combinations.
2. Find the probability of each combination.
3. Use the probabilities to estimate the profit/loss per player.

So far, we have completed step 1 and are at step 2, i.e., we are calculating the probability of each combination. For this, we defined a random variable X, which helped us convert the outcomes of our experiment to measurable values. Now, let's find the probability of each of these combinations.



So, we performed the experiment (i.e., played the game) 75 times and then made the **frequency distribution (histogram)**. Now, you may be thinking, "Well, I want to try that out too." Unfortunately, even if you do have a bag with 3 red balls and 2 blue balls, playing the game 300 times would be tedious and difficult. Well, that's not a problem. You can simulate the whole experiment. (**Note**: Please use your laptop browser to view the following distribution if you are unable to do so in the mobile app.)

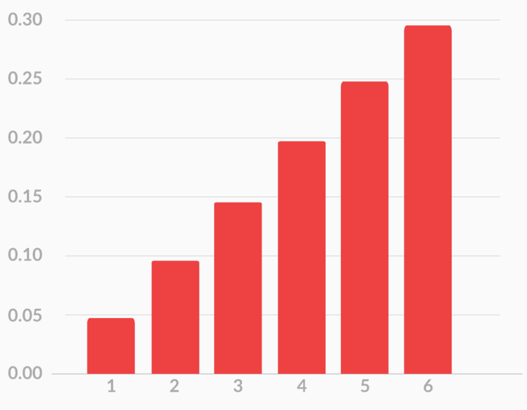
Probability Distributions - II:-

So essentially, a **probability distribution** is a form of representation that tells us the probability for all the possible values of X. It could be any of the following:

- A table

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

- A chart



- An equation

P(x) = x/21
(for x = 1, 2, 3, 4, 5 and 6)

Hence, in a valid, complete probability distribution, there are no negative values, and all the probability values add up to 1. These two conclusions follow from the basic definition of probability.

Also, recall that we discussed that the probability distribution and frequency distribution would be similar in shape, just with different scales. You can try it out in this interactive app. The graph on the left shows the frequency distribution, and the one on the right shows the probability distribution.

Now, let's say that a company's management is contemplating investing in a certain project. Before doing this, it wants to use probability to find whether it can safely expect to make a profit. Whether the company makes a profit or not will actually depend on which economic cycle is going on, i.e., recession, boom, and so on.

Based on the opinions of some experts, the following table is created:

Economic Cycle	Probability
Recession	0.1
Normal	0.7
Boom	0.2

Suppose as an analyst in the investment division, you have been asked to find the answer to the question: "Can the company expect to make a profit or not? Should it invest in this project?"

However, in this form, the table is of no help at all. Hence, let's quantify it using a random variable. Since you are interested in whether the company will profit or not, let's define X as the net revenue of the project.

Now, through some calculations, a fellow analyst of the company has arrived at the net revenue for each of these scenarios. She creates a probability distribution with this data:

X (Net Revenue of Project, in ₹ crore)	P(x)
-305	0.1
+15	0.7
+95	0.2

Now, you finally have a probability distribution for X, the net revenue of the project. Using this probability distribution, you can find the answer to our original question: "Can the company expect a profit from this project? Or, should it expect a loss?". However, to answer this, you will have to learn the concept of expected value, which is what we will cover in the next segment.

Expected Value - I:-

Again, let's go back to the three-step process that we followed to find whether the upGrad red ball game was profitable for the players or for the house:

1. Find all the possible combinations.
2. Find the probability of each combination.
3. Use the probabilities to estimate the profit/loss per player.

Now that we have completed steps 1 and 2, let's move on to step 3, where we will use the probabilities that we calculated to estimate the profit/loss per player.

So, the **expected value** for a variable X is the value of X that we would "expect" to get after performing the experiment an infinite number of times. It is also called the **expectation**, **average** or **mean value**. Mathematically speaking, for a random variable X that can take the values

$$x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n,$$

, the expected value (EV) is given by:

$$EV(X) = x_1 * P(X = x_1) + x_2 * P(X = x_2) + x_3 * P(X = x_3) + \ldots\ldots\ldots\ldots + x_n * P(X = x_n)$$

As you may recall, for our red ball game, the expected value came out to be **2.385**. What does this mean? How does this help us with our original question, which was: How much money on average are the players expected to make?

Let's explore this in the following video:

The **expected value** should be interpreted as the **average value** you get after the experiment has been conducted an **infinite number of times**. For example, the expected value for the number of red balls is 2.385. This means that if we conduct the experiment (play the game) infinite times, the average number of red balls per game would end up being 2.385. You can try it out in this interactive app.

As you can clearly see, after a large number of simulations, the average value does, in fact, converge towards the expected value, which is 2.385. In the next segment, you will learn more about expected value.

Expected Value - II:-

Let's try calculating the expected values for a different activity. Suppose you are throwing a die. You've defined X as the number obtained upon throwing it once. By calculations, you find that the expected value for this is 3.5. Let's see what our simulations show:

Recall the problem you saw earlier, where we were asked by a company to suggest whether it should invest in a given project or not. We had made this probability distribution for X, the net revenue of the project:

X (Net Revenue of Project, in ₹ Crore)	P(x)
-305	0.1
+15	0.7
+95	0.2

Now we are in a position to find the expected value for X, the return of the project. This is called the **expected return.** If it comes out to be negative, we can say that the project is not worth investing in.

The expected value of X, which is also called the expected return, is equal to:

(-305) * P(X = -305) + (+15) * P(X = +15) + (+95) * P(X = +95) = (-305) * 0.1 + (+15) * 0.7 + (+95) * 0.2 = -₹1 crore.

So, the expected return of the project is -₹1 crore. Hence, we can conclude that the project is not worth investing in.

You can find more examples of expected returns here. In the next segment, you will summarize what you have learned so far in this session.

Summary: Basics of Probability:-

In the first section, you learnt how to **quantify the outcomes** of events using **random variables**.

For example, recall that we quantified the balls of a particular colour that we would get after playing our game by assigning a value of X to each outcome. We did so by defining **X as the number of red balls** we would get after playing the game once.

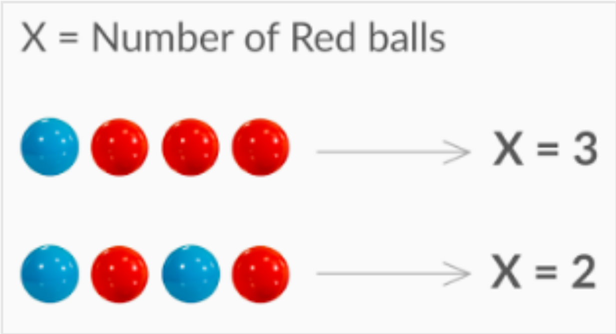


Figure 4 - Quantifying Using Random Variables

Next, we found the **probability distribution**, which was a **distribution giving us the probability for all the possible values of X**.

We created this distribution in a **tabular form:**

x	P(x)
0	0.027
1	0.160
2	0.347
3	0.333
4	0.133

Figure 5 - Tabular Form of Probability Distribution

We also created it in a **bar chart form:**

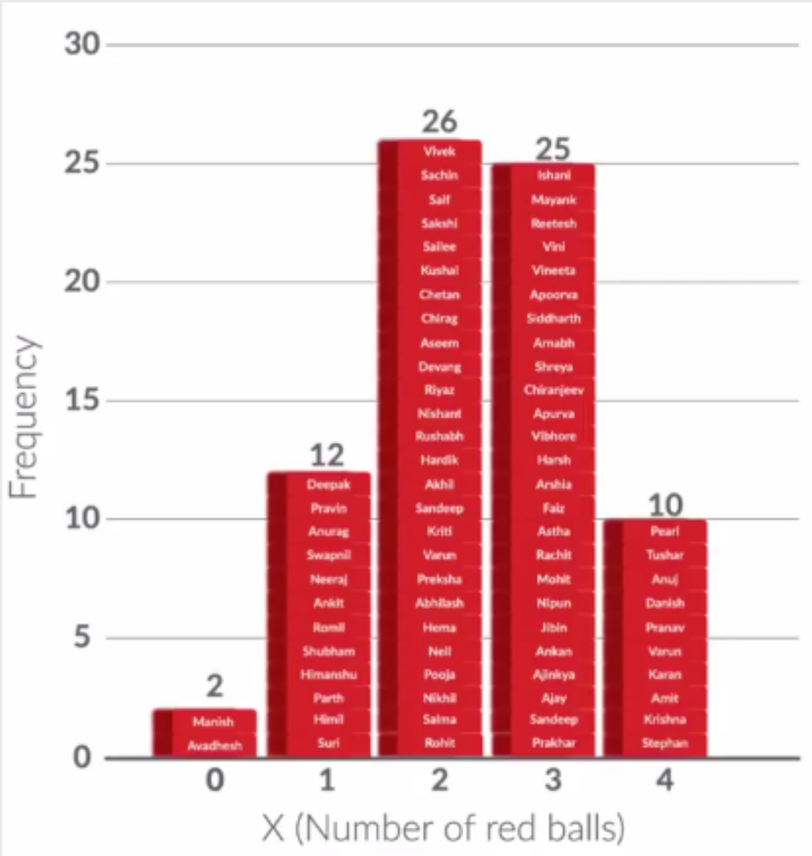


Figure 6 - Bar Chart Form of Probability Distribution

You saw that in the bar chart form, we were able to visualise the probability in a much better way. Thus, this form is used more widely as it helps you see trends easily.

Then, we went on to find the **expected value** for X, the money won by a player after playing the game once. The expected value (EV) for X was calculated using the following formula:

$$EV(X) = x_1 * P(X = x_1) + x_2 * P(X = x_2) + \ldots\ldots\ldots\ldots + x_n * P(X = x_n)$$

Another way of writing this is as follows:

$$EV(X) = \sum_{i=1}^{i=n} x_i * P(X = x_i)$$

Calculating the answer this way, we found the expected value to be +11.28.

In other words, if we conduct the experiment (play the game) **infinite times**, the **average money** won by a player would be ₹11.28. Hence, we decided that we should either decrease the prize money or increase the penalty to make the expected value of X negative. A negative expected value would imply that on average, a player is expected to lose money and the house profits.

Module 3 : Discrete Probability Distributions

Introduction: Discrete Probability Distributions:-

Welcome to the session on '**Discrete Probability Distributions**'. In the last session, you learnt some basic concepts of probability such as **random variables**, **probability distributions**, and **expected value**. Let's now learn some slightly more advanced concepts.

In this session

You will learn about some probability distributions that are commonly used for discrete random variables, such as **binomial probability distribution** and **uniform probability distribution**. In addition, you will learn the concept of **cumulative probability**, which will be useful in our next session on continuous probability distributions.

Probability Without Experiment - I:-

In the last session, we found the probability for certain events by conducting experiments.

Specifically, we asked 75 people to play the upGrad red ball game. Based on the data gathered from these people, we created a histogram or, in other words, a frequency distribution. Then, using this histogram, we created the probability distribution.

However, this is a lengthy process. Is there a shorter process for finding the probabilities, perhaps one that doesn't need repeated experiments? Let's see.

When we have 2-Blue Balls and 3-Red Balls:-
p(B) =2/5=0.4
p(R) =3/5=0.6

p(1B3R) = 0.4 *(0.6)*(0.6)*(0.6)=0.0846
Now this 1B and 3R balls can occurs in 4 different ways so total probability are as = 4 * p(1B3R) = 4*0.0846 is the answer.

Thus, we have gone through the exercise of finding the **probability without conducting any experiment**. You saw that these theoretical (calculated) values of probability are actually quite close to the experimental values. The small differences exist because of the low number of experiments conducted.

Probability Without Experiment - II:-

Recall that we claimed that if we had conducted the upGrad experiment several times, then the resulting experimental probability distribution would have been even closer to the theoretical one.

As you can see, after playing the game several times (~500 times), the experimental probability distribution starts to look similar to the theoretical one.

Similarly, you can simulate a coin flip and compare the theoretical distribution with the experimental one.

Now that you know how to find the probability without an experiment, you can calculate the probability for various combinations without much effort. For example, what if the bag for our game had, say, 4 red balls and only 1 blue ball? You don’t need to perform an experiment 100 or 500 times to find the answer. You can find it using a small calculator, the concepts of probability and, of course, your brain!

In the next segment, you will get to know about the binomial distribution.

Binomial Distribution:-

Earlier, we found the theoretical probability for our game and compared it with the experimental one. Finding the probability without conducting an experiment means that we can find the probability using just pen and paper and with minimal effort.

Now, let’s try to generalise it — let’s say that the probability of getting 1 red ball in one trial is equal to p. In that case, what would be the probability of all 4 balls being red? Let’s see in the following video.

So, the probability distribution for X (i.e., the number of red balls drawn after 4 trials) if the **probability of getting a red ball in 1 trial is 'p' is as follows**:

x	P(X=x)
0	$(1-p)^4$
1	$4p(1-p)^3$
2	$6p^2(1-p)^2$
3	$4p^3(1-p)$
4	p^4

Figure 1 - Probability Distribution for General Probability p

In the following video, we will see how this can be generalised even further.

So, the formula for finding **binomial probability** is given by:

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

Where **n** is **the number of trials**, **p** is the **probability of success**, and **r** is the **number of successes after n trials**.

However, there are some **conditions** that need to be met in order for us to be able to apply the formula.

1. The **total number** of trials is **fixed** at **n**.
2. Each trial is **binary**, i.e., it has **only two possible outcomes:** success or failure.
3. **Probability of success** is the **same** in all trials, denoted by **p**.

When to use binomial :-

CONDITIONS FOR BINOMIAL PROBABILITY DISTRIBUTION

1. Total number of trials is fixed at n
2. Each trial is binary, i.e. has only two possible outcomes - success or failure
3. Probability of success is same in all trials, denoted by p
4. P(X=r) = P(Getting r successes in n trials) = nCr(p)r (1-p)n-r

In the next segment, you will see some examples of binomial distributions.

Binomial Distribution (Examples):-

In the previous section, we listed down some conditions that are to be met for the binomial distribution to be applicable. Let’s take a few examples to understand these conditions in detail.

Binomial Distribution Applicable	**Binomial Distribution Not Applicable**
Tossing a coin 20 times to see how many tails occur	Tossing a coin until a head occurs
Asking 200 randomly selected people if they are older than 21 or not	Asking 200 randomly selected people how old they are
Drawing 4 red balls from a bag, putting each ball back after drawing it	Drawing 4 red balls from a bag, not putting each ball back after drawing it

If you toss a coin 20 times to see how many times you get tails, you are following all the conditions required for a binomial distribution. The total number of trials is fixed (20), and you can only have two outcomes, i.e., tails or heads. The probability of getting a tail is 0.5 each time you toss a coin.

In a way, this is similar to drawing 20 balls out of a bag, replacing each ball after drawing it, and seeing how many of the balls are red. Here, the probability of getting a red ball in one trial is 0.5.

When you toss a coin until you get heads, the total number of trials is not fixed. This is similar to taking out balls from the bag repeatedly until you draw a red ball. You can still find the probability of getting heads in 1 trial, 2 trials, 3 trials etc. and so on, but you cannot use binomial distribution to find that probability.

In the second example, where binomial distribution is not applicable, the experiment does not have only two outcomes, but several. It is similar to taking out balls from a bag that contains red, blue, black, orange, and other-coloured balls. The probability distribution for this experiment cannot be made using binomial distribution.

In the final example, the probability of trials is not equal to each other. For example, the probability of drawing a red ball in the first trial is 3/5. Now, in the second trial, the probability of drawing a red ball would be equal to 2/4 not 3/5, as the red ball taken out in the first trial was not put back. Hence, the probability of getting the combination red-red-red-blue, for example, would be 3/5*2/4*1/3*2/2, which is not the value we got while deriving binomial distribution (3/5*3/5*3/5*2/5). Again, you cannot use binomial distribution to find the probability in this case.

In other words, binomial distribution is applicable in situations where **there are a fixed number of yes or no questions, with the probability of a yes or a no remaining the same for all questions.**

So, you now understand that **binomial distribution** is a very powerful distribution. To get an idea of what this probability distribution looks like, you can use the interactive app provided below. This app shows you the probability distribution for a binomial distribution with n = 5 and p = 0.5. However, you can **play around with** the values of **n and p** to see how that changes the probability distribution. Don't forget to zoom out or zoom in, as needed.

Cumulative Probability:-

In the previous example, we only discussed the probability of getting an exact value. For example, we know the probability of X = 4 (4 red balls). But what if the house wants to know the probability of getting $\leq$ 3 red balls, as the house knows that for $\leq$ 3 red balls, the players will lose and the house will make money?

Sometimes, talking in terms of **less than** is more useful. For example — how many employees can get to work in less than 40 minutes? Let's explore how you can find the probability for such cases.

Clarification: The cumulative probability distribution table, shown in the above video (02:22 to 03:11) should be a follows:

x	**F(x) = P(X≤x)**
0	0.0256
1	0.1792
2	0.5248
3	0.8704
4	1.0000

(**Side Note**: In the question for calculating the probability of weights between 60 and 65, you might have noticed that we calculated P(X ≤ 65)-P(X ≤ 60) to find P(60 ≤ X ≤ 65).
However, this is not entirely correct if you consider discrete distributions, because P(X=60) would also get subtracted from the value P(X<=65) when we evaluate P(X ≤ 65)-P(X ≤ 60)
The assumption we made, that both P(X=60) and P(X=65) would be equal to 0, is necessary for the calculation to hold true. However, an important concept that you will learn further
is that the "weight" variable is generally considered to be continuous and not discrete, and for continuous variables, the value of P(X=x), where X is a random variable and x is a value, is always 0. Hence, the logic provided in the video holds. So, for cumulative distribution tables where the probabilities of individual

values are not given, you can just use a similar analogy to calculate the probability between the two values.)

So, the **cumulative probability of X**, denoted by **F(x)**, is defined as **the probability of the variable being less than or equal to x**.

In mathematical terms, you would write cumulative probability **$F(x) = P(X \le x)$**. For example, $F(4) = P(X \le 4)$, $F(3) = P(X \le 3)$.

Summary: Discrete Probability Distributions:-

We started with learning how to find **probability without experiment**, using basic concepts such as the addition and the multiplication rule of probability.

As a demonstration, we calculated the probability of getting 0, 1, 2, 3 and 4 red balls for our upGrad red ball game. Then, we compared the values we got through theoretical calculations with the ones we got after the experiments in the first session. Recall that the values were quite similar.

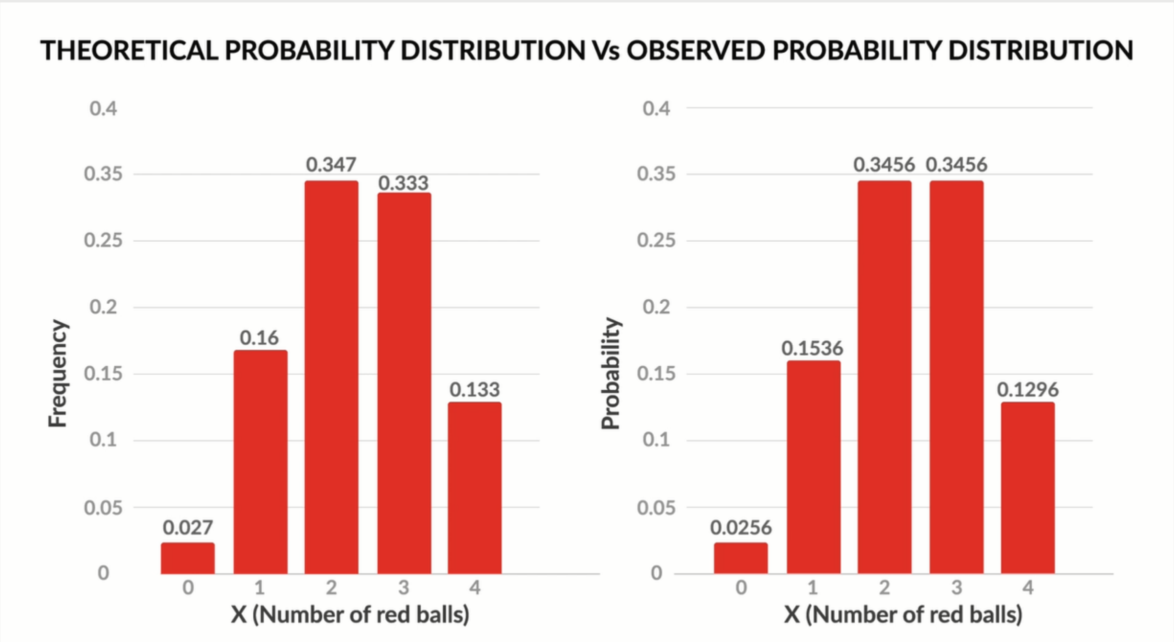


Figure 2 - Observed vs Theoretical Probability Distribution

However, they were not exactly the same due to the low number of experiments conducted (75). If more experiments were performed, the values would have been exactly the same (the values are exactly the same if the number of experiments approaches infinity).

Next, we **generalised this probability**. Specifically, we talked about the probability of getting **r red balls** after drawing **n balls** from a bag. Here, the probability of drawing **a red ball in 1 trial** was equal to **p**.

The probability distribution for this case is given by the following table (X = number of red balls drawn after playing the game once).

BINOMIAL PROBABILITY DISTRIBUTION

x	P(X=x)
0	${}^nC_0(p)^0(1-p)^n$
1	${}^nC_1(p)^1(1-p)^{n-1}$
2	${}^nC_2(p)^2(1-p)^{n-2}$
3	${}^nC_3(p)^3(1-p)^{n-3}$
.	.
.	.
.	.
.	.
n	${}^nC_n(p)^n(1-p)^0$

In general,

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

This distribution is called the **binomial distribution**. It can be used to find the probability of any kind of event, if that event is **a series of yes or no questions, with the probability of yes being the same for all questions**.

Phrasing the conditions more formally, the binomial distribution can be used if, for an experiment:

- The **total number** of trials is **fixed.**
- Each trial is **binary**, i.e., it has **only two possible outcomes:** success or failure.
- The **probability of success** is the **same** for all the trials.

Next, we discussed the **cumulative probability of x**, denoted by **F(x)**, which is the **probability that the random variable X takes a value less than or equal to x**.

For example, we found F(2), the probability of getting 2 or fewer red balls in our upGrad game. It was calculated as:

$$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.0256 + 0.1536 + 0.3456 = 0.5248$$

The cumulative probability is a concept that we will use extensively in our next session on continuous random variables.

Comprehension: Expected Value:-

In the last session, you learnt about expected values. Now that you know how to calculate probability using the binomial distribution, it would be better to revisit the concept and practice some questions.

Calculating the expected value is a three-step process:

1. Define the random variable (X).
2. Calculate the probability distribution P(X). You'll need to calculate it on your own.
3. Plug the above two terms in the following formula:

$$E[X] = \sum(X \times P(X))$$

Let's understand the process with an **example**. Suppose you're playing a game involving a 6-sided die. Could you tell what is the average outcome that you'd expect each time the die is thrown? Answering this question requires us to calculate the expected value.

Let's solve this problem step by step:

1. The first step is defining the random variable. The random variable (X) is the outcome of a die throw. So, X = {1, 2, 3, 4, 5, 6}
2. The second step is to calculate the probabilities related to each outcome. The probability of each outcome is 1/6 in a die throw.

Now, you have X and P(X). If you plug these values in the formula

$$E[X] = \sum(X \times P(X))$$

,you'll get 3.5 as the expected value. So how to interpret this number? This means if you were to throw the die a large number of times, the average of those numbers will tend towards 3.5.

So, why do we need the expected value at all? Well, the expected value lets you reason about real-world random phenomenon more rationally. For instance, investing in stock markets is a popular use-case where this concept is used.

Suppose you're interested in investing in the stock market. It's always better to invest in multiple stocks rather than one stock. You can calculate the expected value of your returns using the concept of expected value. Let's take a simple hypothetical situation. Our random variable, in this case, can take the expected return of each stock. Then, to calculate the expected return, you need the probability of returns for each stock. This way, you can calculate the expected return of your entire portfolio, which will allow you to invest wisely.

Similarly, the expected value is often used in business settings where a stakeholder has multiple options.

Module 4 : Continuous Probability Distributions

Introduction: Continuous Probability Distributions:-

Welcome to the session on '**Continuous Probability Distributions**'. In the last session, you learnt about **binomial distribution** and **cumulative probability**.

In this session

You will learn about **cumulative probability** in detail. You will see how the **probability of a continuous variable** is expressed and how it is different from the way the probability of a discrete variable is expressed. You will then learn about **normal distribution**, which is a commonly used probability distribution among continuous random variables.

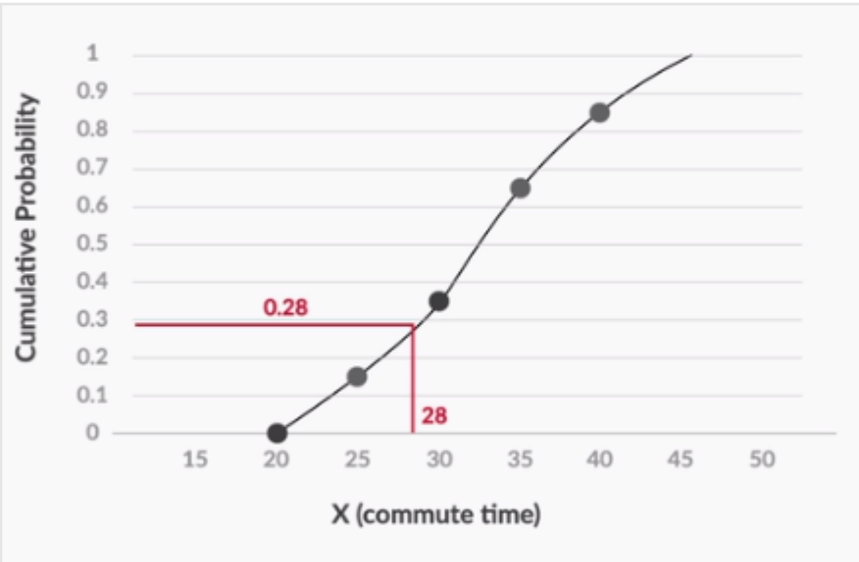
Probability Density Functions - I:-

In the last section, you saw how to find the probability of certain events using the multiplication and addition rules of probability. Also, for some specific cases, you saw that probability distributions like binomial distribution and uniform distribution can be used to calculate probability.

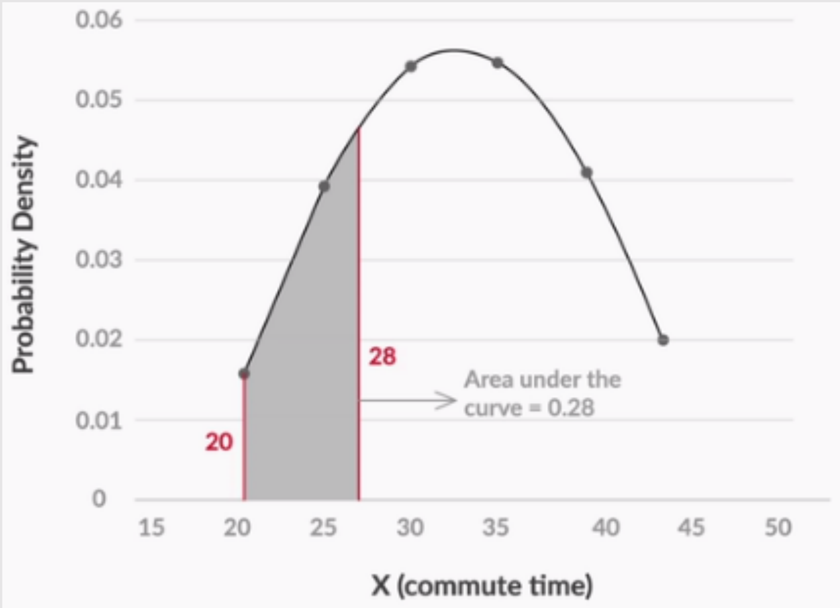
So far, we have only talked about discrete random variables, e.g., number of balls, number of patients, cars, wickets, pasta packets, etc. What happens when we talk about the **probability of continuous random variables**, such as time, weight etc.? Is there any difference? Let’s find out.

Now you know what **CDF** and **PDF** are. Since these two functions talk about probabilities **in terms of intervals** rather than the exact values, it is advisable to use them when talking about continuous random variables, not the bar chart distribution that we used for discrete variables.

Recall that a **CDF**, or a **cumulative distribution function**, is a distribution that plots the cumulative probability of X against X.

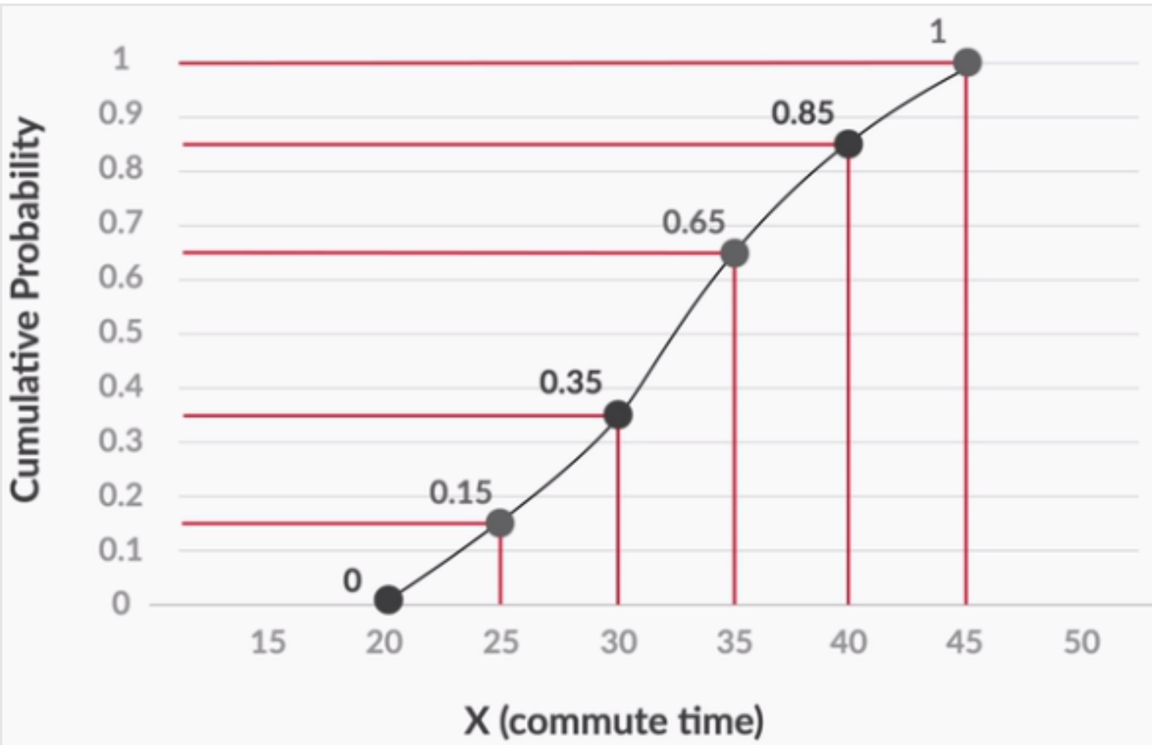


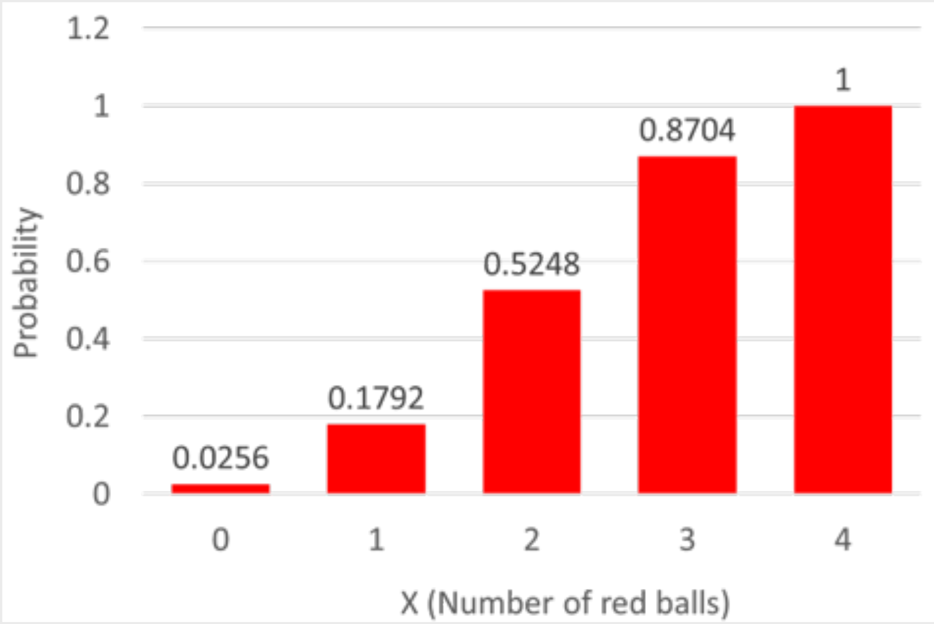
A **PDF**, or a **Probability Density Function**, however, is a function in which the area under the curve gives you the cumulative probability.



For example, the area under the curve between 20, the smallest possible value of X, and 28 gives the cumulative probability for X, which is equal to 28.

The main difference between the cumulative probability distribution of a continuous random variable and a discrete one lies in the way you plot them. While a continuous variables' cumulative distribution is a curve, a distribution for discrete variables looks more like a bar chart.





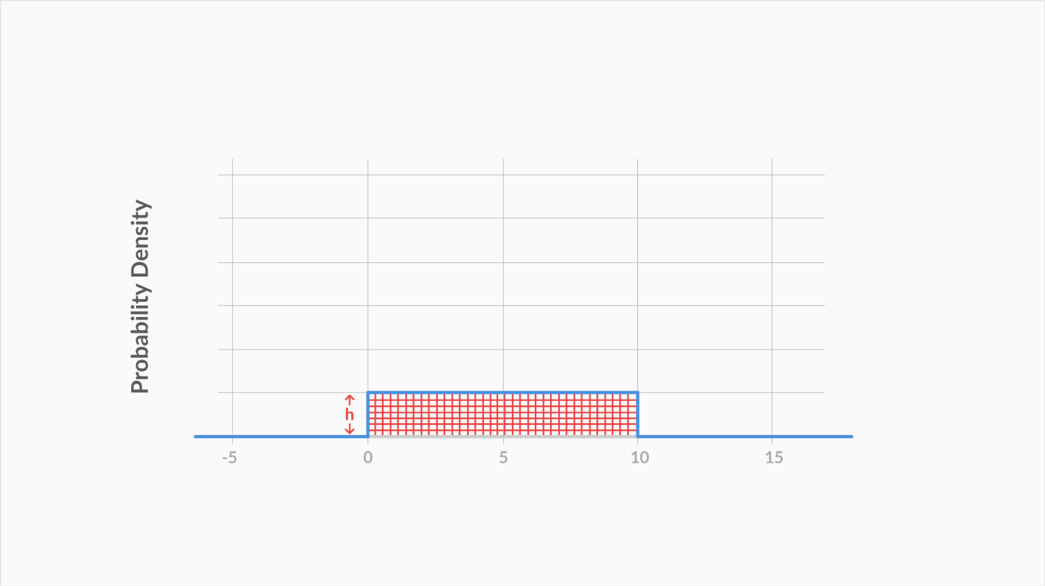
The reason for the difference is that for discrete variables, the cumulative probability does not change very frequently. In the discrete variable example, we only care about what the probability is for 0, 1, 2, 3 and 4. This is because the cumulative probability will not change between, say, 3 and 3.999999. For all values between these two, the cumulative probability is equal to 0.8704.

However, for the continuous variable, i.e., the daily commute time, you have a different cumulative probability value for every value of X. For example, the value of cumulative probability at 21 will be different from its value at 21.1, which will again be different from the one at 21.2, and so on. Hence, you would show its cumulative probability as a continuous curve, not a bar chart.

Probability Density Functions - II:-

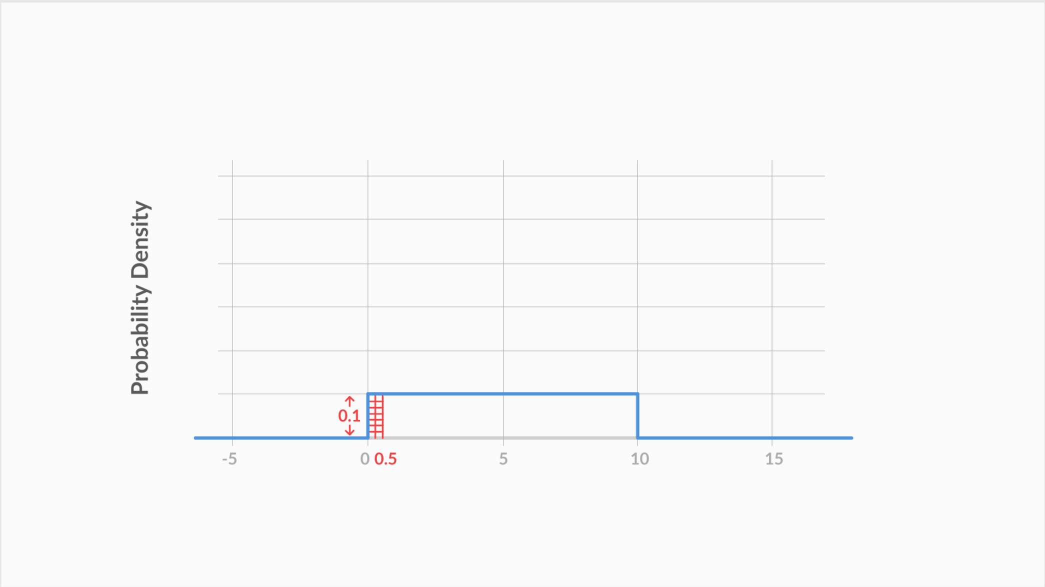
A commonly observed type of distribution among continuous variables is a **uniform distribution**. For a continuous random variable following a uniform distribution, the value of probability density is equal for all possible values. Let's explore this distribution a little more.

Since all possible values are between 0 and 10, the area under the curve between 0 and 10 is equal to 1.



Clearly, this area is the area of a rectangle with length 10 and unknown height h. Hence, you can say that 10 * h = 1, which gives us h = 0.1. So, the value of the PDF for all values between 0 and 10 is 0.1.

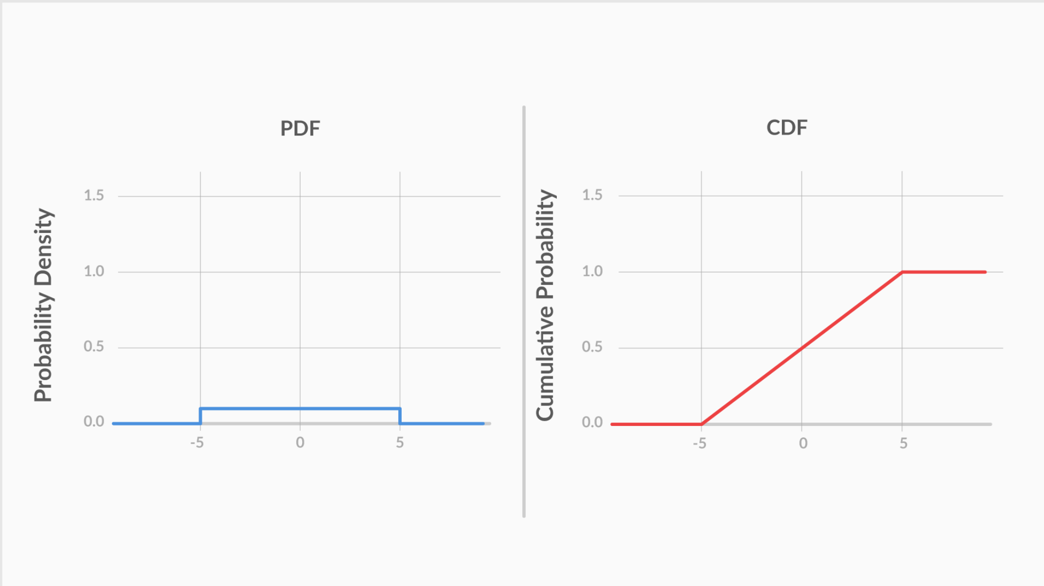
The cumulative probability for X = 0.5 is equal to the area under the curve between X = 0, the lowest possible value, and X = 0.5.



This area = 0.1 * 0.5 = 0.05.

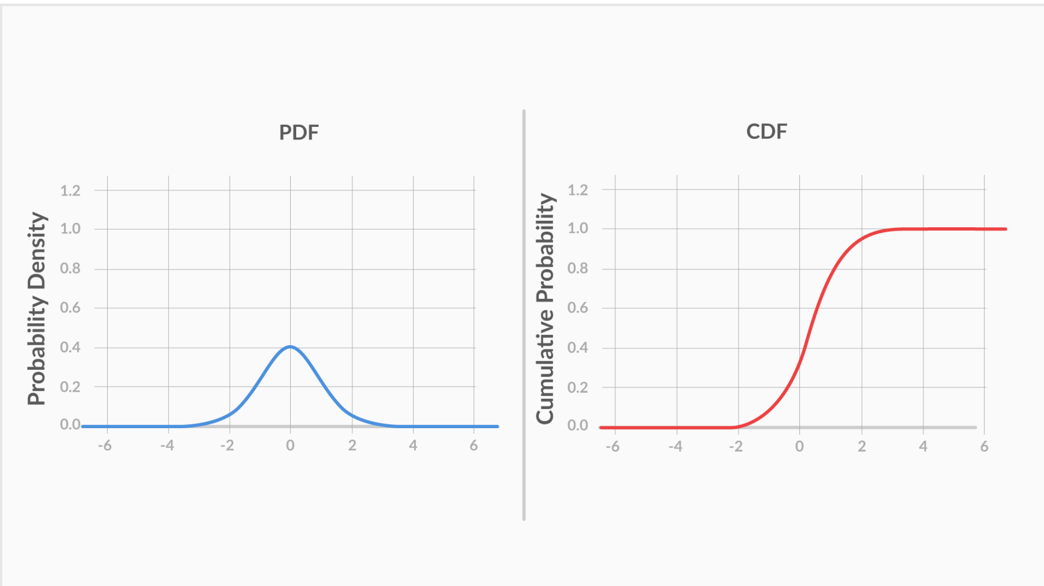
Now you must be wondering when to use PDFs and when to use CDFs. They are both good for continuous variables, but which one is used more in real-life analyses?

Well, PDFs are more commonly used in real life. The reason is that it is much **easier to see patterns in PDFs** as compared to CDFs. For example, here are the PDF and the CDF of a uniformly distributed continuous random variable:



The **PDF clearly shows uniformity**, as the probability density's value remains constant for all possible values. However, the **CDF does not show any trends** that help you identify quickly that the variable is uniformly distributed.

Now, let's look at the PDF and the CDF of a symmetrically distributed continuous random variable:



Again, it is clear that the symmetrical nature of the variable is much more apparent in the PDF than in the CDF.

Hence, generally, PDFs are used more commonly than CDFs.

In the next segment, you will learn more about the normal distribution.

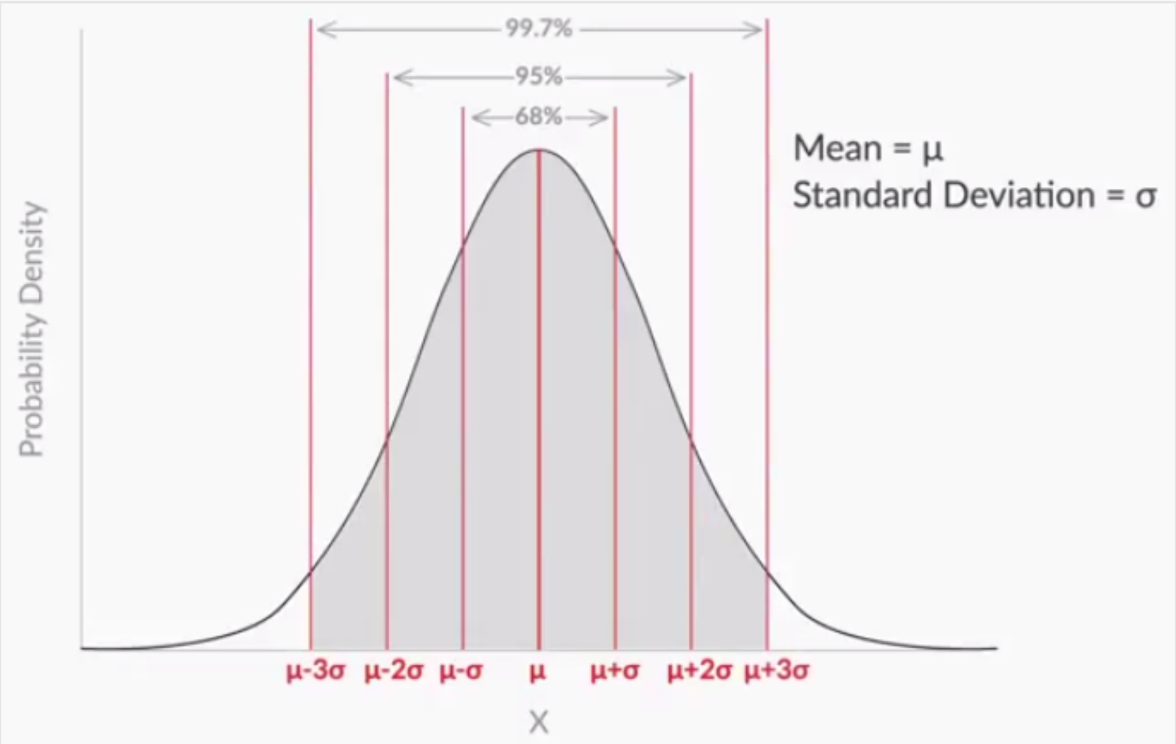
Normal Distribution:-

You saw how the probability distributions of continuous random variables differ from those of discrete random variables.

But can you think of some examples of continuous distributions? Which is the most commonly used continuous probability distribution? Which distribution occurs most commonly in nature? Let’s hear from Professor Tricha on this.

Normally distributed data follows the **1-2-3 rule**. This rule states that there is a:

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

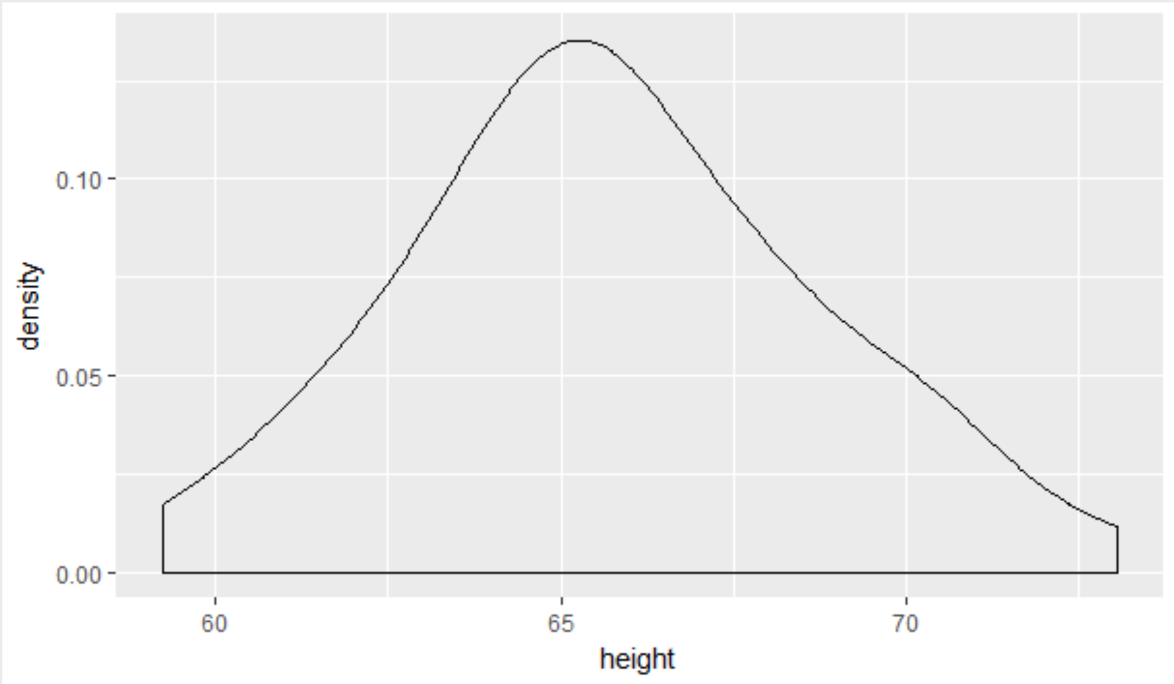


This is similar to this case: If you buy a loaf of bread every day and measure it, where the mean weight = 100 g and the standard deviation = 1 g, then:

1. For 5 days every week, the weight of the loaf that you bought that day will be within 99 g (100-1) and 101 g (100+1).
2. For 20 days every 3 weeks, the weight of the loaf that you bought that day will be within 98 g (100-2) and 102 g (100+2).
3. For 364 days every year, the weight of the loaf that you bought that

day will be within 97 g (100-3) and 103 g (100+3).

A lot of naturally occurring variables are normally distributed. For example, the heights of a group of adult men would be normally distributed. To try this out, we took the heights of 50 male employees at the upGrad office and then plotted the probability density function using that data.



As you can see, the data is roughly normal.

You can visualise the PDF and CDF for different normal distributions using the interactive app given below. From the various options in the drop-down menu, select '**Normal'**. You will then get to see the probability distribution for a normal distribution with μ = 0 and σ = 1. In fact, you can **play around with the value of μ and σ** to see how that changes the distribution. Using the **green slider** below the distribution, you can visualise the distribution's **cumulative probability**.

Standard Normal Distribution:-

As you learnt in the previous question, it doesn't matter what the values of μ and σ are. If you want to find the probability, all you need to know is how far the value of X is from μ, and specifically, what **multiple of σ** is the **difference between X and μ**.

Let's see how you can find this.

As you just learnt, the **standardised random variable** is an important parameter. It is given by:

$$Z = \frac{X - \mu}{\sigma}$$

Basically, it tells you **how many standard deviations away from the mean** your random variable is. As you just saw, you can find the cumulative probability corresponding to a given value of Z, using the **Z table**:

Number in the table represents P(Z≤z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Number in the table represents P(Z≤z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Alternatively, you can use the following equation to find the cumulative probability:

$$F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{\frac{-t^2}{2}} \, dt$$

You can also use **Excel** to find the cumulative probability for Z. For example, let's say you want to find the cumulative probability for Z = 1.5. In the Excel sheet, you will type:

= NORM.S.DIST(1.5, TRUE)

Basically, the syntax is:

= NORM.S.DIST(z, TRUE)

Here, z is the value of the Z score for which you want to find the cumulative probability. TRUE = find cumulative probability, FALSE = find probability density.

Also, you can find the probability without standardising. Let's say that X is normally distributed, with mean (μ) = 35 and standard deviation (σ) = 5. Now, if you want to find the cumulative probability for X = 30, you would type:

= NORM.DIST(30, 35, 5, TRUE)

Basically, the syntax is:

= NORM.DIST(x, mean, standard_dev, TRUE)

As you can see, the value of σ is an indicator of how wide the graph is. This will be true of any graph, not just a normal distribution. A **low** value of σ means that the graph is **narrow**, while a **high** value implies that the graph is **wider**. This is because a wider graph has more values away from the mean, resulting in a high standard deviation.

Again, some more probability distributions are commonly seen among continuous random variables. They are not covered in this course, but if you want to go through some of them, you can use the links below:

1. https://towardsdatascience.com/what-is-exponential-distribution-7bdd08590e2a

2. https://towardsdatascience.com/gamma-distribution-intuition-derivation-and-examples-55f407423840

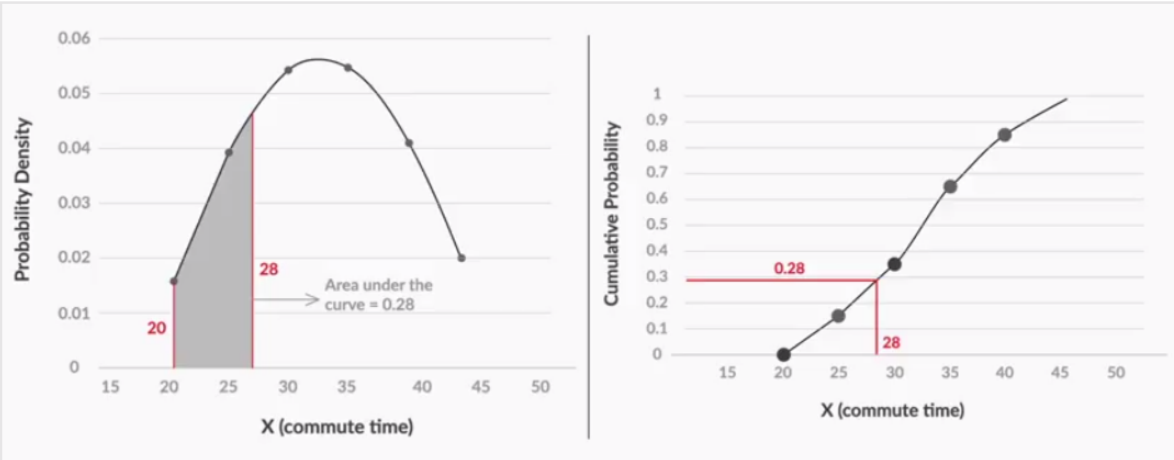
3. http://onlinestatbook.com/2/chi_square/distribution.html

Summary: Continuous Probability Distributions:-

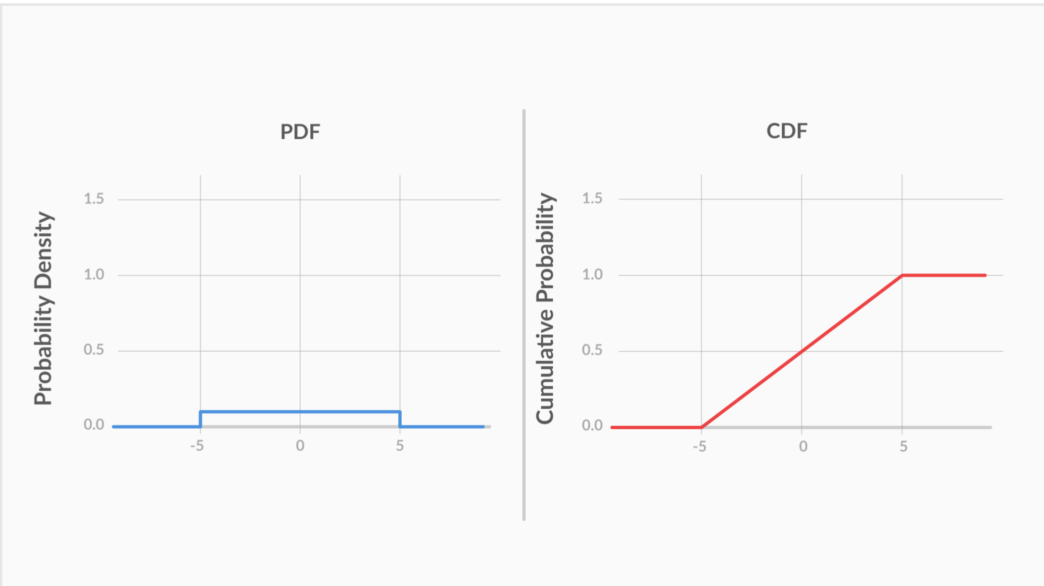
You started this session by learning that for a **continuous random variable**, the **probability of getting an exact value is** very low, almost **zero**. Hence, when talking about the probability of continuous random variables, you can only talk **in terms of intervals**. For example, for a particular company, the probability of an employee's commute time being exactly equal to 35 minutes was zero, but the probability of an employee having a commute time between 35 and 40 minutes was 0.2.

Hence, for continuous random variables, **probability density functions** (**PDFs**) and **cumulative distribution functions** (**CDFs**) are used instead of the bar chart type of distribution used for the probability of discrete random variables. These functions are preferred because they talk about probability in terms of intervals.

Then, you understood that the major difference between a PDF and a CDF is that in a CDF, you can find the cumulative probability directly by checking the value at x. However, for a PDF, you need to find the area under the curve between the lowest value and x to find the cumulative probability.



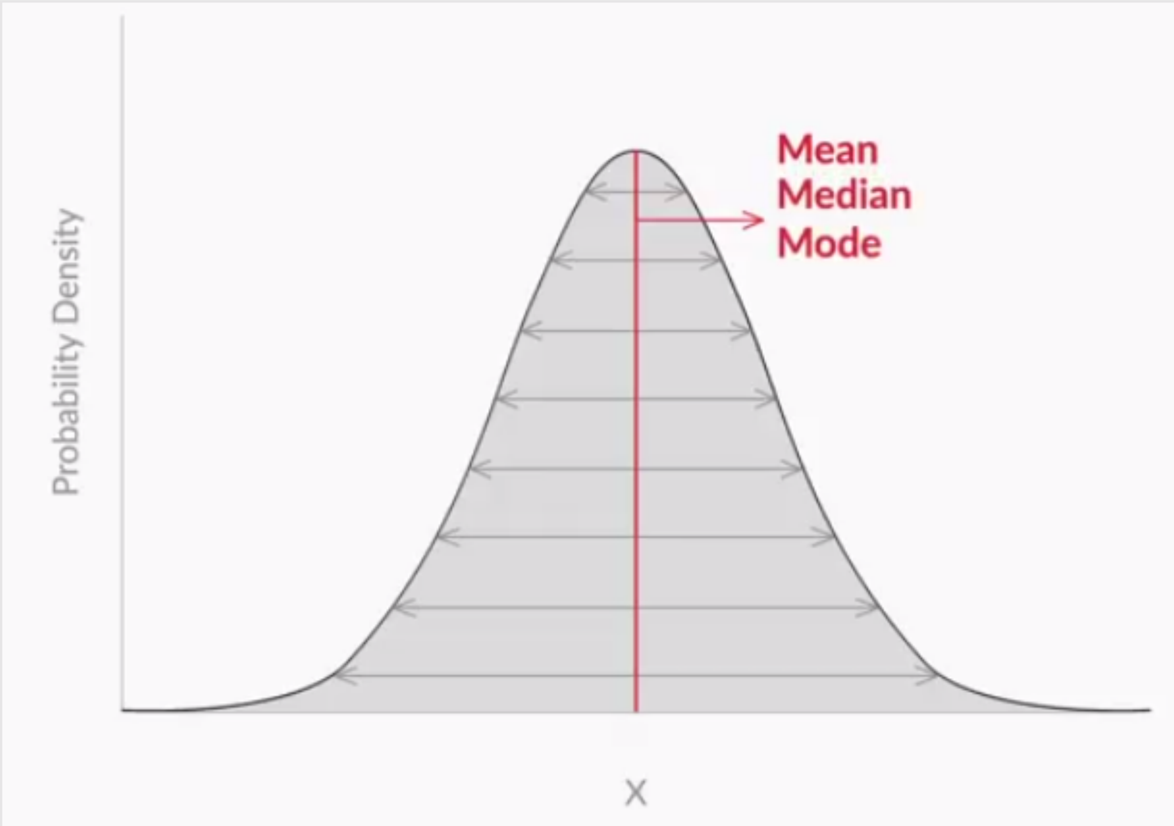
You also learnt that **PDFs are still more commonly used**, mainly because it is very **easy to see patterns** in them. For example, for a uniformly distributed variable, the PDF and CDF look like this:



While the PDF clearly shows that the variable is uniformly distributed, the CDF does not offer any such quick insights.

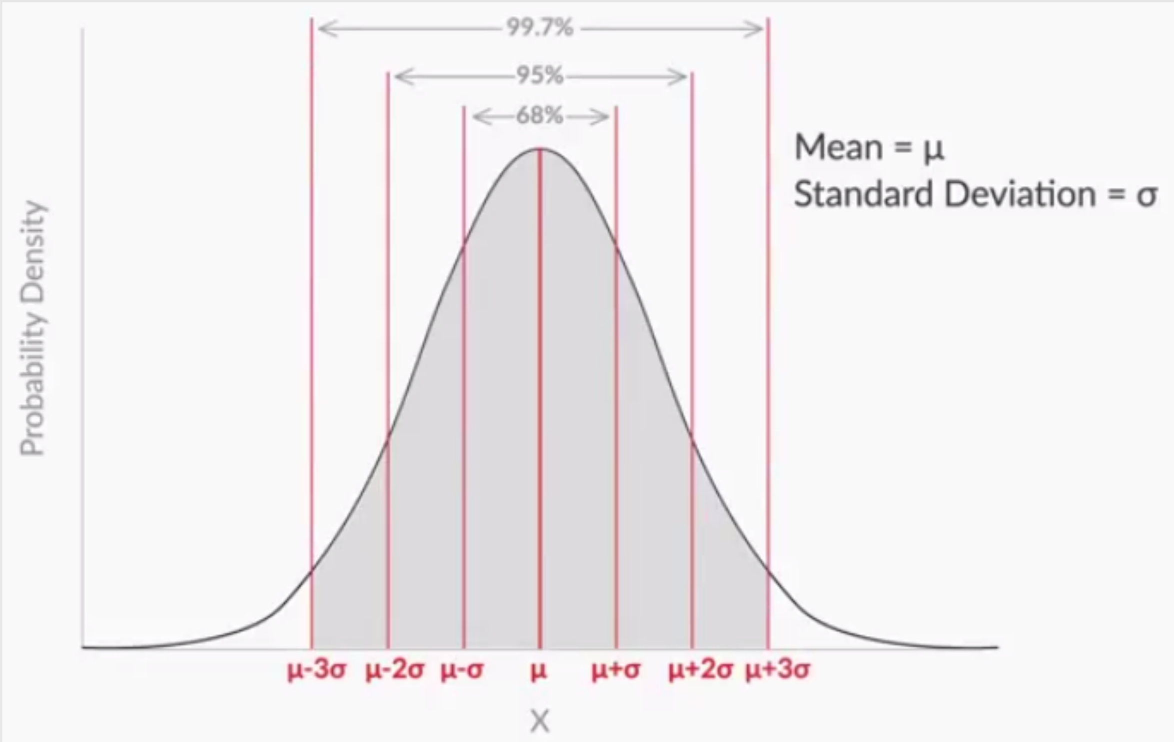
Next, you learnt about a very famous probability density function: the **normal distribution**. You saw that it is **symmetric**, and its **mean, median and**

mode lie at the **centre**.



You also learnt the **1-2-3 rule**, which states that there is a:

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



Then, you learnt that to find the probability, you do not need to know the value of the mean or the standard deviation; just knowing **the number of standard deviations away from the mean** your random variable is suffices. That is given by:

$$Z = \frac{X-\mu}{\sigma}$$

This is called the **Z score**, or the **standard normal variable**.

Finally, you learnt how to find the cumulative probability for various values of Z using the **Z table**. For example, you found the cumulative probability for Z = 0.68 using the Z table.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

The intersection of row '0.6' and column '0.08', i.e., 0.7517, is your answer.

Also, you learnt how to use Excel to find this probability. For example, the cumulative probability for Z =1.5 can be found using Excel by typing:

= NORM.S.DIST(1.5, TRUE)

Also, you can find the probability without standardising. The syntax for that is:

= NORM.DIST(x, mean, standard_dev, TRUE)

A normal distribution finds use in many statistical analyses. In the next session, you will learn about its use in the central limit theorem, which is, in turn, useful for understanding the next module on hypothesis testing.

Module 4 : Central Limit Theorem

Introduction: Central Limit Theorem:-

Welcome to the session on 'Central Limit Theorem'. In the last session, you learnt about probability density functions, specifically normal and standard normal distributions.

In this session

You will learn what a sample is and why it is so error-prone. You will then understand how to quantify this error made in sampling using a popular theorem in statistics, called the central limit theorem.

Prerequisites

There are no prerequisites for this session, other than, of course, your knowledge of what was discussed in the previous three sessions.

Samples:-

So far, you have conducted analyses for data on 75 people, 3,000 people, and so on. But what if you need to analyse a very large amount of data, e.g., data on 3,00,000 people? Or, what if you need to do this for, say, the entire Indian population?

Suppose for a business application, you want to find out the average number of times people in urban India visited malls last year. That's 400 million (40 crore) people! You cannot possibly go and ask every single person how many times they visited the mall. That's a costly and time-consuming process. How can you reduce the time and money spent on finding this number?

To reiterate, these are the notations and formulas related to populations and their samples:

Population/Sample	Term	Notation	Formula
Population ($X_1, X_2, X_3, \ldots\ldots, X_N$)	Population Size	N	Number of items/elements in the population
	Population Mean	μ	$\frac{\sum_{i=1}^{i=N} X_i}{N}$
	Population Variance	σ^2	$\frac{\sum_{i=1}^{i=N}(X_i - \mu)^2}{N}$
Sample ($X_1, X_2, X_3, \ldots\ldots, X_n$) (Sample of Population)	Sample Size	n	Number of items/elements in the sample
	Sample Mean	$\bar{X}$	$\frac{\sum_{i=1}^{i=n} X_i}{n}$
	Sample Variance	S^2	$\frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n-1}$

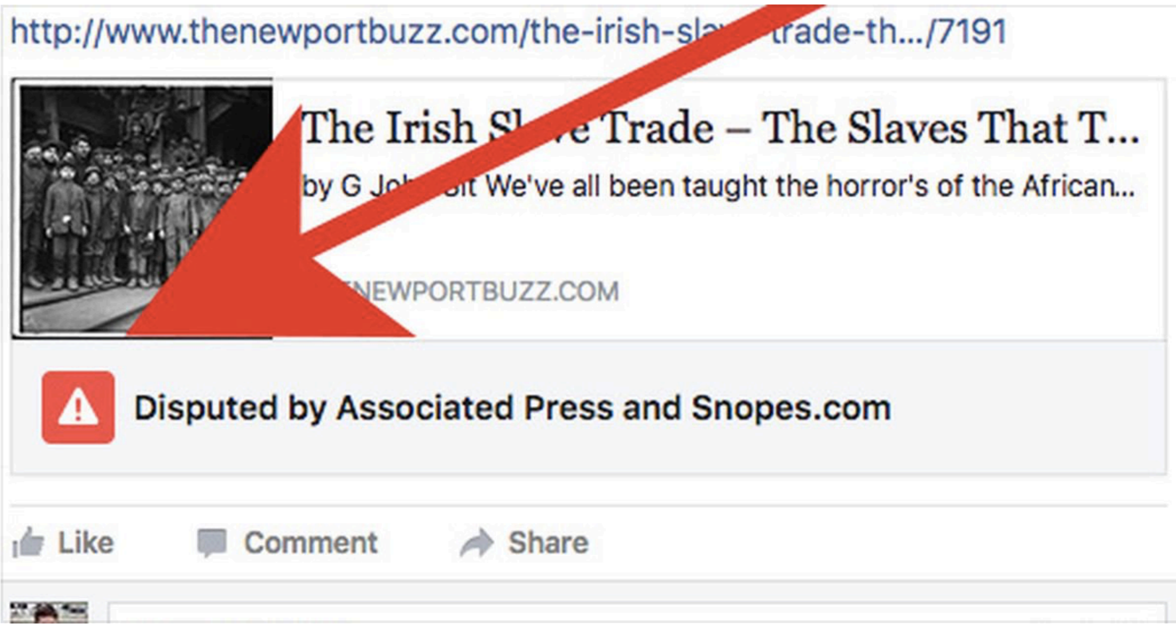
The reason for dividing by n-1 and not n is beyond the scope of this course. However, if you want to learn more about it, please refer to this link.

For an upcoming government project, you want to find the average height of the students in Class VIII of a given school. Instead of asking each student, suppose you took a few students as your sample and wrote the data down:

Roll Number	Height
8012	121.92 cm
8045	133.21 cm
8053	141.34 cm
8099	126.23 cm
8125	175.74 cm

Let's go through another example of sampling.

In order to counter fake news, let's say that Facebook is planning to include a new feature in its timeline. Below each post, a fact-checking warning will be provided, like this:



In case you want to read more about this feature, please refer to this link.

Before changing the timelines of all Facebook users to include this feature, Facebook first wants to evaluate how its users would react to this new feature.

So, it lets a small sample (~10,000 users) try out the new timeline. Then, it asks the 10,000 users whether they prefer the new timeline (Feature B) or the old timeline (Feature A).

Let’s say that one such survey shows that 50.5% of the people prefer feature B to feature A. Based on this, Facebook can say that feature B is preferred by more people than feature A, and hence, B should replace A.

By conducting the exercise on a sample and not the entire population, it **saved time and money** and **avoided risks** that could arise if it rolled out an untested feature.

But hold on! How can you be sure that the insights inferred for the sample hold true for the population as well? In other words, just because 50.5% of the people in the sample preferred feature B, is it fair to infer that 50.5% of the people in the population (1.86 billion Facebook users) will also prefer feature B to A?

You cannot answer this question with the information that you have right now. However, after the next few lectures, which will cover sampling distributions, central limit theorem and confidence intervals, you will be equipped with the knowledge required to answer this question.
In the next segment, you will learn more about sampling distributions.

Sampling Distributions:-

Now, let's move on to sampling distributions, whose properties, as we said earlier, will help you estimate the population mean from the sample mean.

So, the sampling distribution, specifically the sampling distribution of the sample means, is a probability density function for the sample means of a population.

This distribution has some very interesting properties, which will later help you estimate the sampling error. Let's take a look at these properties.

The **sampling distribution’s mean is denoted by** μ/¯X **,** as it is the mean of the sampling means. Let’s see what it is for this sampling distribution.

Properties of Sampling Distributions:-

We’ve been saying that the sampling distribution has some interesting properties that will later help you estimate the error in your samples. Let’s finally see what these properties are.

Again, to recap, let’s see what the notations and formulas are for populations, samples and sampling distributions

Population/Sample	Term	Notation	Formula
Population	Population Size	N	Number of items/elements in the population
$(X_1, X_2, X_3, \ldots\ldots, X_N)$	Population Mean	μ	$\frac{\sum_{i=1}^{i=N} X_i}{N}$
	Population Variance	σ^2	$\frac{\sum_{i=1}^{i=N}(X_i - \mu)^2}{N}$
Sample	Sample Size	n	Number of items/elements in the sample
$(X_1, X_2, X_3, \ldots\ldots, X_n)$	Sample Mean	$\bar{X}$	$\frac{\sum_{i=1}^{i=n} X_i}{n}$
(Sample of Population)	Sample Variance	S^2	$\frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n-1}$
Sampling Distribution of the Sample Mean	Sampling Distribution's Size	No convention (We have used k, but that is not a norm)	
$(\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots\ldots, \bar{X}_k)$	Sampling Distribution's Mean (mean of sample means)	$\mu_{\bar{X}}$	$\mu_{\bar{X}} = \mu$
(k Sample Means)	Sampling Distribution's Standard Deviation	S.E. (Standard Error)	S.E. = $\sigma/\sqrt{n}$

So, there are two important properties of a sampling distribution of the mean:

1. **Sampling distribution's mean** (μ^-X) = **Population mean** (μ)
2. Sampling distribution's standard deviation (**Standard error**) = σ/√n , where σ is the population's standard deviation and n is the sample size

Recall that in the last video, we created a sampling distribution and found the values of its mean and standard deviation as 2.348 and 0.4248, respectively. These values were very close to the values suggested by the formula, i.e., 2.385 and 0.44.

In the next segment, you will learn more about the central limit theorem.

Central Limit Theorem:-

You understood the third property of sampling distributions, which talks about their shape. Basically, it says that for n > 30, the sampling distributions become normally distributed. Let's recall all the three properties that you have learnt so far for sampling distributions.

So, the central limit theorem says that for any kind of data, provided a high number of samples has been taken, the following properties hold true:

1. **Sampling distribution's mean** ($\mu_{\bar{X}}$) = **Population mean** (μ),
2. Sampling distribution's standard deviation (**standard error**) = $\frac{\sigma}{\sqrt{n}}$, and
3. **For n > 30**, the sampling distribution becomes a **normal distribution**.

We made two sampling distributions (the upGrad game and the banking data set) and saw that they follow the aforementioned three properties.

Now, let's listen to Professor Tricha as she verifies the central limit theorem for some more population distributions.

Professor Tricha verified the CLT by performing simulations on different kinds of data. In case you want to try out these simulations yourself, you can go to this link. Press the "Begin" button in the top-right corner to get started.

In the next two segments, you will take a look at a detailed demonstration in python through which you'll be verifying the properties of Central Limit Theorem.

CLT Demonstration: I:-

In the previous lectures, you were formally introduced to what sampling distributions are and how they possess certain interesting properties which will help us in making inferences about the population parameters from a given sample. In this segment, you'll verify those properties using a Python demonstration to solidify your understanding of those concepts. Please download the necessary datasets and Jupyter notebook from the link given below:

[**Note**: This code walkthrough is for demonstration purposes only and therefore you will not be evaluated on this. The main aim of this demo is to get you to understand the difference between sample, population, sample means, the sampling distribution and their underlying properties.]

For this demonstration, you'll be using a dataset containing information about NBA players.
First, you'll be seeing how the population mean and the population standard deviation are related to those of the sampling distribution generated.

Note : *At "1:22" and "4:31" the instructor uses the code **df.Weight.std()** but it should be **df.Weight.std(ddof= 0)** as default value for ddof in pandas is 1 and while calculating population standard deviation we divide the numerator with n and not n-1.*

As you saw in the video above, the sampling distribution possesses 2 interesting properties that are related to the population parameters. Specifically, you verified that

- **Sampling distribution's mean** ($\mu_{\bar{X}}$) = **Population mean** (μ),
- Sampling distribution's standard deviation (**standard error**) = $\frac{\sigma}{\sqrt{n}}$

For the above example, we computed the mean weight of all the basketball players that were available in the dataset. This value came out to be **220.67**. In the context of this experiment, this value is our **population mean**.

Now to verify the first property, you picked around 1000 random samples of size 30 from the entire dataset and then calculated the mean of each sample. You plotted the distribution of all these sample means. This is your **sampling distribution**.

When you computed the mean of this sampling distribution (or in other words, the mean of all the sample means that you had taken earlier) you observed that this value came out to be **220.69**. As you can see this value is pretty close to the original population mean of **220.67**.

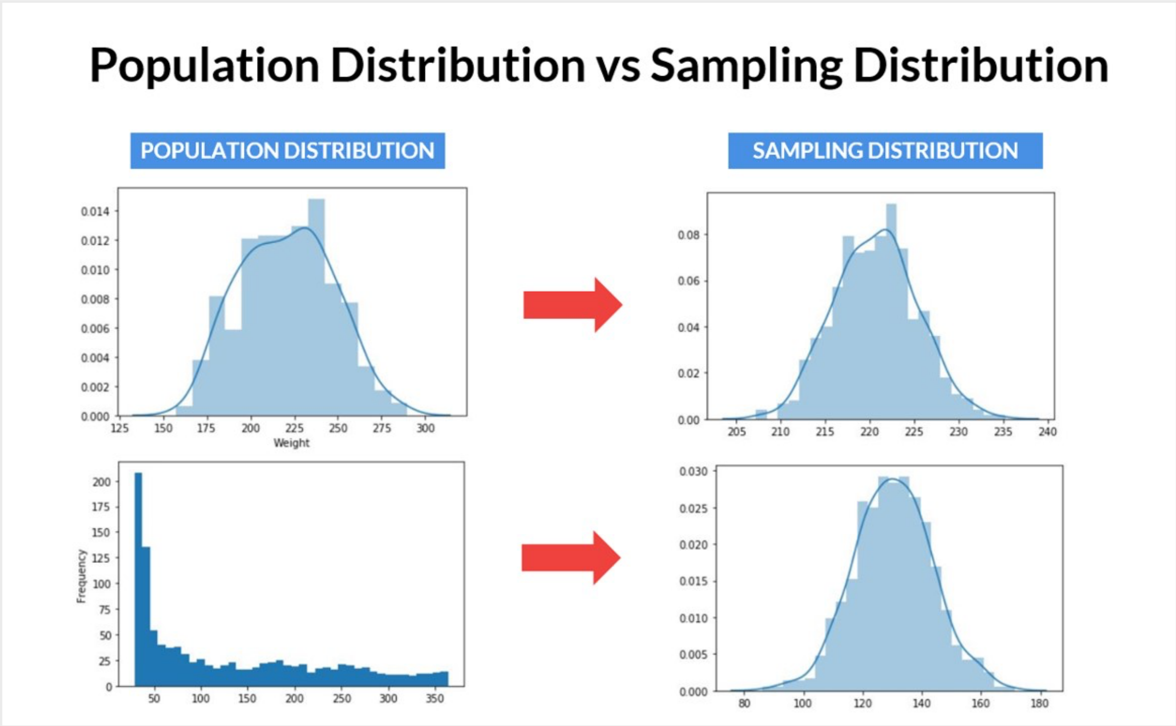
Similarly, when you computed the standard deviation of the sampling distribution, you observed the following relationship

$$Sampling\ Distribution\ Standard\ Deviation = \frac{Population\ Standard\ Deviation}{\sqrt{Sample\ Size}}$$

Now that these two properties are verified, we move onto the next property, which is verifying that the sampling distribution would be a **normal distribution.**

In the above example, you already saw that the sampling distribution was nearly a normal distribution. However, you may wonder that since the original distribution of the population was also normal, therefore this leads to the normal behaviour for the sampling distribution. In order to verify whether it is indeed true or not, take a look at the next lecture.

Thus, no matter the parent population distribution, when you take samples, compute their means and find the sampling distribution, it **will always be normal, or at least nearly normal.** This is one of the most important implications of the Central Limit Theorem. The following image summarises the results obtained in the above video.

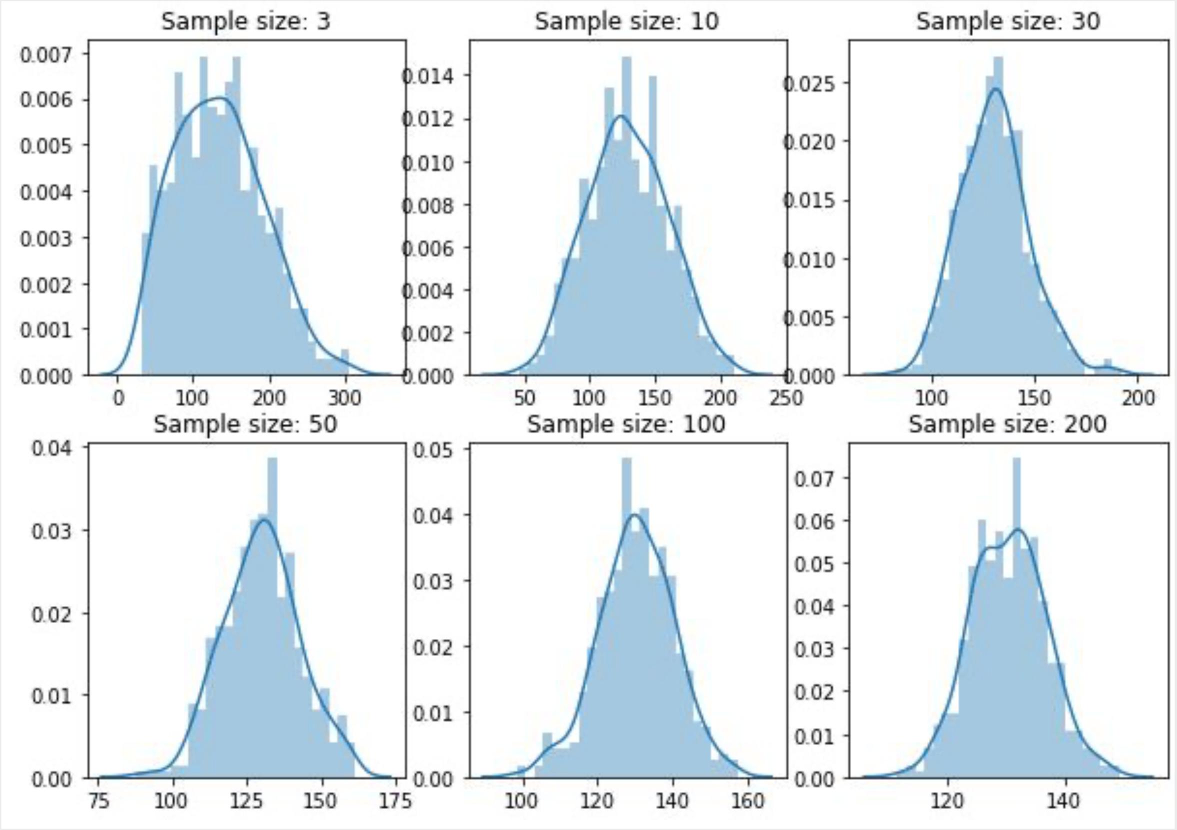


This concludes the first part of the demonstration. In the second part, you will observe the effect of sample size on the resulting sampling distribution.

CLT Demonstration: II:-

There is only one more thing to consider now, which is the sample size. Here, we will observe that as the sample size increases, the underlying sampling distribution will approximate a normal distribution even more closely.

As you saw, **a sample size of 30** or above is ideal for concluding that the sampling distribution is nearly normal and further inferences can be drawn from it. The following image portrays the sampling distribution's nature for different sample sizes



Now, let's summarise the learnings of this demonstration.

To reiterate, here are the properties of the sampling distribution as given by the Central Limit Theorem for any kind of data provided a high number of samples has been taken

1. **Sampling distribution's mean** ($\mu_{\bar{X}}$) = **Population mean** (μ),
2. Sampling distribution's standard deviation (**standard error**) = $\frac{\sigma}{\sqrt{n}}$, and
3. **For n** > **30**, the sampling distribution becomes a **normal distribution**.

Now that you have understood the properties of Central Limit Theorem, you are well equipped to infer the population mean from the sample mean.

Recall that in the first lecture on samples, we found the mean commute time of 30,000 employees of an office by taking a small sample of 100 employees and finding their mean commute time. This sample's mean was

$\bar{X} = 36.6$

minutes and its standard deviation was S = 10 minutes.

We then said that this sample mean cannot be taken as the population mean, as there might be some errors in the sampling process. However, we can say that the population mean, i.e., the daily commute time of all 30,000 employees

$$\bar{X} = 36.6 \text{ (sample mean) } \pm \text{ some margin of error.}$$

Now, you may be thinking that you can use the standard error for the margin of error. However, keep in mind that although the standard error provides a good estimate of this margin of error, you cannot use it in place of the margin of error. To understand why and how you would find the margin of error in that case, let's move on to the next lecture, where we will use the CLT (central limit theorem) to find the aforementioned margin of error.

Additional Notes

- As mentioned earlier, this code is for demonstration purposes only. However, if you want you can change the target population from 'Weight' to 'Height' and verify the results as well.

Summary: Central Limit Theorem - Part I:-

Let’s summarise everything that has been taught so far in this session, and then, you can move on to the rest of the session.

First, you saw how instead of finding the mean and standard deviation for the entire population, it is sometimes beneficial to find the **mean** and **standard deviation** for only a small representative **sample**. You may have to do this because of time and/or money constraints.

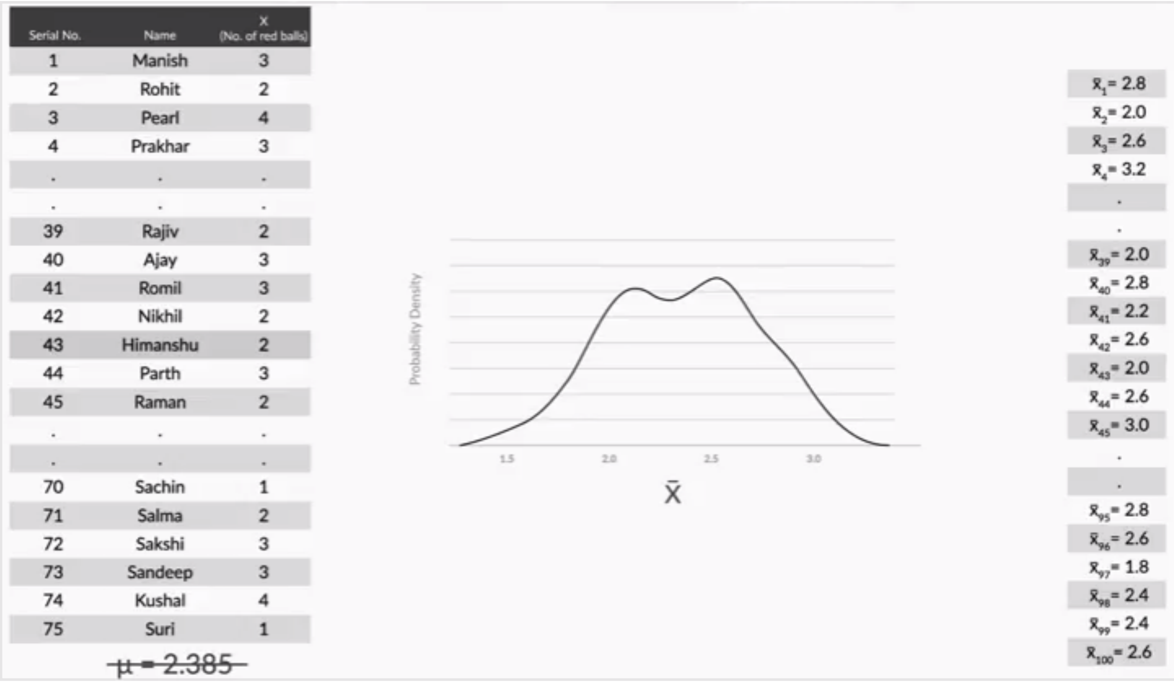
For example, for an office of 30,000 employees, we wanted to find the average commute time. So, instead of asking all employees, we asked only 100 of them and collected the data. We found the mean to be 36.6 minutes and the standard deviation to be 10 minutes.

However, it would not be fair to infer that the population mean is exactly equal to the sample mean. This is because the flaws of the sampling process must have led to some error. Hence, the sample mean’s value has to be reported with some **margin of error**.

For example, the mean commute time for the office of 30,000 employees would be equal to 36.6 ± 3 minutes, 36.6 ± 1 minutes, or 36.6 ± 10 minutes, i.e., 36.6 minutes ± some margin of error.

However, at this point in time, you do not exactly know how to find what this margin of error is.

Then, we moved on to sampling distributions, some of the properties which would help you find this margin of error.



We created a sampling distribution, which was a probability density function for 100 sample means with a sample size = 5.

A sampling distribution, which is essentially the distribution of the sample means of a population, has some interesting properties, which are collectively called the **central limit theorem**. It states that no matter how the original population is distributed, the sampling distribution will follow these three properties:

1. **Sampling distribution's mean** ($\mu_{\bar{X}}$) = **Population mean** (μ),
2. Sampling distribution's standard deviation (**Standard error**) = $\frac{\sigma}{\sqrt{n}}$, where σ is the population's standard deviation and n is the sample size, and
3. **For n** > **30**, the sampling distribution becomes a **normal distribution**.

To verify these properties, we performed sampling using the data collected for our upGrad game from the first session on inferential statistics. The values for the sampling distribution thus created

($\mu_{\bar{X}}$ = 2.348,

S.E. = 0.4248) were quite close to the values predicted by theory ($\mu_{\bar{X}}$ = 2.385, S.E. = 0.44).

To summarise, the notations and formulas for populations, samples and sampling distributions are as follows:

Population/Sample	Term	Notation	Formula
Population ($X_1, X_2, X_3, \ldots\ldots, X_N$)	Population Size	N	Number of items/elements in the population
	Population Mean	μ	$\frac{\sum_{i=1}^{i=N} X_i}{N}$
	Population Variance	σ^2	$\frac{\sum_{i=1}^{i=N} (X_i - \mu)^2}{N}$
Sample ($X_1, X_2, X_3, \ldots\ldots, X_n$) (Sample of Population)	Sample Size	n	Number of items/elements in the sample
	Sample Mean	$\bar{X}$	$\frac{\sum_{i=1}^{i=n} X_i}{n}$
	Sample Variance	S^2	$\frac{\sum_{i=1}^{i=n} (X_i - \bar{X})^2}{n-1}$
Sampling Distribution of the Sample Mean ($\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots\ldots, \bar{X}_k$) (k Sample Means)	Sampling Distribution's Size	No convention (We have used k, but that is not a norm)	
	Sampling Distribution's Mean (mean of sample means)	$\mu_{\bar{X}}$	$\mu_{\bar{X}} = \mu$
	Sampling Distribution's Standard Deviation	S.E. (Standard Error)	S.E. = $\sigma/\sqrt{n}$

Before moving on to the next lecture, let's spend some time attempting a few practice questions.

Estimating Mean Using CLT:-

Now that you have gone through the mid-session summary, let's get back to the rest of the session. Earlier, we tried to estimate the mean commute time of 30,000 employees of an office by taking a small sample of 100 employees and finding their mean commute time. This sample's mean was

$\bar{X}$

= 36.6 minutes and its standard deviation was S = 10 minutes.

Recall that we also said that the **population mean**, i.e., the daily commute time of all 30,000 employees (μ) = 36.6 (**sample mean**) $\pm$ some **margin of error**.

You can find this margin of error using the CLT (central limit theorem). Now that you know the CLT, let’s see how you can find the margin of error.

To summarise, let’s say that you have a sample with sample size n, mean

$\bar{X}$

and standard deviation S. Now, the **y% confidence interval** (i.e., the confidence interval corresponding to a y% confidence level) for

μ

would be given by the range:

$$\text{Confidence interval} = \left(\bar{X} - \frac{Z^*S}{\sqrt{n}}, \bar{X} + \frac{Z^*S}{\sqrt{n}}\right),$$

where, **Z* is the Z-score associated with a y% confidence level**. In other words, the population mean and the sample mean differ by a **margin of error** given by

$$\frac{Z^*S}{\sqrt{n}}.$$

Some commonly used Z* values are given below:

Confidence Level	Z*
90%	±1.65
95%	±1.96
99%	±2.58

At this point, it is important to address a common misconception. Sampling distributions are just a theoretical exercise; you’re not actually expected to make one in real life. If you want to estimate the population mean, you will just take a sample. You will not create an entire sampling distribution.

You must be wondering why you studied sampling distributions if this is the case. To understand the reason for this, let's go through the actual process of sampling. Recall that you are doing **sampling because you want to find the**

population mean, albeit in the form of an interval. The three steps to follow are as follows:

1. First, take a sample of size n.
2. Then, find the mean $\bar{X}$ and standard deviation S of this sample.
3. Now, you can say that for a y% confidence level, the confidence interval for the population mean μ is given by $(\bar{X} - \frac{Z^*S}{\sqrt{n}}, \bar{X} + \frac{Z^*S}{\sqrt{n}})$.

However, as you may have seen in the video above, **you cannot finish step 3 without the CLT**. The CLT lets you assume that the sample mean would be normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (approx. $\frac{S}{\sqrt{n}}$). Using this assumption, it is possible to find the margin of error, confidence interval, etc.

Thus, you learnt about sampling distributions so that you could learn more about the CLT and be able to make all the assumptions as stated above. In the next segment, you will learn how to estimate confidence intervals.

Confidence Interval - Example:-

In the following video, Professor Tricha will walk you through a real-life example of how confidence intervals can be used to make decisions.

Let's consider another example of how you can use confidence intervals to make a decision. Recall the Facebook example discussed in the last session. Let's find the 90% confidence interval (confidence interval for a 90% confidence level) for that case.

Recall that 50.5% of the 10,000 people surveyed preferred feature B to feature A. So, if X = the proportion of people that prefer feature B to feature A, then, for this sample,

$\bar{X}$

= 0.505 (50.5%) and n = 10,000. In addition to this, you've been told that the sample's standard deviation S = 0.2(20%).

Also, you know that the actual population mean

μ

lies between

$\bar{X}$

± margin of error. However, now it is vital for us to find this margin of error.

If this margin of error is, say, 1%, then that means that the population mean

μ

, which is the proportion of people that prefer feature B to feature A, lies between the range (50.5 - 1)% to (50.5 + 1)%, i.e., 49.5 % to 51.5%. This means that you cannot say with certainty that

μ

would be more than 50%. So, even though the proportion of people that prefer feature B to feature A is more than 50% in our sample, you would not be able to say with certainty that this proportion would be more than 50% for the entire population.

On the other hand, if the margin of error is, say, 0.3%, then you will be able to say that the population mean lies within (50.5 - 0.3)% and (50.5 + 0.3)%, i.e., 50.2% to 50.8%. So, you will be able to say with certainty that the proportion of people that prefer feature B to feature A is more than 50% in our sample and for the entire population too.

Now, the margin of error corresponding to a 90% confidence level would be given by

$$\frac{Z^*S}{\sqrt{n}} = \frac{1.65 * 0.2}{\sqrt{10,000}} =$$

0.0033 (0.33%), and the population mean lies between 50.17% and 50.83%.

Hence, you can say that feature B should replace feature A with 90% confidence.

Summary: Central Limit Theorem - Part II:-

First, you learnt how you can use your knowledge of the **CLT** to **infer the population mean from the sample mean**.

We estimated the mean commute time of 30,000 employees of an office by taking a sample of 100 employees, finding their mean commute time. Specifically, you were given a sample with a sample mean $\bar{X}$ = 36.6 minutes and a sample standard deviation S = 10 minutes.

Using the CLT, you concluded that the sampling distribution for the mean commute time would have the following:

1. Mean = μ {unknown}
2. Standard error = $\frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$
3. Since n(100) > 30, the sampling distribution is a normal distribution

Using these properties, you were able to claim that the probability that the population mean μ lies between 34.6 (36.6 - 2) and 38.6 (36.6 + 2) is 95.4%.

Then, you learnt the following terminology related to the claim:

1. The probability associated with the claim is called the **confidence level**. (Here, it is 95.4%.)
2. The maximum error made in a sample mean is called the **margin of error**. (Here, it is 2 minutes.)
3. The final interval of values is called the **confidence interval**. [Here, it is the range (34.6, 38.6).]

You then generalised the whole process. Let's say you have a sample with a **sample size n, mean** $\bar{X}$ and **standard deviation S**. You learnt that a **y% confidence interval** (i.e., a confidence interval corresponding to a y% confidence level) for μ will be given by the range:

Confidence interval = $(\bar{X} - \frac{Z^* S}{\sqrt{n}}, \bar{X} + \frac{Z^* S}{\sqrt{n}})$,

Where, **Z*** is the Z-score associated with a **y% confidence level**.