Module 3 Topic 2 Video 1

Random Variables and Probability Distribution

Learning Objective

- What are random variables?
- What are probability distributions?
- Types of probability distributions: Discrete and Continuous
- PDF Vs PMF Vs CDF

Random Variable

```
1  movies.Rating
```

```
0       5.6
1       2.2
2       5.0
3       6.2
4       6.5
        ...
5270    6.0
5271    6.0
5272    5.9
5273    4.2
5274    7.5
```

Note that the values have decimal places a.k.a. float or continuous

- *Movie rating* is a variable that seems to have random values
- The rating is unknown until the movie is released and rated by critics. So the rating for each movie is a random variable
- A random variable is a variable that takes values that appear to be random in nature
- Since the values are continuous or float, this is an example of continuous random variable

Discrete Random Variable

```
1  movies.Rating
```

```
0       5
1       2
2       5
3       6
4       6
       ..
5270    6
5271    6
5272    5
5273    4
5274    7
```

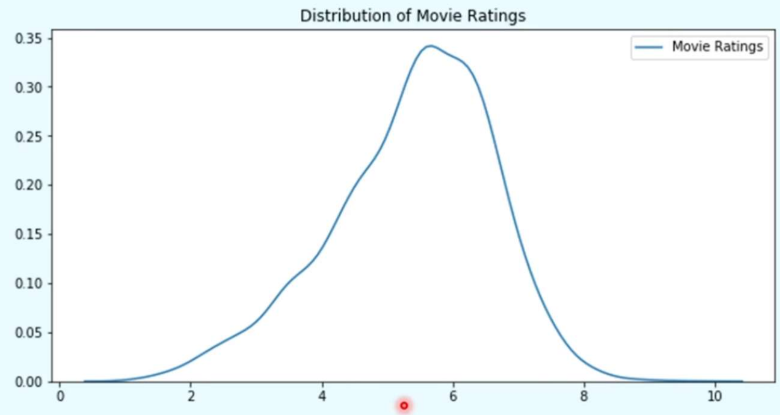
Note that the values do not have decimal places a.k.a. integer or discrete

- If we convert the rating to integer then it becomes discrete
- Then the random variable *movies rating* is a discrete random variable
- Some variables are discrete by nature e.g. *Genre*
- However, some continuous variables such as *rating* can be expressed as both continuous as well as discrete accounting for some information loss

Distribution of Continuous Random Variable

```
1  movies.Rating
```

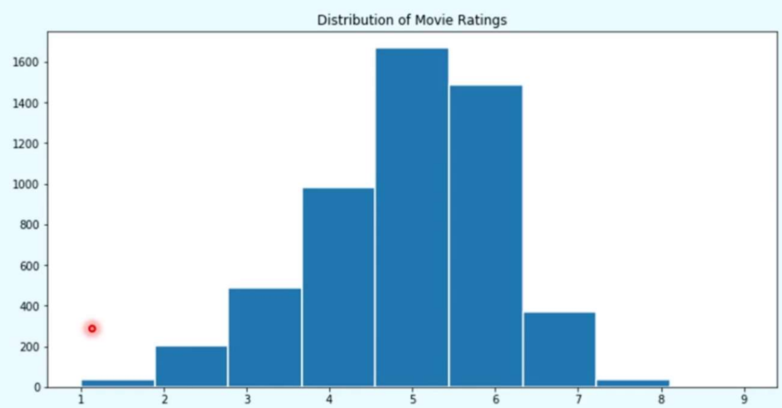
```
0       5.6
1       2.2
2       5.0
3       6.2
4       6.5
        ...
5270    6.0
5271    6.0
5272    5.9
5273    4.2
5274    7.5
```



Distribution of Discrete Random Variable

```
1  movies.Rating
```

```
0       5
1       2
2       5
3       6
4       6
        ..
5270    6
5271    6
5272    5
5273    4
5274    7
```

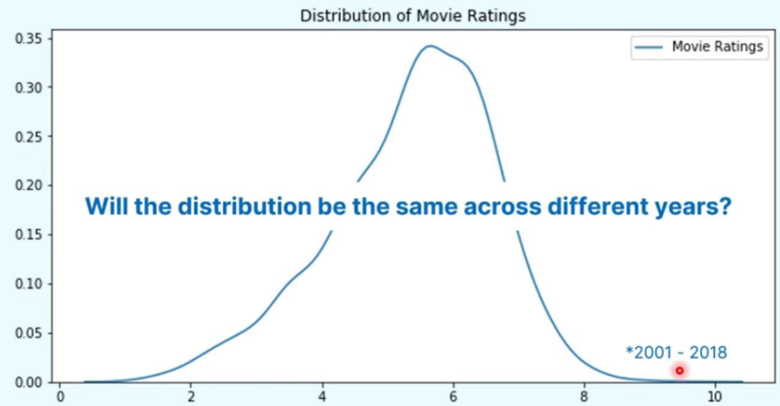


Distribution of Continuous Random Variable

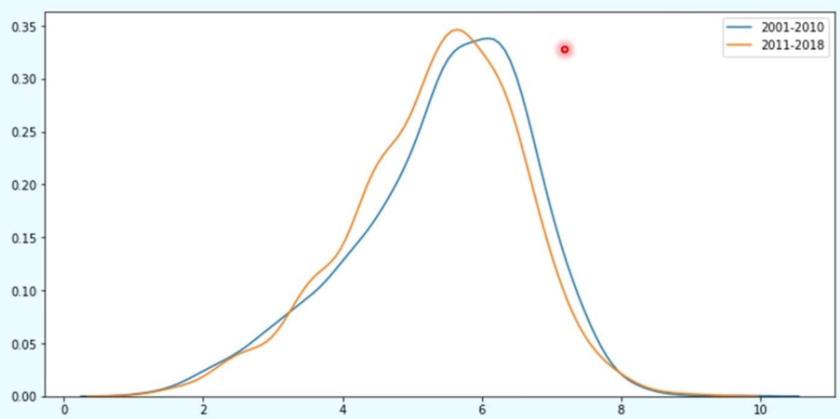
```
movies.Rating
```

```
0       5.6
1       2.2
2       5.0
3       6.2
4       6.5
       ...
5270    6.0
5271    6.0
5272    5.9
5273    4.2
5274    7.5
```

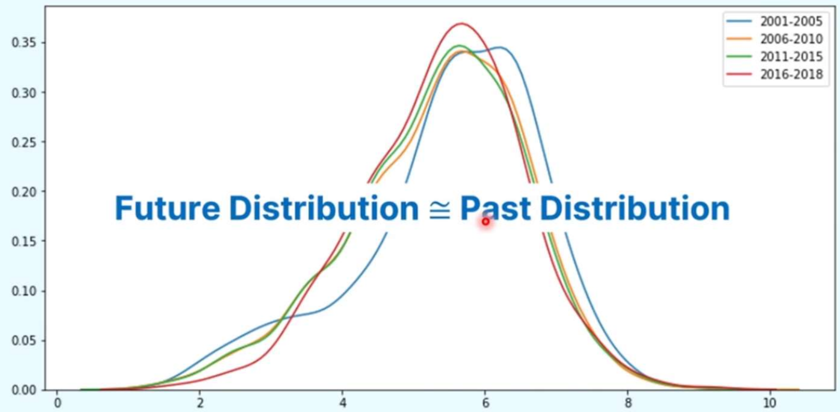
Note that the values have decimal places a.k.a. float



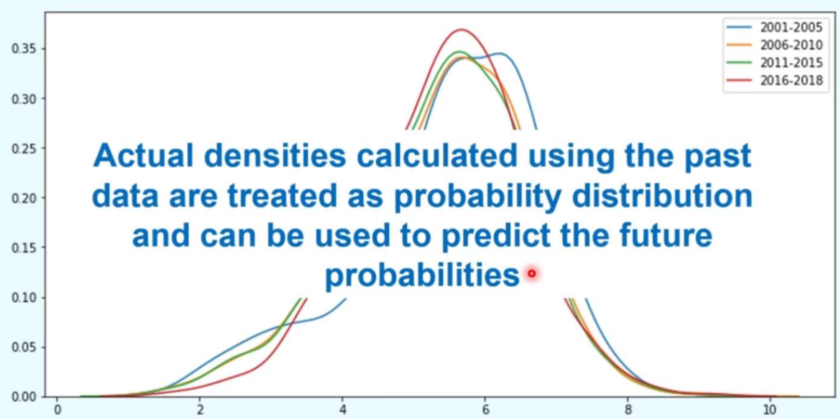
Movies Rating across Multiple Years



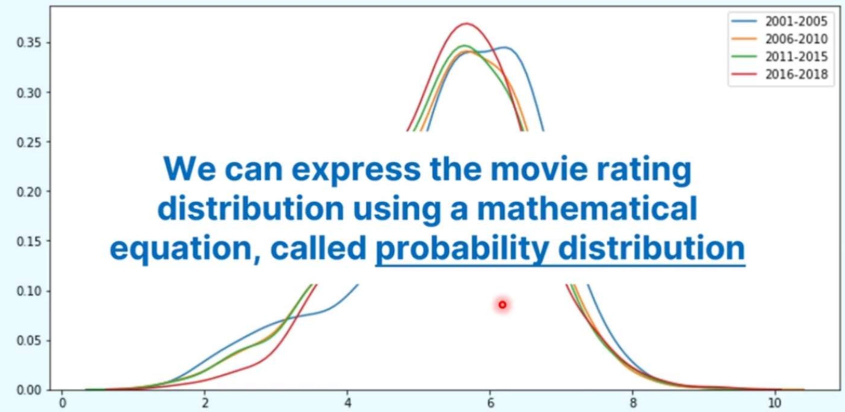
Movie Rating across Multiple Years



Movie Rating across Multiple Years

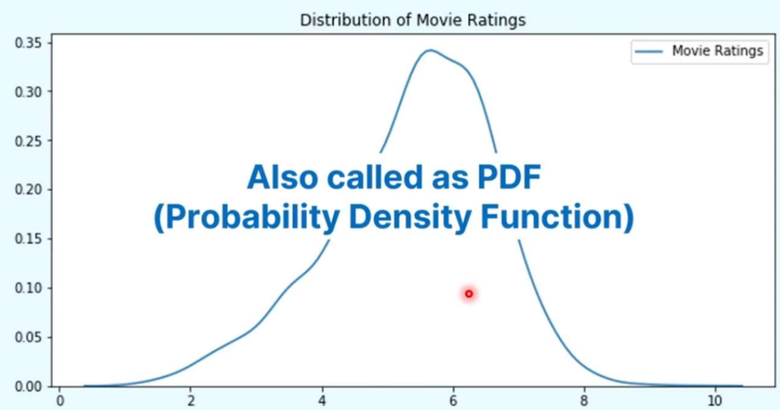


Movie Rating across Multiple Years



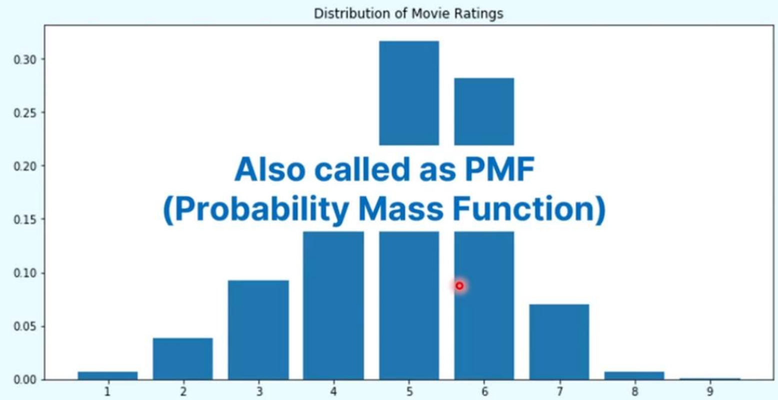
We can express the movie rating distribution using a mathematical equation, called <u>probability distribution</u>

Probability Density Function

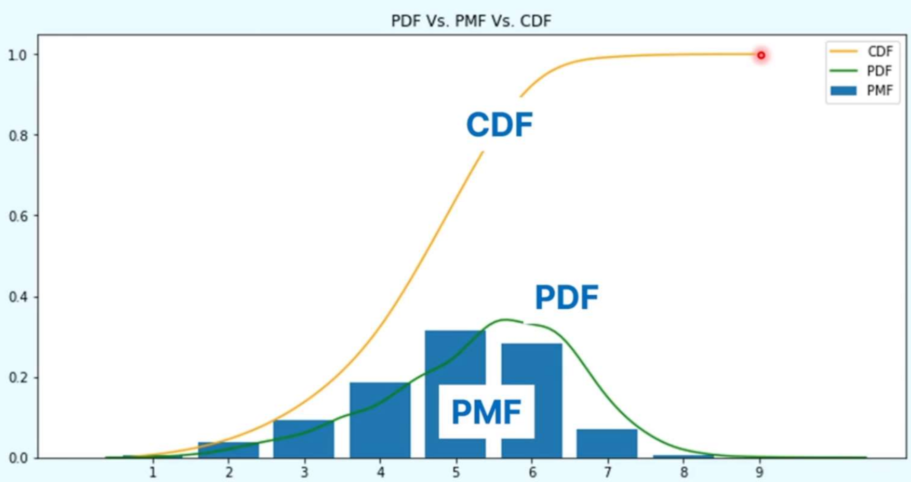


Also called as PDF (Probability Density Function)

Probability Mass Function



Cumulative Distribution Function



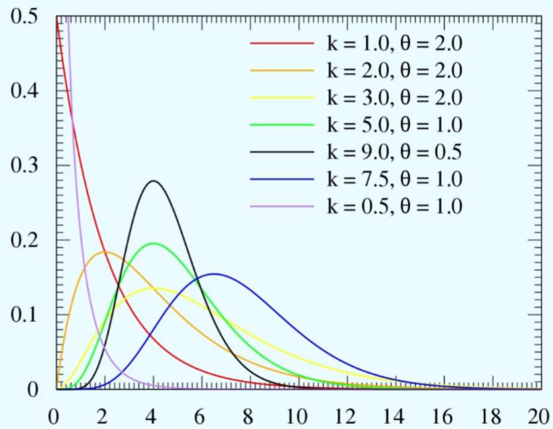
What did we learn?

- A random variable is a variable that takes values that appear to be random in nature
- Random variables can be discrete and continuous
- Actual densities calculated using the past data can be treated as probability distribution to predict the future probabilities
- The distribution of a random variable can also be expressed as a mathematical expression called as probability distribution
- PDF – PMF - CDF

Module 3 Topic 2 Video 2

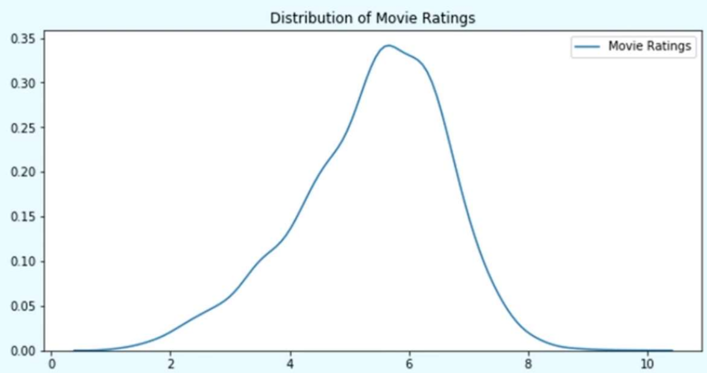
Using Probability Distribution to Estimate Probabilities

There are many distribution functions



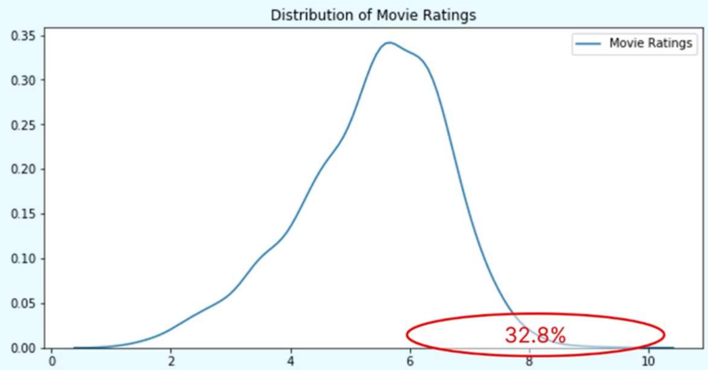
- Uniform distribution
- Binomial distribution
- Poisson distribution
- Normal distribution
- T-distribution
- Chi-square distribution
- F distribution
- etc

P(Rating > 6) from Movie Rating Distribution



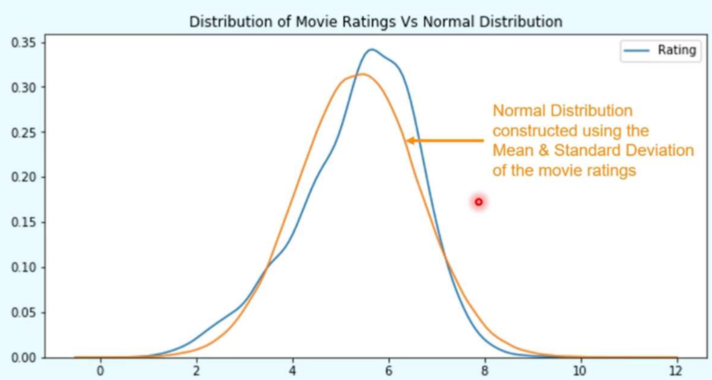
$$P(Rating > 6) = \frac{Total\ No.of\ Movies\ with\ a\ Rating > 6}{Total\ No.of\ Movies}$$

P(Rating > 6) from Movie Rating Distribution



$$P(Rating > 6) = \frac{Total\ No.\, of\ Movies\ with\ a\ Rating > 6}{Total\ No.\, of\ Movies}$$

P(Rating > 6) from Normal Distribution



What did we learn?

- The distribution of a random variable can also be expressed as a mathematical expression called as probability distribution
- We do not need the full data, we just need the parameters such as mean, standard deviation etc. to estimate probabilities.

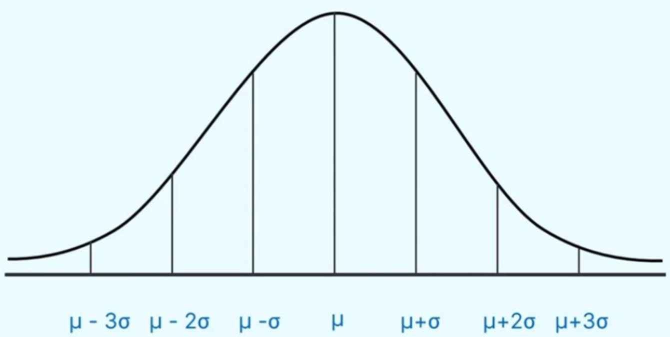
Module 3 | Topic 2 | Video 3

Normal Distribution

Learning Objective

- What is Normal distribution?
- Properties of Normal distribution
- Parameters that define a Normal distribution
- Estimate probabilities using Normal distribution in Python

Normal Distribution

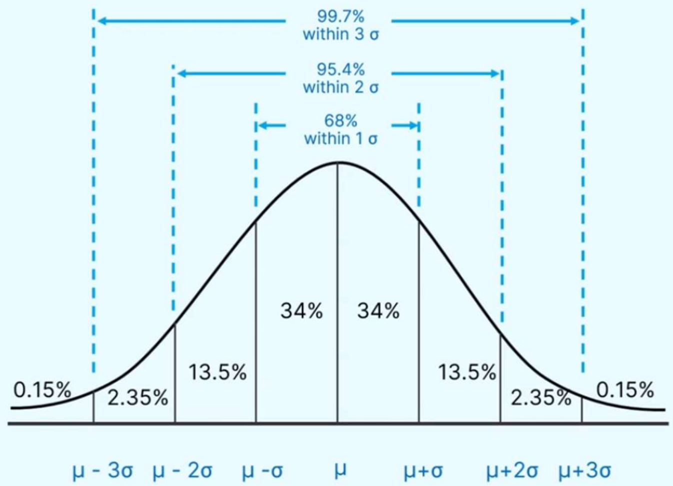


- Follows a bell curve
- Mean = median = mode
- Symmetric about its mean
- Asymptotic curve

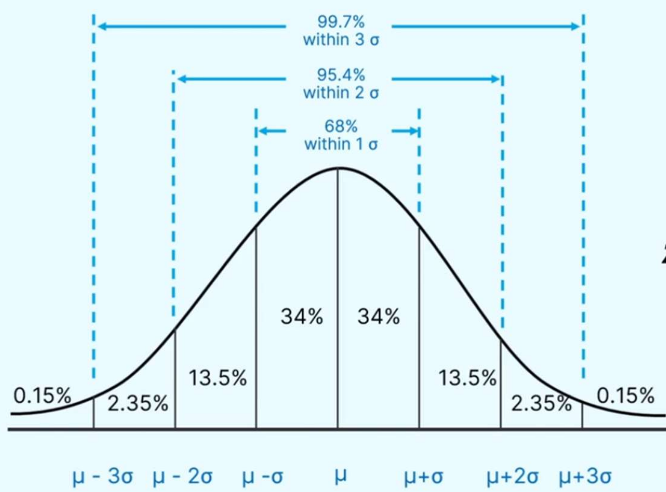
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-[\frac{(x-\mu)^2}{2 \cdot \sigma^2}]}$$

- Parameters of a normal distribution:
- μ = mean value of the population
- σ = standard deviation of the population

Normal Distribution



Standard Normal Distribution



- $\mu = 0$
- $\sigma = 1$

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

What did we learn?

- Normal distribution is bell curve with mean, median and mode all being same
- Normal distribution is defined using the mean **(μ)** and standard deviation **(σ)**
- In a Normal distribution, 68% of the data falls within the **$\mu + \sigma$,** 95.4% of the data falls within **$\mu + 2\sigma$** and 99.7% data falls within **$\mu + 3\sigma$**
- In a standard normal distribution mean, **$\mu = 0$** and standard deviation **$\sigma = 1$**

Learning Objective

- T-distribution Vs. Normal distribution
- Estimate probabilities using T-distribution

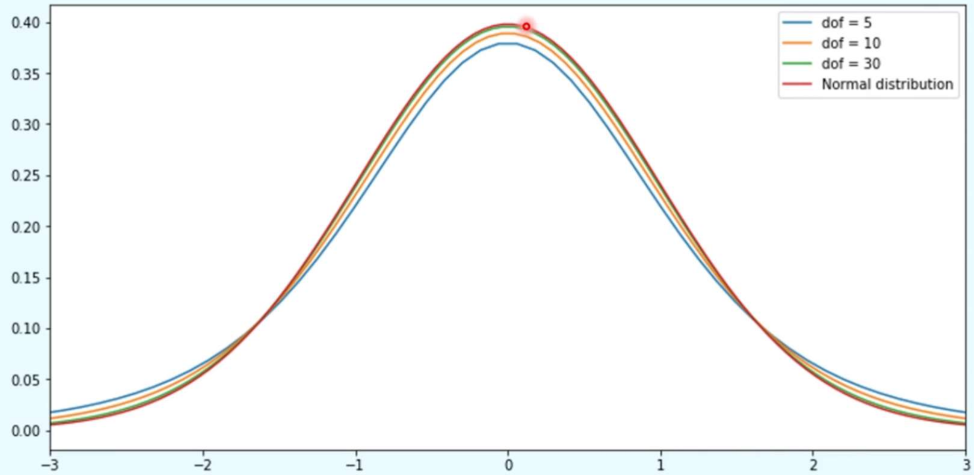
Module 3 Topic 2 Video 5

T-Distribution

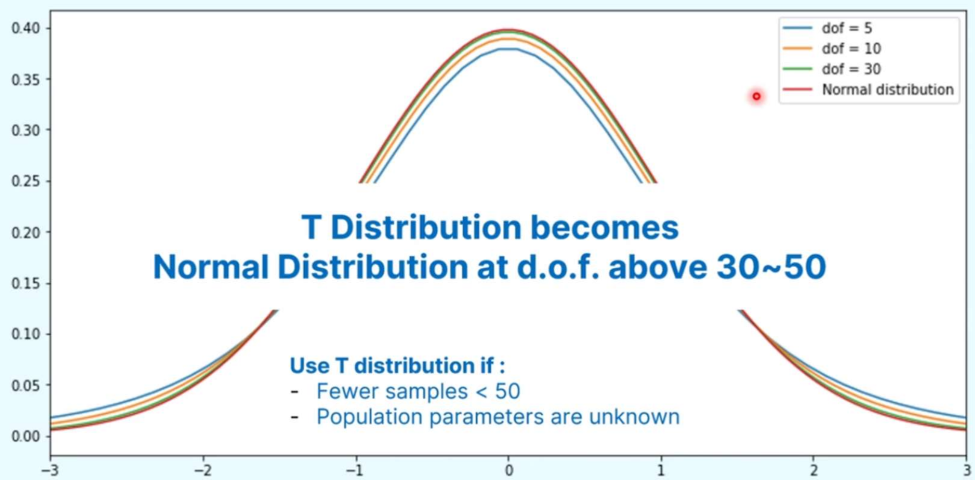
Learning Objective

- T-distribution Vs. Normal distribution
- Estimate probabilities using T-distribution

T-Distribution Vs Normal Distribution



T-Distribution Vs Normal Distribution



FreshCo

After seeing a possible loss of $80 per shipment due to the delays, you have asked your process improvement team to improve the process and cut down the transport time.

After making several changes, the team recorded order to delivery time in minutes for the last 10 deliveries as follows:

528, 566, 589, 495, 582, 573, 545, 593, 592, 664

What is the probability of a delivery being rejected after the implementation of the new process?

Do you see an improvement?

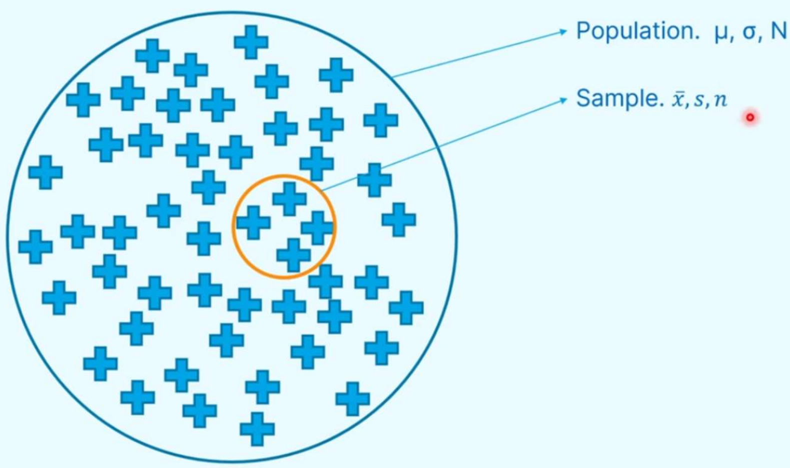
Module 3 Topic 3 Video 1

Sampling Distribution and Central Limit Theorem

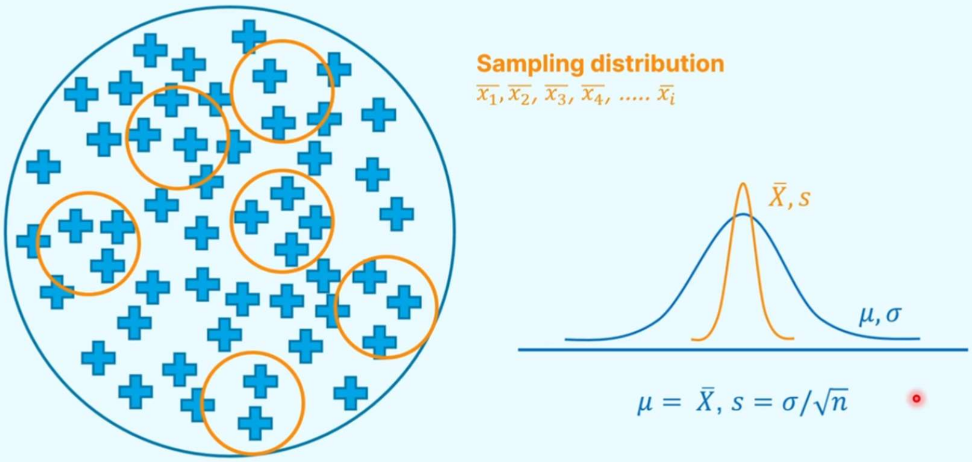
Learning Objective

- Review Sampling Theory
- Sampling Distribution
- Properties of Sampling Distribution
- Central Limit Theorem

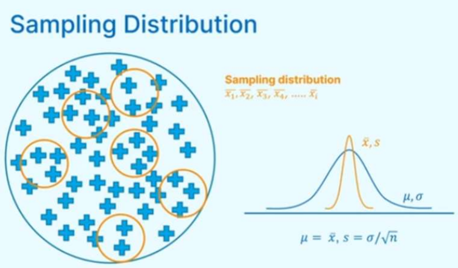
Population Vs Sampling



Sampling Distribution



Properties of Sampling Distributions



Distribution	Population	Sample dist.
Mean	μ	$\bar{X}$
Variance	σ^2	s^2
Standard deviation (SE)	σ	$s = \frac{\sigma}{\sqrt{n}}$
Size	N	n

Properties of Sampling Distribution

1. Sample means are normally distributed about the true population mean
2. If sufficiently large samples (n) are taken, irrespective of the shape of the distribution of the population, the sampling distribution will always follow normal distribution
3. Sampling mean is an unbiased estimator of the population mean (i.e. average of all sample means equals the population mean
4. The standard deviation of the sampling distribution is called the standard error $(\sigma/\sqrt{n})$
5. With more the samples n, the sampling distribution would become less spread $(\sigma/\sqrt{n})$

> Central Limit Theorem <

04:35

Module 3 Topic 3 Video 2

Central Limit Theorem

Central Limit Theorem

1. Sample means are normally distributed about the true population mean
2. If sufficiently large samples (n) are taken, irrespective of the shape of the distribution of the population, the sampling distribution will always follow normal distribution
3. Sampling mean is an unbiased estimator of the population mean (i.e. average of all sample means equals the population mean
4. The standard deviation of the sampling distribution is called the standard error ($\sigma/\sqrt{n}$)
5. With more the samples n, the sampling distribution would become less spread ($\sigma/\sqrt{n}$)