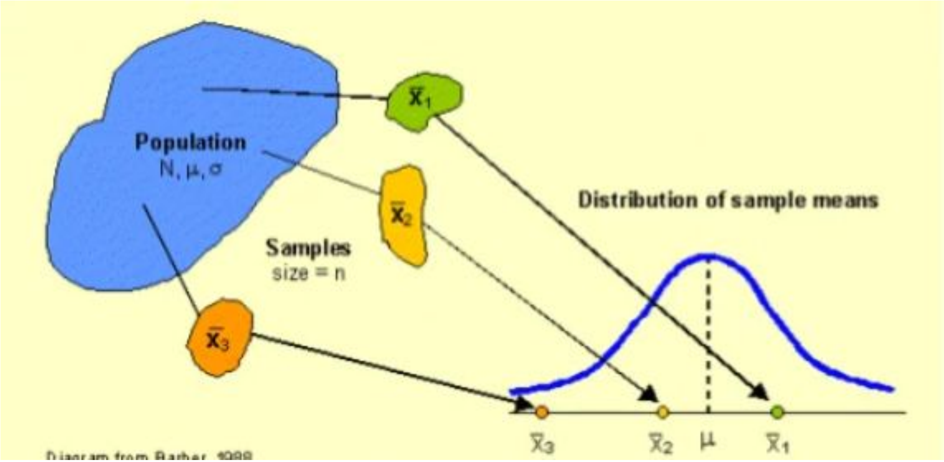
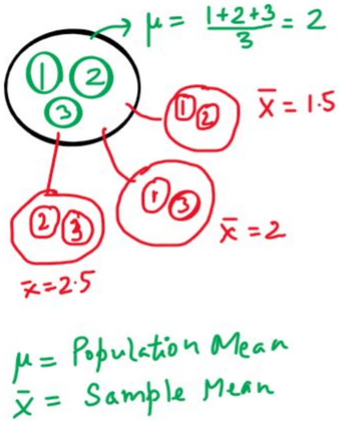
Inferential Analysis

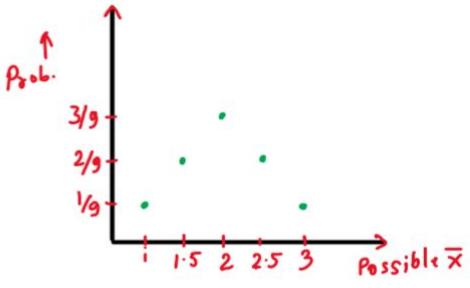
Part 3

Sampling Distribution

- A sampling distribution is a probability distribution of a statistic (such as the mean) that results from selecting an infinite number of random samples of the same size from a population. The sampling distribution of a given population is the distribution of frequencies of a range of different outcomes that could possibly occur for a statistic of a population.
- A sampling distribution is a statistic that is arrived out through repeated sampling from a larger population.
- It describes a range of possible outcomes that of a statistic, such as the mean or mode of some variable, as it truly exists a population.
- The majority of data analyzed by researchers are actually drawn from samples, and not populations.







Sampling distribution for the sample mean for sample size 2

Types of Sampling Distribution

As we have already discovered about Sampling Distribution, we will now learn about the various types of Sampling Distribution in statistics. To begin with, there are 3 types of Sampling Distribution. They are as follows-

Sampling Distribution of Mean

The first and foremost type of sampling distribution is of the mean. This type focuses on calculating the mean average of all sample means which then lead to sampling distribution. The average of every sample is put together and a sampling distribution mean is calculated which reflects the nature of the whole population. With more samples, the standard deviation decreases which leads to a normal frequency distribution or a bell-shaped curve on the graph.

Sampling Distribution of Proportion

When it comes to the second type of Sampling Distribution, the population's samples are calculated to obtain the proportions of a population. Herein, the mean of all sample proportions is calculated, and thereby the sampling distribution of proportion is generated. As the proportion of a population is defined by a part of the population that possesses a certain attribute, the sampling distribution of proportion aims to achieve a mean of all sample proportions that involve the whole population.

T-Distribution

Third of all, T-Sampling Distribution is considered to involve a small size of the population that gives about no information about standard deviation. Under this type of sampling distribution, the population size is very small that, in turn, leads to a normal distribution. The frequent distribution in this type is the most near to the mean of the sampling distribution. Only a handful of samples are far off from the mean value of the whole population.

Content Source: www.analyticssteps.com

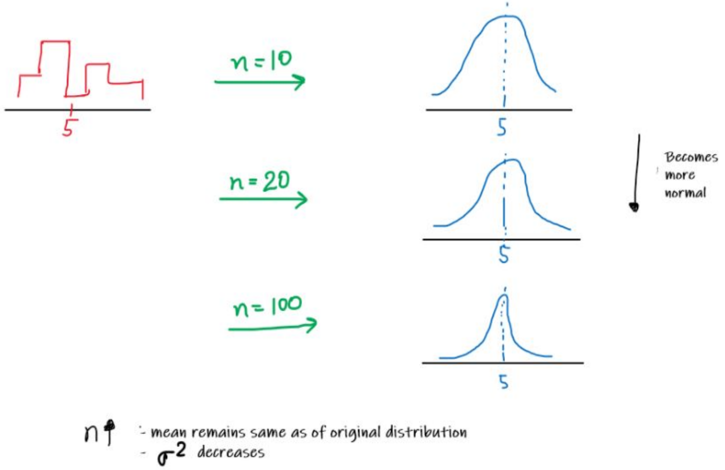
Central Limit Theorem

• The Central Limit Theorem states that the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger - no matter what the shape of the population distribution.

• This fact holds especially true for sample sizes over 30.

• All this is saying is that as you take more samples, especially large ones, your graph of the sample means will look more like a normal distribution.

• The standard deviation of the sample means will be smaller than the population standard deviation and will be equal to the standard deviation of the population divided by the square root of the sample size.



- To determine the probability of sample mean falls within a particular range, we use,

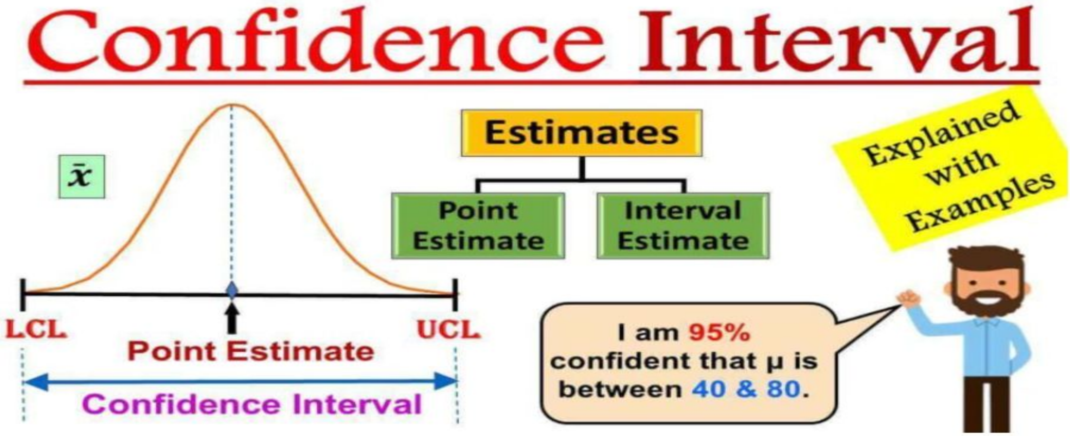
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- The sample mean $\bar{x}$ is a random variable and its probability distribution is called the sampling distribution of $\bar{x}$
- The sampling distribution is denoted by Z
- Where, Z follows Normal distribution with E($\bar{x}$) and standard deviation $\sigma_{\bar{x}}$
- Here, E($\bar{x}$) = μ and $\sigma_{\bar{x}} = \sigma / \sqrt{n}$

Confidence Interval

- A confidence interval, in statistics, refers to the probability that a population parameter will fall between a set of values for a certain proportion of times.
- A confidence interval displays the probability that a parameter will fall between a pair of values around the mean.
- Confidence intervals measure the degree of uncertainty or certainty in a sampling method.
- They are most often constructed using confidence levels of 95% or 99%.

Confidence Interval = "Point estimate" +/- "Margin of Error"



Example 1 : CI for Single population Mean

A random sample of n = 50 males showed a mean average daily intake of dairy products equal to 756 grams with a standard deviation of 35 grams. Find a 95% and 99% confidence interval for the population average U ?

Ans: x¯= 756 grams ; n= 50 , σ = 35 ,

From Z table for 95 % confidence interval we get value as 1.96.

Interval estimate is = 756±1.96×35/√50

746.30 ≤ m ≤ 765.70 grams for 95 % confidence interval

$$\boxed{\bar{X} \pm z \cdot \frac{\sigma}{\sqrt{n}}}$$

Example: https://digitalelearnings.com/

Example 2 : CI for Single population Proportion

Billing statement for 1000 patients discharged from a particular hospital were randomly selected for error. Out of 1000 billing statements, 102 were found to contain errors. Using this formation lets construct 99% confidence interval ?

Ans: p^ = 102 / 1000 = 0.102

From Z table for 99 % confidence interval we get value as 2.576

$$\boxed{\hat{p} \pm z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$$\hat{p} \; \frac{x}{n}$$

p^ = is random and varies from sample to sample

$$0.102 \pm 2.576 * \sqrt{\frac{0.102(1-0.102)}{1000}}$$

0.077 ≤p^≤0.127for 99 % confidence interval