# 1. Introduction

Welcome to the series of E-learning modules on Confidence Intervals for the means. In this module, we are going to cover the Interval Estimation- procedure to estimate the population mean when the population standard deviation is known and unknown.

By the end of this session, you will be able to:

- Explain the interval estimation
- Explain the confidence interval for the population mean when variance is known
- Explain the confidence interval for the population mean when variance is unknown
- Explain how to apply interval estimation technique for the estimation of unknown population mean

Statistical inference – specifically, decision making and prediction - centuries old and plays a very important role in most peoples' lives.
For example:

- The Government needs to predict short and long term interest rates
- A broker wants to forecast the behaviour of the stock market
- A consumer wants to estimate the selling price of his/her house before putting it on the market

There are many ways to make these decisions or predictions, some subjective and some more objective in nature. How good our predictions or decisions be? Although you may feel that your own built in decision making ability is quite good, experience suggests that this may not be the case.

It is the job of the mathematical statistician to provide methods of statistical inference making that are better and more reliable than just subjective guesses. As discussed earlier, inference is concerned with making decisions or predictions about parameters – the numerical descriptive measures that characterize a population.

These parameters may be population mean mu, the population standard deviation sigma, the binomial proportion P. In statistical inference, a practical problem is restated in the framework of population with a specific parameter of interest. One of the methods of making inference is estimation- Estimating or predicting the value of the parameter.

A statistical problem, which involves planning, analysis and inference making, is incomplete without a measure of the goodness of the inference. That is how accurate or reliable is the method you have used?
If a stock broker predicts that the price of a stock will be eighty dollars next Monday, will you be willing to take actions to buy or sell your stock without knowing how reliable his/her prediction is? Will the prediction be within one dollar, two dollars or ten dollars of the actual price of the next Monday?

Statistical procedures are important because they provide the following two types of information:

- Methods of making the inference
- A numerical measure of the goodness or reliability of the inference

As we discussed earlier, in a decision-making process the parameters of interest can be estimated using either point estimation or interval estimation techniques. If we have to estimate the population mean using point estimation either we go for the method of moments or the method of maximum likelihood, which gives a single number as an estimate of the population mean.

However, an interval estimator is a rule for calculating two numbers say A and B. To create an interval that we are fairly certain contains the parameter of interest that is population mean mu. Hence, a confidence interval gives an estimated range of values for the population average, which is likely to include an unknown population mean, the estimated range being calculated from a given set of sample data.

In statistics, a **Confidence Interval** (**CI**) is a kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. We measure the probability using the confidence coefficient; designated by (1 minus alpha). A confidence interval can be used to describe the reliability of the survey results.

For example, the circuits in computers and other electronic equipments consist of one or more Printed Circuit Boards (PCB) and computers are often repaired by replacing one or more defective PCBs. In an attempt to find a proper setting of plating process applied to one side of PCB, a production supervisor might estimate the average thickness of copper plating on PCBs using samples from several days of operation.

Before observing the production process, the production supervisor has no knowledge of the average thickness mu and he/she is in an estimation problem. He/she can estimate the average thickness of copper plating on PCBs either using point estimation techniques or interval estimation techniques.
Already we are aware of the point estimation techniques to estimate an average of the population. Now, let us see how to get an interval estimate for the unknown population average.

Confidence Interval for a Mean
Practical problems very often lead to estimation of mu, the mean of the population. A confidence interval for a mean specifies a range of values within which the unknown population parameter, in this case the mean, may lie. For example:

- The average achievement of college students
- The average number of deaths per age category
- The average demand for a new cosmetic product

When the sample size n is large, the sample mean y bar is the best point estimator for the population mean mu. Since its sampling distribution is approximately normal, it can be used to construct a confidence interval according to the general approach.

# 2. Confidence Intervals for Mean and Known Standard Deviation

1) Confidence Intervals for Mean and Known Standard Deviation
Assumptions

- Population standard deviation is known
- Population is normally distributed
- If population is not normal, use large sample

Suppose y one, y two, up to $y_n$ are the random samples drawn from a population size N with mean mu and variance sigma square, we know that
Population mean mu is equal to summation Yi by N
And Sample mean y bar is equal to summation yi by n

Z is equal to (y bar minus mu) by Standard Error of y bar, which follows normal distribution with mean zero and variance one.
We can always find two quantities minus Z alpha by two and plus Z alpha by two from standard normal variate tables such that
Probability of minus Z alpha by two less than or equal to Z less than or equal to Z alpha by two is equal to 1 minus alpha
Probability of minus Z alpha by two less than or equal to (y bar minus mu) by standard error of y bar less than or equal to Z alpha by two is equal to 1 minus alpha

Probability of minus Z alpha by two into standard error of y bar less than or equal to (y bar minus mu) less than or equal to Z alpha by two into standard error of y bar is equal to 1 minus alpha
Probability of y bar minus Z alpha by two into standard error of y bar less than or equal to mu less than or equal to y bar plus Z alpha by two into standard error of y bar is equal to 1 minus alpha
Therefore, hundred into (one minus alpha) percent confidence interval for the population mean mu when the variance is known as sigma square is given by
[y bar minus Z alpha by two into standard error of y bar, y bar plus Z alpha by two into standard error of y bar]

However, standard error of y bar equals to square root of variance of y bar equals to square root of sigma square by n.
Ninety five percent confidence interval:
For ninety five percent confidence, alpha is equal to point zero five and alpha by two is equal to point zero two five. The value of Z as point zero two five is found by looking in the standard normal table. This area in the table is associated with a Z value of one point nine six.
In other words, of all the possible y-bar values along the horizontal axis of the normal distribution curve, ninety five percent of them should be within a Z score of one point nine six from the mean.

Ninety five percent confidence interval for the population mean mu is
[y bar minus one point nine six into square root of sigma square by n, y bar plus one point nine six into square root of sigma square by n]
Note that this interval is exact only when the population distribution is normal. For large samples from other population distributions, the interval is approximately correct by the Central Limit Theorem.

For example:
In a certain problem, the student calculated the sample mean of the diameter of cables to be hundred and one point eight two with its standard error point four nine. The critical value for a ninety five percent confidence interval is one point nine six and a ninety five percent confidence interval for the unknown mean mu is
(hundred and one point eight two minus (one point nine six into point four nine), hundred and one point eight two plus (one point nine six into point four nine))
Is equal to (hundred point eight six, hundred and two point seven eight)

As the level of confidence decreases, the size of the corresponding interval will also decrease.
The student was interested in a ninety percent confidence interval for the cable diameters. In this case, one minus alpha is equal to point nine zero and alpha by two is equal to point zero five. The critical value Z for this level is equal to one point six four five, so the ninety percent confidence interval is
(hundred and one point eight two minus (one point six four five into point four nine), hundred and one point eight two plus (one point six four five into point four nine))
Is equal to (hundred and one point zero one, hundred and two point six three)

# 3. Confidence Intervals for Mean and Unknown Standard Deviation

2. Confidence Intervals for Mean and Unknown Standard Deviation
Assumptions:

- Population standard deviation is unknown
- Population is normally distributed
- If population is not normal, use student's *t* distribution

In most practical research, the standard deviation for the population of interest is unknown. In this case, the standard deviation sigma is replaced by the estimated standard deviation *s*. Since the standard error is an estimate for the true value of the standard deviation, the distribution of the sample mean y bar is no longer normal with population mean mu and standard deviation sigma by root n.

Instead, the sample mean follows the *t distribution* with mean mu and standard deviation s by root n. The *t* distribution is also described by its *degrees of freedom*. For a sample of size *n*, the *t* distribution will have *n minus 1* degrees of freedom. The notation for a *t* distribution with *k* degrees of freedom is *t of k*. As the sample size *n* increases, the *t* distribution becomes closer to the normal distribution, since the standard error approaches the true standard deviation sigma for large *n*.

Suppose y one, y two, up to $y_n$ are the random samples drawn from a population of size N with unknown mean mu and unknown variance sigma square, we know that
Population mean mu is equal to summation Yi by N
And Sample mean y bar is equal to summation yi by n
s square is equal to summation (yi minus y bar) whole square by n minus one
Chi square is equal to summation (yi minus y bar) whole square by sigma square which is equal to (n minus one)into s square by sigma square which follows chi- square distribution with (n minus one) degrees of freedom

We have Z by square root of chi square (n minus 1) by n minus one is equal to y bar minus mu by sigma by root n divided by square root of (n minus one) into s square by sigma square into n minus one which is equal to y bar minus mu by s by root n follows t distribution with (n minus 1) degrees of freedom
When standard error of y bar is equal to s by root n
Y bar minus mu by standard error of y bar follows t alpha with (n minus one) degrees of freedom

We can always find two quantities minus t alpha (n minus 1) and plus t alpha (n minus 1) from students t-distribution table such that
Probability of minus t alpha (n minus 1) less than or equal to t less than or equal to t alpha (n minus 1) is equal to 1 minus alpha
Probability of minus t alpha (n minus 1) less than or equal to (y bar minus mu) by standard error of y bar less than or equal to t alpha (n minus 1) is equal to 1 minus alpha

Probability of minus t alpha (n minus 1) into standard error of y bar less than or equal to (y bar minus mu) less than or equal to t alpha (n minus 1) into standard error of y bar is equal to 1 minus alpha
Probability of y bar minus t alpha (n minus 1) into standard error of y bar less than or equal to mu less than or equal to y bar plus t alpha (n minus 1) into standard error of y bar is equal to 1 minus alpha

Standard error of y bar is equal to square root of variance of y bar which is equal to square root of s square by n
Therefore, 100 into 1 minus alpha percent C.I for the population mean μ when the variance is unknown is given by
[y bar minus t alpha (n minus 1) into square root of s square by n, y bar plus t alpha (n minus 1) into square root of s square by n]
Ninety five percent Confidence Interval for the population mean mu is
[y bar minus t point zero five (n minus 1) into square root of s square by n, y bar plus t point zero five (n minus 1) into square root of s square by n]
t alpha (n minus one) is the t distribution value with (n minus 1) degrees of freedom at alpha equal to zero point zero five (because one minus alpha is equal to point nine five, which implies alpha is equal to zero point zero five)

# 4. Steps to Obtain Confidence Interval for Mean

**Steps to obtain confidence interval for mean**

- First obtain the point estimate of mu, that is, the sample mean y bar
- If *n* is large, then the Central Limit Theorem can be used and y bar is normally distributed with mean mu and standard deviation sigma by root n
- Select a confidence level. The most common level is ninety five percent
- Obtain the margin of error associated with the confidence level. For a normal distribution, the interval from Z is equal to minus one point nine six to Z is equal to one point nine six contains ninety-five percent of the area under the curve or of the sample means.
- Then, the confidence interval is [y bar minus one point nine six into standard error of (y bar), y bar plus one point nine six into standard error of (y bar)]
- When the population standard deviation sigma is unknown and n is small say (less than thirty) then, the confidence interval is [y bar minus t alpha (n minus 1) into standard error of y bar, y bar plus t alpha (n minus 1) into standard error of y bar]
- When reporting a confidence interval, make sure you report both the interval and the confidence level. One without the other is meaningless

# 5. Example

Example:

An auditor is faced with a population of one thousand vouchers and wants to estimate the average value of the population. A sample of fifty vouchers is selected with average voucher amount of one thousand and seventy six point three nine dollars, standard deviation of two hundred and seventy three point six two dollars. Set up the ninety-five percent confidence interval estimate of the average amount for the population of vouchers.

Solution:
Given n is equal to fifty, sample mean y bar is equal to one thousand and seventy six point three nine dollars and s is equal to two hundred and seventy three point six two dollars
A ninety five percent CI for the population mean is
[y bar minus t alpha (n minus 1) into s by square root of n, y bar plus t alpha (n minus 1) into s by square root of n]
1 minus alpha is equal to point nine five, which implies alpha is equal to point zero five and since n is greater than thirty, t alpha (n minus 1) is equal to one point nine six.

By substituting the given values we get,
[one thousand seventy six point three nine minus (one point nine six into two hundred and seventy three point six two by root fifty, one thousand seventy six point three nine plus (one point nine six into two hundred and seventy three point six two by root fifty]
Which is equal to [one thousand and twenty six point three nine, one thousand one hundred and twenty six point three nine]
The ninety five percent confidence interval for the population average amount of the vouchers is between one thousand and twenty six point three nine and one thousand one hundred and twenty six point three nine.

Here’s a summary of our learning in this session, where we understood:

- The derivation of confidence limits for the unknown population when sigma is known
- The derivation of confidence limits for the unknown population when sigma is unknown
- The illustration of interval estimation of population mean by examples