1. Introduction

Welcome to the series of e-learning modules on Confidence Limits for the estimation of population mean, total and estimation of sample size. In this module we are going cover the basic concept of interval estimation, confidence coefficient, properties, confidence limits for population mean, total and estimation of sample size.

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

- Interval Estimation
- Confidence limits for population mean when variances are known and unknown
- Confidence limits for population total
- Estimation of sample sizes under Simple Random Sampling (SRS)

Interval estimates can be contrasted with point estimates.
A point estimate is a single value given as the estimate of a population parameter that is of interest. For example, the mean of some quantity.
A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

In statistics, a confidence interval is a kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate.
It is an observed interval, that is it is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest, if the experiment is repeated.
How frequently the observed interval contains the parameter is determined by the confidence level or confidence coefficient.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage the interval will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95, but we can produce 90 percent, 99 percent, 99.9 percent confidence intervals for the unknown parameter. The width of the confidence interval gives us an idea about how uncertain we are about the unknown parameter.A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.

For example, a confidence interval can be used to describe the reliability of survey results.

In an election poll of voting-intentions, the result might be that 40% of respondents intend to vote for a certain party.

A 90% confidence interval for the proportion in the whole population having the same intention on the survey date might be 38% to 42%.

From the same data one may calculate a 95% confidence interval, which might in this case be

36% to 44%.

A major factor determining the length of a confidence interval is the size of the sample used in the estimation procedure.

Here for instance is, the number of people taking part in the survey.

When applying standard statistical procedures, there will often be standard ways of constructing confidence intervals.

These are devised so as to meet certain desirable properties which will hold good, given the assumptions on which the procedure relies on are true.

The desirable properties may be described as Validity, Optimality and Variance.

- ‘Validity’ is most important property, followed closely by ‘Optimality’.
- ‘Invariance’ may be considered a property of the method of derivation of a confidence interval, rather than as a rule for constructing the interval.

- In case of non-standard applications, the same desirable properties would be sought.

2. Elements of Confidence Interval Estimation

Elements of Confidence Interval Estimation
Level of confidence

- Precision: Closeness to the unknown parameter
- Cost required to obtain a sample of size ‘n’

Confidence limits are defined as the the lower and upper boundaries or values of a confidence interval.
That is, they are the values which define the range of a confidence interval.

The confidence level is the probability value - 1 minus alpha, associated with a confidence interval.
It is often expressed as a percentage.
For example, say alpha equals to .05, equals to 5 percent.
Then the confidence level is equal to – 1 minus .05, equals to .95.
That is, a 95% confidence level.

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, A Producer may wishes to estimate his mean daily output;
A medical researcher may wish to estimate the mean response by patients to a new drug; etc.

The confidence interval size can be affected by certain factors including size of the sample, level of confidence and population variability.A larger sample size normally will lead to a better estimate of the population parameter

Now we shall look at calculating the confidence intervals for mean and known standard deviation.

First, we shall make the following three assumptions:

1. Population standard deviation is known
2. Population is normally distributed, and
3. If population is not normal, use large sample

Suppose Y1, Y2 etc Yn are the simple random samples drawn from a population of size ‘N’ with mean Y bar, and variance sigma square.

We know that,

Population mean Y bar 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 Y bar divided by the Standard Error,SE of y bar, which follows Normal distribution N, with mean zero and variance one.

We can always find two quantities: minus Z alpha by two & 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 Y bar divided 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 Y bar, 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 Ybar, 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 Y bar 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.

Under simple random sampling with replacement, 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 the population mean Y bar 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

Under simple random sampling without replacement, Standard Error of y bar equals to Square root of N minus n into S square by N into n.
Under simple random sampling without replacement, 95% Confidence Interval for the

population mean is:
y bar minus 1.96 into square root of N minus n into S square by N into n, y bar plus 1.96 into square root of N minus n into S square by N into n

In a certain problem, the student calculated the sample mean of the boiling temperatures to be 101.82, with its standard deviation 0.49. The critical value for a 95% confidence interval is 1.96.

A 95% confidence interval for the unknown mean Y bar is:101.82 minus 1.96 into 0.49, 101.82 plus 1.96 into 0.49,Equals to 100.86, 202.78.

As the level of confidence decreases, the size of the corresponding interval also decreases.

Suppose a student was interested in a 90 percent confidence interval for the boiling temperature, then in this case, 1 minus alpha is equal to 0.90 and alpha by 2 is equal to 0.05.

The critical value Z for this level is equal to 1.645, so the 90 percent confidence interval is:
101.82 minus 1.645 into 0.49,101.82 plus 1.645 into 0.49,
Equals to 101.01, 102.63.

3. Confidence Intervals for Unknown Mean and Standard Deviation

Next, here we shall look at confidence intervals for unknown mean and standard deviation. The three assumptions we will make in this case are:

1. Population standard deviation is unknown
2. Population is normally distributed, and
3. If population is not normal, use student's 't' distribution.

In most practical research, the standard deviation for the population of interest is not known. In such cases, the standard deviation sigma is replaced by an estimate standard deviation 's' also known as the Standard Error.
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 mue and standard deviation

Instead, the sample mean follows the *t distribution* with mean Y bar and standard deviation s by square root of 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 a large sample size.

For a population with unknown mean Y bar and unknown standard deviation, a confidence interval for the population mean, based on a simple random sampling with replacement of size *n*, is:
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.
Under simple random sampling without replacement,
y bar minus t alpha, n minus 1 into square root of N minus n into S square by N into n, y bar plus t alpha, n minus 1 into square root of N minus n into S square by N into n.

An estimate of population total is N into sample mean y bar

Suppose y1, y2 etc y-n are the simple random samples drawn from a population of size N with mean Y bar and variance sigma square, we know that:

Sample mean y bar is equal to summation yi by n

Z is equal to Y cap minus Y divided by Standard Error of Y cap 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 cap minus Y divided by Standard Error of Y cap, 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 cap less than or equal to y cap minus Y, less than or equal to Z alpha by two into Standard error of Y cap is equal to 1 minus alpha

Probability of Y cap minus Z alpha by two into Standard Error of Y cap less than or equal to Y less than or equal to Y cap plus Z alpha by two into Standard error of Y cap is equal to 1 minus alpha

Therefore, Hundred into one minus alpha percent Confidence Interval for the population total when the variance is known as sigma square is given by:

N into ybar minus Z alpha by two into Standard Error of Y cap, N into y bar plus Z alpha by two into Standard error of Y cap.

Under Simple Random Sampling With Replacement, Standard error of Y cap equals to square root of N square into sigma square by n.

95 percent Confidence Interval for the population total is:

N into ybar minus 1.96 into square root of N square into sigma square by n, N into y bar plus 1.96 into square root of N square into sigma square by n.

Under Simple Random Sampling Without Replacement, Standard Error of Y cap equals to Square root of N square into N minus n into S square by N into n

95 percent Confidence Interval for the population total is N into ybar minus 1.96 into square root of N square into N minus n into S square by N into n N into ybar plus 1.96 into square root of N square into N minus n, into S square by N into n.

4. Size of the Simple Random Sample

Size of the Simple Random Sample:
Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample.
In order to estimate the population parameters with a specified degree of precision one has to take decisions about the sample size.

Now suppose it is desired to estimate the mean of a population, given a confidence interval.

One of the first questions that arise is, how large should the sample be? This question must be given serious consideration, because it is a waste of time and resource to take a larger sample than what is needed, to achieve the desired result.

Similarly, a sample that is too small may lead to results that are of no practical value.

Three questions must be answered to determine the sample size:
1. Standard deviation of the population:
It is rare that a researcher knows the exact standard deviation of the population. Typically, the standard deviation of the population is estimated, a - From the results of a previous survey; b - From a pilot study; c - From secondary data, or d - The judgment of the researcher.

The next question is about the maximum acceptable difference.
This is the maximum amount of error that you are willing to accept.
And finally the third: Desired percentage confidence level.
The confidence level is the level of certainty that the sample mean does not differ from the true population mean by more than the maximum acceptable difference.

5. Degree of Precision

The degree of precision is usually determined in terms of :

- The margin of error permissible in the estimate
- The confidence coefficient with which we want this estimate to lie within the permissible margin of error

If Y bar denotes the mean of a population observation of size N and y bar denotes the mean of the sample observations of size n,
we know that y bar is an unbiased estimate of population mean Y bar.
If the permissible error in estimating Y bar is e, and the confidence coefficient is 1 minus alpha, then the sample size n is determined by the equation:
Probability of modulus of y bar minus Y bar less than or equal to e is equal to 1 minus alpha. Call this as (1)
This implies, Probability of modulus of y bar minus Y bar greater than e is equal to alpha. Now call this as (2)
Where alpha is the level of significance.

If n is sufficiently large and we consider a simple random sample without replacement, then the statistic,
Z is equal to y bar minus Y bar divided by S into square root of N minus n by N into n which follows Normal distribution with mean zero and variance one.
Accordingly if we take alpha equals to zero point zero five, then we have,
Probability of modulus ybar minus Y bar divided by S into square root of N minus n by N into n greater than 1.96 is equal to zero point zero five.
Comparing this with equation (2) we get
e is equal to 1.96 into S into square root of N minus n divided by N into n.

Which implies n is equal to N into S square into 1.96 square whole divided by N into e square plus 1.96 square into S square.
The formula gives the sample size in simple random sample without replacement for estimating Y bar, with the confidence coefficient 95 percent and the margin of error ‘e’, provided n is large.

Now, if n is small, the statistic is defined as:
Student’s t distribution with n minus 1 degrees of freedom,
If alpha is the level of significance and t alpha is the significant value for n minus 1 degrees of freedom then n is given by the equation:
n equals to N into t square into S square whole divided by N into e square plus t square into S square
which implies n equals to t square into S square whole divided by e square plus t square into S square by N
Under simple random sampling with replacement:
n equals to 1.96 square into sigma square by e square.

Here's a summary of our learning in this session:

- Interval estimation process
- Confidence interval estimation for the mean (σ known and unknown) under SRS
- Discussed confidence interval estimation for the total (σ known and unknown) under SRS
- Determining sample size