1. Introduction

Welcome to the series of E-learning modules on Interval Estimation. In this module, we are going to cover the basic concept of interval estimation, confidence limits, coefficient, properties and cautions to be taken in interval estimation.

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

- Explain about interval estimation
- Explain the basic principle behind the technique
- Explain the confidence limits, level and coefficient
- Explain the central confidence intervals
- Explain the cautions to be taken in the process of interval estimation

In a decision making process, the parameters of interest can be estimated either using point estimation or interval estimation techniques.
Let us consider a population with probability density function or probability mass function f of (x, theta), where the parameter theta may be real valued or vector valued. The methods of estimation discussed so far are called point estimation methods because a single point estimates the parameter.

In general, an estimator cannot be expected to coincide with the actual value of the parameter. In fact, the estimator in general has a continuous distribution and, the probability that it is equal to a particular value is zero. It is therefore desirable to give an interval in which the population parameter may be expected to lie with a specified degree of confidence.

Such a kind of estimation where an interval expected to contain the true value of the parameter is obtained is called interval estimation. However, a point estimate is only as good as the sample it represents. If other random samples are taken from the population, the point estimates derived from those samples are likely to vary.
Because of variation in sample statistics, estimating a population parameter with a confidence interval is often preferable to use a point estimate.

Thus, 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 (CI) 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 (i.e. it is calculated from the observations) and 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.

More specifically, the meaning of the term "confidence level" is that, if confidence intervals are constructed across many separate data analyses of repeated (and possibly different) experiments, the proportion of such intervals that contain the true value of the parameter will approximately match the confidence level. This is guaranteed by the reasoning underlying the construction of confidence intervals.

The lower and upper limits of the confidence intervals are called confidence limits. The amount of confidence (in terms of probability) with which a parameter is expected to lie in the confidence interval is known as confidence coefficient.
In interval estimation, if we say that we obtained ninety-five percent confidence interval, it means that if a large number of samples is taken and if the confidence interval for each of the sample is obtained, then in ninety five percent of the cases the parameter values can be expected to lie in the confidence interval obtained.

Hence, if independent samples are taken repeatedly from the same population, and a confidence interval is calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated because this percentage is ninety-five percent. However, we can produce ninety percent, ninety eight percent, ninety nine percent, ninety nine point nine percent (or whatever) confidence intervals for the unknown parameter.

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 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.
- An alternate method: multiply the confidence interval, ninety five percent by half (since the distribution is symmetric and the intervals are equal on each side of the population mean).

The width of the confidence interval gives us some 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 how reliable survey results are. In a poll of election voting-intentions, the result might be that forty percent 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 thirty eight percent to forty two percent.

From the same data, one may calculate a ninety-five percent confidence interval, which might in this case be thirty six percent to forty four percent. A major factor determining the length of a confidence interval is the size of the sample used in the estimation procedure. For example, the number of people taking part in a survey.

2. Principle of Interval Estimation

Principle of Interval Estimation
Consider a distribution with p.d.f f of (x, theta), where theta is an unknown parameter. Our problem is to find a confidence interval for the parameter theta. In an interval estimation problem of finding confidence interval for the parameter theta with certain amount of confidence (1 minus alpha), we need to find two quantities A and B based on the sample observations such that
Probability of [A less than theta less than B] is equal to 1 minus alpha. Call this as (1)

In general, there may be many sets of values for A and B, which satisfy the equation (1). However, A and B are found such that they are at equal distance from the true value of the parameter theta. Intervals so obtained are called central intervals and the others non-central. If A and B are numbers such that P of [theta less than A] is equal to 1 minus alpha and P of [theta greater than B] is equal to 1 minus beta then A and B are called one sided upper confidence limit for theta with confidence coefficient (1 minus alpha) and one sided lower confidence limit with confidence coefficient (1 minus beta).

Central Confidence Interval
Let us consider a population with probability density function or probability mass function f of (x, theta) where theta is the parameter. Let x one, x two... xn be a sample of size n taken from the above population. Let A and B be two statistics.
If P of [A greater than theta] is equal to alpha one by two and P of [B less than theta] is equal to alpha two by two such that P of [A less than theta less than B] is equal to 1 minus alpha where alpha is equal to alpha one by two plus alpha two by two, confidence intervals of this type are called central confidence intervals.

Example:
- A business analyst for Cellular Telephone Company takes a random sample of eighty-five bills for a recent month and from these bills, it computes a sample mean of one hundred and fifty three minutes. If the company uses the sample mean of one hundred and fifty three minutes as an estimate for the population mean, then the sample mean is being used as a Point Estimate. History and similar studies indicate that the population standard deviation is forty-six minutes.

- The value of Z is decided by the level of confidence desired. A confidence level of ninety five percent has been selected
- The confidence interval is constructed from the point estimate, one hundred and fifty three minutes, and the margin of error of this estimate, plus or minus nine point seven eight minutes
- The resulting confidence interval is one hundred forty three point two two less than or equal to mu less than or equal to one sixty two point seven eight
- The cellular telephone company business analyst is ninety five percent confident that the average length of a call for the population is between one hundred forty three point two two and one sixty two point seven eight minutes

Interpreting C.I

For the above ninety five percent confidence interval, the following conclusions are valid:

- I am ninety five percent confident that the average length of a call for the population mu, lies between one hundred forty three point two two and one sixty two point seven eight minutes
- If I repeatedly obtained samples of size eighty-five, then ninety five percent of the resulting confidence intervals would contain μ and five percent would not

But be Careful! The following statement is NOT true:

“The probability that mu lies between one hundred forty three point two two *and* one sixty two point seven eight *is point nine five”.*

Once you have inserted your sample results into the confidence interval formula, the word Probability can no longer be used to describe the resulting confidence interval.

3. Desirable Properties

Desirable properties

When applying standard statistical procedures, there will often be standard ways of constructing confidence intervals. These will have been devised so as to meet certain desirable properties, which will hold given that the assumptions on which the procedure relies are true.

These desirable properties may be described as validity, optimality and invariance.

Of these, "validity" is most important, followed closely by "optimality". "Invariance" may be considered as a property of the method of derivation of a confidence interval rather than of the rule for constructing the interval. In non-standard applications, the same desirable properties would be sought.

- Validity: This means that the nominal coverage probability (confidence level) of the confidence interval should hold, either exactly or to a good approximation
- Optimality*:* This means that the rule for constructing the confidence interval should make as much use of the information in the data set as possible. Recall that one could throw away half of a dataset and still be able to derive a valid confidence interval. One way of assessing optimality is by the length of the interval. So that a rule for constructing a confidence interval is judged better than another if it leads to intervals whose lengths are typically shorter

- Invariance: In many applications, the quantity being estimated might not be tightly defined as such. For example, a survey might result in an estimate of the median income in a population, but it might equally be considered as providing an estimate of the logarithm of the median income, given that this is a common scale for presenting graphical results

It would be desirable that the method used for constructing a confidence interval for the median income would give equivalent results when applied to constructing a confidence interval for the logarithm of the median income. Specifically, the values at the ends of the latter interval would be the logarithms of the values at the ends of former interval.

4. 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
Confidence limits are the lower and upper boundaries or values of a confidence interval, that is, the values that define the range of a confidence interval.

Confidence Level
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 is equal to point zero five, which is equal to five percent. Then, the confidence level is equal to (1 minus point zero five) which is equal to point nine five, that is a ninety five percent confidence level.

Constructing Interval estimates of a parameter

- The general form for the interval estimate of a population parameter is point estimate of parameter plus or minus Margin of error
- The margin of error is an amount that is added to and subtracted from the point estimate of a statistic to produce an interval estimate of the parameter

- The size of the margin of error depends on:
  - The type of sampling distribution for the sample statistic
  - The percentage of the area under the sampling distribution that a researcher decides to include – usually ninety percent, ninety five percent or ninety nine percent. This is termed a confidence level
- Each interval estimate is an interval constructed around the point estimate along with a confidence level

Confidence Interval for a Mean
A confidence interval for a mean specifies a range of values within which the unknown population parameter may lie (in this case the mean). These intervals may be calculated as for example, a producer who wishes to estimate his mean daily output, a medical researcher who wishes to estimate the mean response by patients to a new drug, etc.

5. Steps to Obtain Confidence Interval for Mean

Steps to obtain confidence interval for mean

- Obtain the point estimate of mu, that is, the sample mean
- Determine the distribution of the sample mean. If *n* is large, then the Central Limit Theorem can be used and x 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 [x bar minus one point nine six into standard error of x bar, x bar plus one point nine six into standard error of x bar]
- When reporting a confidence interval, make sure you report both the interval and the confidence level. One without the other is meaningless

Determination of sigma

- In order to construct an interval estimate, it is necessary to obtain some estimate of sigma, the variability of the population from which the sample is drawn. This is required to obtain an estimate of the standard error of the sample mean
- Generally, the sample standard deviation *s* is used as an estimate of sigma. For large sample size, assume the CLT holds and assumes a reasonable estimate of sigma. For a small sample, where *n* is less than thirty, the *t*-distribution should be used, again using *s* as an estimate of sigma

- In practice, sigma is rarely known. In addition, as *n* increases, the *t*-distribution approaches the normal distribution. Thus, as long as n is greater than thirty, it is acceptable to use *s* as an estimate of sigma for purposes of constructing an interval estimate
- Certain factors may affect the confidence interval size including size of sample, level of confidence, and population variability. A larger sample size normally will lead to a better estimate of the population parameter

- A confidence interval does not predict that the true value of the parameter has a particular probability of being in the confidence interval given the data actually obtained

(An interval intended to have such a property, called a credible interval, can be estimated using Bayesian methods. However, such methods bring with them their own distinct strengths and weaknesses)

Selecting a confidence level

- There is no one confidence level that is appropriate for all circumstances
- Greater confidence level means greater certainty that the interval estimate of mu actually contains mu. However, for ninety nine percent or ninety nine point nine percent

confidence level, the interval may be very wide

- Smaller confidence levels (Example: eighty percent or ninety percent) produce smaller margins of error and seemingly more precise interval estimates, but they are less likely to contain mu

- Use the level requested or the level others have used when researching similar issues
- By tradition, the default level is ninety-five percent
- Issues such as manufacturing products have to be safe for human use. For example, foods should require high confidence levels (ninety nine point nine percent plus). However, this may increase costs of manufacture and checking for safety.

Cautions about interval estimates

- There are many assumptions involved in interval estimation:
  - The sample is randomly selected from a population
  - The sample size is sufficiently large to use the central limit theorem
  - The population standard deviation is known or *s* is a good estimate of sigma
  - The selection of a confidence level is an arbitrary process
  - The population is not too skewed

- As a result, interval estimates are not precise, but are estimates or approximations
- Larger n, repeated sampling, comparisons with other studies, careful sampling and survey design and practice can improve the quality of the estimates

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

- The principle of interval estimation process
- The basic concepts such as confidence limits, confidence coefficients and central confidence intervals
- The procedure for interval estimation and its properties
- The cautions about interval estimates