1. Introduction

Welcome to the series of E-learning modules on Estimator and Estimate. In this module we are going cover the basic concepts of Estimator and Estimate, Role of Estimators and estimates, properties, Types of estimators or estimates, Biased estimates and standard error.

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

- About an estimator and an estimate
- Types of estimates or estimators
- Properties of estimators or estimates
- Biased estimates and standard error

It would not be unusual to see in a daily news paper that "a recent poll of one thousand five hundred randomly chosen Americans indicates that twenty two percent of the entire U.S population is presently dieting with a margin of error plus or minus 2 percent". Perhaps you have wondered about such claims. For instance what exactly does "with a margin of error of plus or minus 2 percent" mean?

Also how is it possible, in a nation of over one hundred and fifty million adults that the proportion of them presently on diets can be ascertained by a sampling of only one thousand five hundred people? In this topic we will find answers to these questions. In general we will consider how one can learn about the numerical characteristics of a population by analyzing results from a sample of this population

Whereas the numerical values of the members of the population can be summarized by a population probability distribution, this distribution is often completely not known. For instance: certain of its parameters such as its mean and its standard deviation may be unknown. A fundamental concern in Statistics relates to how one can use the results from a sample of the population to estimate these unknown parameters.

For instance: If the items of the population consist of newly manufactured computer chips, then we may be interested in learning about the average functional lifetime of these chips. That is we would be interested in estimating certain parameters of the population distribution. To accomplish this we will show how to use estimators and the estimates they give rise to.

One of the objectives of drawing samples from a population is to extract maximum information about the population based on a sample from the population and to get the best estimates of the parameters of the population. When the population is very large or when it is impossible to obtain the characteristics of the population we try to estimate the parameters using sampling theory such that the estimated value of the parameter lies almost closer to the actual value of the parameter.

An estimator of a population parameter is a random variable that depends on the sample information and whose realizations provide approximations to the unknown parameter of the population. A specific realization of that random variable is called an estimator and estimate.

An estimator is a statistic whose value depends on the particular sample drawn. The value of the estimator called the estimate is used to predict the value of a population parameter.

For instance: If we want to estimate the mean life time of a chip, then we would employ the sample mean as an estimator of the population mean. If the value of the sample mean were hundred and twenty three hours then the estimate of the population mean would be hundred and twenty three hours.

It seems reasonable to base our conclusion on the sample mean life time so we say that the estimator of the population mean is the sample mean and let us suppose that the sample mean is x bar. Naturally x bar is the estimator and the value of this sample mean is the estimate of the population mean (mu).

Suppose we have a random sample x one, x two, etc, x_n on a variable X whose distribution in the population involves an unknown parameter theta. It is required to find an estimate of theta on the basis of sample values.

2. Estimation and Estimate

Estimation
Estimation is the process by which sample data are used to indicate the value of an unknown quantity in a population. Results of estimation can be expressed as a single value, known as a point estimate, or a range of values, known as a confidence interval.

What is an estimate?
When we have observed a specific value of our estimator we call that value an estimate. In other words an estimate is a specific observed value of a statistic. We form an estimate by taking a sample and computing the value taken by our estimator in that sample. An estimate is an indication of the value of an unknown quantity based on observed data.

More formally, an estimate is the particular value of an estimator that is obtained from a particular sample of data and used to indicate the value of a parameter.

Example
1) Suppose the manager of a shop wanted to know the mean expenditure of customers in his shop in the last year. He could calculate the average expenditure of the hundreds (or perhaps thousands) of customers who bought goods in his shop, that is, the population mean. Instead he could use an estimate of this population mean by calculating the mean of a representative sample of customers. If this value was found to be twenty five dollars, then twenty five dollars would be his estimate.

2) Suppose that we calculate the mean odometer reading (mileage) from a sample of used taxis and find it to be Ninety eight thousand miles. If we use this specific value to estimate the mileage for a whole fleet of used taxis, the value Ninety eight thousand would be an estimate.

3. Types of Estimates

Types of Estimates:
An estimate of the population parameters can be expressed in two ways:

1) Point Estimate and 2) Interval Estimate

Point estimate.
A point estimate of a population parameter is a single value of a statistic. For example, the value of the sample mean **y bar is** a point estimate of the population mean mu. Similarly, the value of the sample proportion ***p*** is a point estimate of the population proportion ***P***.

A point estimate is a single number that is used to estimate an unknown parameter. If while watching the first members of a football team come into a field you say “Why, I bet their line must average two hundred and fifty pounds” You have made a point estimate. A department head would make a point estimate if he said “Our current data indicate that this course will have three hundred and fifty pounds students in the fall”

A point estimate is often insufficient, because it is either right or wrong. If you are told only that his point estimate of enrolment is wrong, you do not know how wrong it is, and you cannot be certain of the estimates reliability. If you learn that it is off by only ten students you would accept three hundred and fifty students as a good estimate of future enrolment. Therefore point estimate is much more useful if it is accompanied by an estimate of the error that might be involved.
Point estimator of a population parameter is a function of the sample information that yields a single number. The corresponding realization is called the point estimate of the parameter

Notation for the parameters , point estimators and estimates
Population parameter Mean (mu,) - estimator is x bar - estimate is x bar
Parameter Variance (sigma square)- estimator is s x square and estimate is s square
For Standard deviation (sigma) it is s_x and s
For Proportion (P or π) it is p_x and p

The term **estimator** refers to the formula or expression used to calculate the **estimate**, the actual numerical value estimate of the population parameter in a particular problem. When we speak generically, it is conventional to represent the population parameter being estimated by the symbol theta, the Greek letter 'theta', and to represent the estimator by the symbol theta cap, the same Greek letter, but with a caret on top.

Interval estimate.
Assume that we have a sample (x one, x two etc xn) from a given population. All parameters of the population are known except some parameter theta.
We want to determine from the given observations unknown parameter - theta.
In other words we want to determine a number or range of numbers from the observations that can be taken as a value of an unknown parameter theta.

An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, **“**a” **less than “**x” less than **“**b” is an interval estimate of the population mean μ. It indicates that the population mean is greater than **“**a” **but** less than “**b”**.

An interval estimate is a range of value used to estimate a population parameter. It indicates an error in two ways.
1) By the extent of its range and
2) The probability of the true population parameter lying within the range.
In general, point estimation should be contrasted with interval estimation. Such interval estimates are typically either confidence intervals in the case of frequentist inference, or credible intervals in the case of Bayesian inference.
However, because of the presence of sampling error, sometimes it is more useful to start with the point estimate, and then establish a range of values both above and below the point estimate. Next, by using the probability-numbers characteristic of normally distributed variables, we can state the level of confidence we have that the actual population mean will fall somewhere in our range. This process is known as "constructing a confidence interval”

Problem of statistics is not to find estimates but to find estimators. Estimator is not rejected because it gives one bad result for one sample. It is rejected when it gives bad results in a long run. That is, it gives bad result for many, many samples. Estimator is accepted or rejected depending on its sampling properties. Estimator is judged by the properties of the distribution of estimates it gives rise.

Since estimator gives rise to an estimate that depends on sample points (x one, x two, etc, xn) estimate is a function of sample points. Sample points are random variable therefore estimate is random variable and has probability distribution. We want that estimator to have four desirable properties like consistency, unbiasedness, efficiency and sufficiency

4. Properties of Estimators

Properties of **Estimators**
1 Unbiased - The expected value (mean) of the estimator's sampling distribution is equal to the underlying population parameter; that is, there is no upward or downward bias.
2. Efficiency - While there are many unbiased estimators of the same parameter, the most efficient has a sampling distribution with the smallest variance. Good estimator has smaller standard error than other estimators. And it falls closer than other estimators to parameter. Sample mean has a smaller standard error than sample median

3. Consistency - Larger sample sizes tend to produce more accurate estimates; that is, the sample parameter converges on the population parameter. An estimator is said to be consistent if the variance of its sampling distribution decreases with increasing sample size. This is a good property because it means that if you make the effort to collect data from a larger random sample, you should end up with a more accurate estimate of the population parameter.

4. Sufficiency: If an estimate consists of sufficient information about the population parameter being estimated then estimate is said to be sufficient. Sufficient estimates are obtained by making so much of use of information from the sample that no other estimate could contain the additional information from the sample about the population parameter being estimated.
How well a specific estimator satisfies each of these criteria depends very much on the details of the population distribution.

Parameter Estimates – Fit
Methods that enable us to estimate with reasonable accuracy the population proportion (the proportion of the population that possesses the given characteristic) and the population mean. To calculate the exact proportion or the exact mean would be an impossible goal.

Even so, we will be able to make an estimate , make a statement about the error and will probably accompany this estimate and implement some controls to avoid as much of the errors as possible. As decision makers we will be forced to at times to rely on blind hunches.
In practice, we would prefer to have one estimator (theta cap that is always the best.
However, theta cap is a function of the observed xi's so that actual value of the parameter Theta is equal to theta cap plus estimation error

Thus, we may identify the "best estimator" as the one with:
Least bias (unbiased)
Minimum variance of estimation error (increase the likelihood that the observed parameter estimate represents the true parameter)
In practice, confidence interval estimates are used more commonly by far than point estimates. Nevertheless, since point estimates are used in certain important ways in statistics, and carry with them some important concepts and terms, we need to look at them briefly.

5. Standard Error

Standard Error
Incidentally, because of it being a measure of the scale of error in an estimator, the standard deviation of a sampling distribution is often referred to as the **standard error of the estimate**. The variability of a statistic is measured by the Standard Error of Sample Estimators.

The standard error is computed from known sample statistics, and it provides an unbiased estimate of the standard deviation of the estimator.

Both the precision and credibility of the estimates improves with the increasing quality and quantity of the sample.
The values of population parameters are often unknown, making it impossible to compute the standard deviation of a statistic. When this occurs, we use the standard error.

The care that the statisticians must take in picking an estimator

A given sample statistic is not always the best estimator of its analogous population parameter. Consider a symmetrically distributed population in which the values of the median and then mean coincide. In this instance a sample mean would be an unbiased estimator of a population median also a consistent estimator of a population median. Because as the sample size increase the value of the sample mean would tend to come very close to the population median.

And sample mean would be more efficient estimator of the population median than the sample median itself because in large samples the sample mean has a smaller standard error than the sample median. At the same time the sample median in a symmetrically distributed population would be an unbiased and consistent estimator of the population mean but not the most efficient estimator because in large samples its standard error is larger than that of the sample mean.

<u>Bias and its effects</u>
In sample survey theory it is necessary to consider biased estimators for two reasons
1. In some of the most common problems , particularly in the estimation of ratios, estimators that are otherwise convenient and suitable are found to be biased
2. Even with estimators that are unbiased in probability sampling, errors of measurement and nonresponsive may produce biases in the numbers that we are able to compute from the data.

Bias is a term which refers to how far the average statistic lies from the parameter it is estimating, that is, the error which arises when estimating a quantity. Errors from chance will cancel each other out in the long run, those from bias will not.

These are the Systematic errors produced by our sampling procedure. For example, if you sample people and ask them whether they watched Olympic games, but the percentage always comes out too high (maybe because you have interviewed your friends and your whole group really likes sports and games)

The Mean Square Error

In order to compare an biased estimator with an unbiased estimator or two estimators with different amounts of bias, a useful criterion is the Mean Square Error (M.S.E) of the estimator measured from the population value that is being estimated which gives an accuracy of an estimator

M.S.E of theta cap = Expected value of theta cap minus theta) whole square

Sampling Distribution:

It is possible to draw more than one sample from the same population and the value of an estimator will in general vary from sample to sample. For example, the average value in a sample is a statistic. The average values in more than one sample, drawn from the same population, will not necessarily be equal.

The sampling distribution describes probabilities associated with an estimator when a random sample is drawn from a population.The sampling distribution is the probability distribution or probability density function of an estimator.

A procedure of “guessing” properties of the population from which data are collected is known as a statistical estimation and an estimate of an unknown parameter is a value that represents a “guess” of the properties of the population. Ultimate aim of Statistical Estimation is to provide a best estimate for the population under study satisfying almost all the properties of a good estimates without any bias .

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

- Statistical definition of an estimator and an estimate
- Types of estimators or estimates
- Difference between point and interval estimates (estimators)
- Properties of estimates or estimators
- Biased estimates and the standard error