BLOCK I:
SAMPLING DISTRIBUTION AND THEORY OF ESTIMATION

Unit-1
Sampling Distribution

Unit Structure:

1.1 Introduction

In case of any statistical investigation, our interest lies in studying the various characteristics of a particular collection of objects or observations usually called the target population, simply population or universe. By definition, the collection of all the observations under study in any statistical investigation, is called population or universe for that specific study. The number of observations included in a population is termed as the size of the population or population size.

Again, a subcollection of the population is known as sample. In other words, a sample may be difined as a part of a population so selected with a view to represent the population. The number of units in a sample is called sample size and the units forming the sample are known as "Sampling Units". Again, a detailed and complete list of all the sampling units is termed as a "Sampling Frame". It is a must to have an updated sampling frame complete in all respects before the samples are actually drawn.

Any statistical characteristics such as mean, median, quartile, standard deviation, moments etc. of the population under study, is called a parameter, while any statistical characteristics such as above of a sample drawn from a population is called a statistic. Very often, the values of various parameters are unknown and these are estimated by the corresponding statistic. For example, sample mean $\bar{x}$ is used as an estimator of population mean μ, sample standard deviations is used as an estimator of population standard deviation σ, etc. The difference between a statistic and the corresponding parameter is known as sampling error. The study of Sampling theory as well as theory of estimation help us to estimate the true value of the population parameters by minimizing the sampling errors.

1.2 Objectives

After going through thid unit, you will be able to-

- know the concept of sampling distribution;
- explain the methods of estimation;
- discuss the technique of solving practical problems.

1.3 Sampling Fluctuations

The value of parameter is considered as constant. But, if we compute the value of a statistic, say mean or median or mode or s.d., etc, it is quite natural that the value of the sample statistic may vary from sample to sample as the sampling units of one sample may be different from that of another sample even if the sample sizes are same. The variation in the values of a statistic from sample to sample is termed as "Sampling Fluctuations" or "Sampling Variation".

1.4 Sampling Distribution of a Statistic

If it is possible to obtain the values of a statistic (t) from all the possible samples of a fixed sample size along with the corresponding probabilities, then we can arrange the values of the statistic, which is to be treated as a random variable, in the form of a probability distribution. Such a probability distribution is known as the sampling distribution of the statistic.

Starting with a population of N units, we can draw many a sample of a fixed size n. In case of sampling with replacement, the total number of samples that can be drawn is N^n and consequently we shall get N^n different values of any statistic (t) like mean, median, S.D. etc. computed for N^n samples. Again, when sampling is done without replacement of the sampling units, the total number of samples that can be drawn is $N_{C_n = m}$ (say). We can compute any statistic (t) like mean, median, S,d. etc. for these m samples resulting in m values of the statistic. These N^n values of the statistic (t) in case of sampling with replacement and $N_{C_n = m}$ values in case of sampling without replacement may be arranged in the form of a probability distribution known as the sampling distribution of the statistic.

The sampling distribution, just like a theoretical probability distribution possess different properties. One of these is the 'Law of Large Numbers' which asserts that a positive integer n can be determined such that if a random sample of size n or large is drawn from a population having mean μ, the probability that the sample mean $\bar{x}$ will deviate from μ by less than any arbitrarily small quantity can be made to be as close to 1. This implies that a fairly reliable inference can be made about an infinite population by taking only a finite sample of sufficiently large size. Another interesting result in this connection is the 'Central Limit theorem' which is discussed elaborately in the Unit 2.

Check Your Progress

1. What do you mean by Sampling Fluctuations?
2. What is 'Law of Large Numbers'?
3. Distinguish between parameter and statistic.

1.5 Standard Error of a Statistic

The standard deviation of a statistic is termed as standard error. We know that, the population standard deviation describes the variation among values of members of the population, whereas the standard deviation of sampling distribution measures the variability among the values of the statistic (such as mean values, median values, etc) due to sampling errors. Thus knowledge of sampling distribution of a statistic enables us to find the probability of sampling error of the given magnitude. Consequently standard deviation of sampling distribution of a sample statistic measures sampling error and is also known as standard error of the statistic. If t be any statistic calculated for different samples, then the standard error of the statistic t is generally denoted by S.E. (t).

The S.E.(t) measures not only the amount of chance error in the sampling process but also the accuracy desired in estimation of population parameters. Some of the common results of standard error of different statistic are given below :

1. $S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ (*sample drawn with replacement*),

$$S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}}\left(\frac{N-n}{N-1}\right)(sample - drawn - without - replacement)$$

2. $S.E.(p) = \sqrt{\frac{PQ}{n}}$ (*sample drawn with replacement*),

$$S.E.(p) = \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}(sample\ drawn\ without\ replacement)$$

3. $S.E.(s) = \frac{\sigma}{\sqrt{2n}}$

4. $S.E.(sample\ median) = \sqrt{\frac{\pi}{2n}}\quad \sigma = \frac{1.125332\sigma}{\sqrt{n}}$

$(\therefore \pi = 3.1416)$

5. $S.E.(r) = \frac{1-y^2}{\sqrt{n}}$

6. $S.E.(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

7. $S.E.(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

8. $S.E.(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$

Example 1. A population comprises 3 members 1,5,3. Draw all possible samples of size two

i) with replacement

ii) without replacement

Find the sampling distribution of sample mean in both cases.

Solution : i) with replacement:

since n=2 and N=3, the total number of possible sample of size 2 with replacement = 3^2 =9. These are exhibited along with the corresponding sample mean in the following table :

Sl. No.	Sample	Sample Mean ($\bar{x}$)
1	1, 1	1
2	1, 5	3
3	1, 3	2
4	5, 1	3
5	5, 5	5
6	5, 3	4
7	3, 1	2
8	3, 5	4
9	3, 3	3

This sampling distribution of the sample mean is given as follows:

$\bar{x}$	1	2	3	4	5	Total
p :	1/9	2/9	3/9	2/9	1/9	1

(ii) Without replacement :

As N = 3 and n = 2, the total number of possible samples without replacement = ${}^{N}C_{n} = {}^{3}C_{2} = 3$. Possible samples of size 2 and corresponding sample means are given below :

Serial No.	Sample	Sample Mean ($\bar{x}$)
1	1, 3	2
2	1, 5	3
3	3, 5	4

The sampling distribution of the sample mean is given as follows:

$\bar{x}$:	2	3	4	Total
p :	1/3	1/3	1/3	1

Example : Compute the standard deviation of sample mean for the last problem. Obtain the SE of sample mean and show that they are equal.

Solution : We consider the following cases :

(i) With replacement :

Let u = $\bar{x}$. The sampling distribution of u is given by

u :	1	2	3	4	5
p :	1/9	2/9	3/9	2/9	1/9

$$\therefore E(u) = \sum p_i \, u_i$$

$$= \frac{1}{9}\times 1 + \frac{2}{9}\times 2 + \frac{3}{9}\times 3 + \frac{2}{9}\times 4 + \frac{1}{9}\times 5$$

$$= 3$$

$$\therefore E(u^2) = \sum p_i u_i^{\,2}$$

$$= \frac{1}{9}\times 1^2 + \frac{2}{9}\times 2^2 + \frac{3}{9}\times 3^2 + \frac{2}{9}\times 4^2 + \frac{1}{9}\times 5^2$$

$$= \frac{31}{3}$$

$$\therefore V(u) = E(u^2) - [E(u)]^2$$

$$= \frac{31}{3} - 3^2 = \frac{4}{3}$$

$$\text{Hence } SE_{\bar{x}} = \frac{2}{\sqrt{3}} \qquad \longrightarrow (1)$$

Again, the population mean (μ) is given by

$$\mu = \frac{1+5+3}{3} = 3$$

and the population variance (σ^2) is given by

$$\sigma^2 = \frac{1}{N}\sum (x-\mu)^2$$

$$= \frac{1}{3}\left[(1-3)^2 + (5-3)^2 + (3-3)^2\right] = \frac{8}{3}$$

$$\therefore SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{8}{3}} \times \frac{1}{\sqrt{2}} = \frac{2}{\sqrt{3}} \qquad \longrightarrow (2)$$

Thus comparing (1) and (2), we are able to verify the validity of the formula.

<u>(ii) Without replacement :</u>

In this case, the sampling distribution of $v = \bar{x}$ is given by

v :	2	3	4
p :	1/3	1/3	1/3

$$\therefore E(\bar{x}) = E(v) = \frac{1}{3}\times 2 + \frac{1}{3}\times 3 + \frac{1}{3}\times 4$$

$$= 3$$

$$V(\bar{x}) = Var(v) = E(v^2) - [E(v)]^2$$

$$= \frac{1}{3}\times 2^2 + \frac{1}{3}\times 3^2 + \frac{1}{3}\times 4^2 - 3^2$$

$$= \frac{29}{3} - 9$$

$$= 2/3$$

$$\therefore SE_{\bar{x}} = \sqrt{\frac{2}{3}}$$

$\therefore$ SE for without replacement for population is given by

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$= \sqrt{\frac{8}{3}} \times \frac{1}{\sqrt{2}} \times \sqrt{\frac{3-2}{3-1}}$$

$$= \sqrt{\frac{2}{3}}$$

and thereby, we make the same conclusion as in previous case.

Example : Construct a sampling distribution of the sample mean for the following population when random samples of size 2 are taken from it (a) with replacement and (b) without replacement. Also find the mean and standard error of the distribution in each case.

Population Unit :	1	2	3	4
Observation :	22	24	26	28

Solution :

The mean and standard diviation of population are

$$\mu = \frac{22+24+26+28}{4} = 25 \text{ and}$$

$$\sigma = \sqrt{\frac{22^2+24^2+26^2+28^2}{4} - 25^2}$$

$$= \sqrt{5} = 2.236 \text{ respectively.}$$

(a) With replacement :

When random samples of size 2 are drawn, we have $4^2 = 16$ samples, shown below :

Simple No	Sample Values	$\bar{x}$
1	22, 22	22
2	22, 24	23
3	22, 26	24
4	22, 28	25
5	24, 22	23
6	24, 24	24
7	24, 26	25
8	24, 28	26
9	26, 22	24
10	26, 24	25
11	26, 26	26
12	26, 28	27
13	28, 22	25
14	28, 24	26
15	28, 26	27
16	28, 28	28

Since all of the above samples are equally likely, therefore, the probability of each value of $\bar{x}$ is $\frac{1}{16}$. Thus, we can write the sampling distribution of $\bar{x}$ as given below :

$\bar{x}$:	22	23	24	25	26	27	28	Total

$p:$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	1

$$\therefore E(\bar{x}) = 22\times\frac{1}{16} + 23\times\frac{2}{16} + 24\times\frac{3}{16} + 25\times\frac{4}{16} +$$

$$26\times\frac{3}{16} + 27\times\frac{2}{16} + 28\times\frac{1}{16}$$

$$= 25$$

$$V(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$$

$$= \left[22^2\times\frac{1}{16} + 23^2\times\frac{2}{16} + 24^2\times\frac{3}{16} + 25^2\times\frac{4}{16} + \right.$$

$$\left. 26^2\times\frac{3}{16} + 27^2\times\frac{2}{16} + 28^2\times\frac{1}{16}\right] - 25^2$$

$$= 627.5 - 625 = 2.5$$

$$\therefore \text{S.E.}(\bar{x}) = \sqrt{v(\bar{x})} = \sqrt{2.5}$$

Which is equal to $\frac{\sigma}{\sqrt{n}}$

(b) Without replacement :

When random samples of size 2 are drawn without replacement, we have ${}^4C_2$ samples, shown below :

Simple No	Sample Values	$\bar{x}$
1	22, 24	23
2	22, 26	24
3	22, 28	25
4	24, 26	25
5	24, 28	26
6	26, 28	27

Since all the samples are equally likely, the probability of each value of $\bar{x}$ is $\frac{1}{6}$. Thus, we can write the sampling distribution of $\bar{x}$ as

$\bar{x}:$	23	24	25	26	27	Total

$$p: \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{2}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad 1$$

$$\therefore E(\bar{x}) = \frac{1}{6}[23 + 24 + 25 \times 2 + 26 + 27]$$

$$= 25$$

$$\therefore V(\bar{x}) = E(\bar{x}^2) - [E(x)]^2$$

$$= \left[\frac{1}{6} \times 23^2 + \frac{1}{6} \times 24^2 + \frac{2}{6} \times 25^2 + \frac{1}{6} \times 26^2 + \frac{1}{6} \times 27^2\right] - 25^2$$

$$= 626.67 - 625$$
$$= 1.67$$

$$\therefore S.E.(\bar{x}) = \sqrt{v(\bar{x})} = \sqrt{1.67} = 1.292$$

Alternatively, population S.E. is given as

$$S.E.(\bar{x}) = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}} = \sqrt{\frac{4-2}{3} \times \frac{5}{2}}$$

$$= \sqrt{1.67} = 1.292$$

1.6 Summing Up

A sampling distribution is an array of sample studies relating to a population. If we select a number of independent random samples of a definite size from a given population and calculate some statistic like the mean, standard deviation etc. from each sample, we shall get an array of values of these statistics. The distribution so obtained by these values of the statistic is called the sampling distribution of that statistic. The standard deviation of the sampling distribution would be called the standard error which is abbreviated as S.E. The concept of Standard Error is very useful in testing statistical hypothesis and in the theory of estimation.

1.7 Model Questions

Objective Questions:

1. A population has N items. Samples of size n are selected without replacement. Find the number of possible samples.

2. Standard error is always non-negative. (True or False)

3. If the mean of population is μ then the mean sampling distribution is .... (fill in the blank)

4. Consider a population containing N items and n are selected as a sample with replacement. Find the numver of possible samples.

5. Sampling distribution describes the distribution of sample .... (fill in the blank)

Descriptive Questions:

1. Explain the concept of sampling distribution of a statistic.

2. A population consists of four numbers 3, 4, 2, 5. Consider all possible distinct samples of size two that can be drawn without replacement and verify that the population mean is equal to the mean of the sample means.

3. A simple random sample of size 36 is drawn from a finite population of 101 units. If the population S.D. is 12.6, find the standard error of the sample mean when the sample is drawn (i) with replacement, (ii) without replacement.

4. Consider a population of 6 units with values 1,2,3,4,5,6. Write down all possible samples of size 2 (without replacement) from this population and construct a sampling distribution of the sample mean. Also find the mean and standard error of the distribution.

5. What do you mean by 'Sampling Fluctuations'? Describe briefly.

1.8 References and Suggested Readings

1. Gupta S.C. & Kapoor V.K.; Fundamentals of Mathematical Statistics; Sultan Chand & Sons.

2. Hazarika P.; Essential Statistics for Economics and Commerce; Akansha Publishing House.

3. Rao Radhakrishna C.; Linear Statistical Inference and its Applications; Wiley Eastern Limited.

******

Unit-2

Statement of Central Limit Theorem, Estimation of the Mean and The Variance of the Sampling Distribution of Sample Mean

Unit Structure:

2.1 Introduction

It is seen that most of the distributions like Binomial, Poisson, etc. tend to normal distribution when the size of the sample is too large. For this reason and for otherwise also the distribution of sample mean, whatever be the nature of the parent population, will approach to the normal distribution as the size of the sample increases. This fact leads to the Central limit theorem, first proved by the French mathematician Pierre-Simon Laplace in 1810.

The theorem is applicable to all the populations in practice except a few which are very much different from the normal. It should be noted that the efficiency of the theorem increases with an increase in the sample size regardless of whether the source population is normal or skewed.

2.2 Objectives

After going through this unit, you will be able to

- know the basic concept of Central Limit Theorem

- understood the applications of Central Limit Theorem
- discuss the technique of finding estimation of the mean and the variance
- explain the concept of standard error

2.3 Central Limit Theorem

This theorem states that :

"If $\{x_1, x_2, ..., x_n\}$ is a random sample of size n from a non-normal population of size N with mean μ and standard deviation σ, then the sampling distribution of sample mean $\overline{x}$ will approach normal distribution with mean μ and standard error $\frac{\sigma}{\sqrt{n}}\left(\text{or } \sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}\right)$ as n becomes larger and larger."

It should be noted that as a general rule, when $n \geq 30$, the sampling distribution of $\overline{x}$ is taken to be normal for practical purposes. Moreover, the larger the sample size the better will be the approximation.

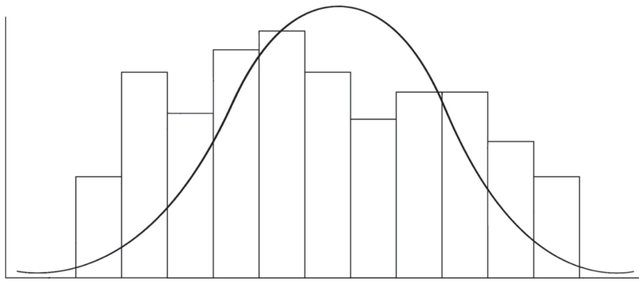
The statement of the above Central limit theorem is actually deducted from the generalized central limit theorem which is given as :

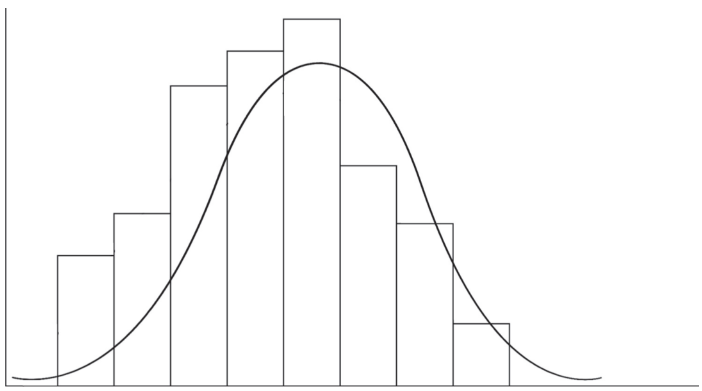
"If $\{x_1, x_2, ..., x_n\}$ are independent random variables following any distribution, then under certain very general conditions, their sum $\sum x = x_1 + x_2 + ... + x_n$ is asymptotically normally distributed, i.e. $\sum x$ follows normal distribution as $n \longrightarrow \propto$."

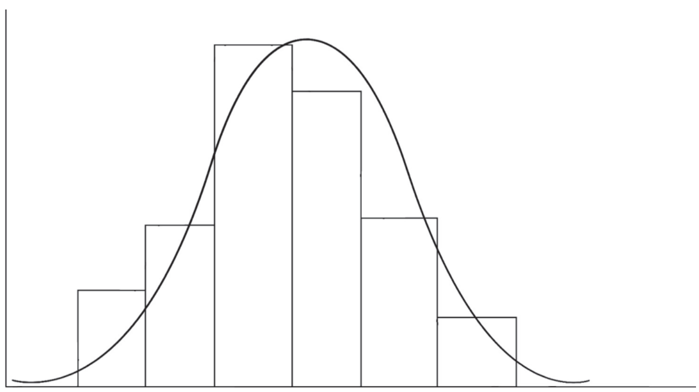
Thus the Central Limit Theorem asserts that for any statistic t, the random variable $Z = \frac{t - E(t)}{S.E.(t)}$ approaches the standard normal distribution of the population as n tends to infinity. This result is extensively used in Large Sample tests and in construction of confidence limits for the parameters provided the samples are relatively large.

2.3.1 How Does the Central Limit Theorem Works?

Basically the probability distributions are based on the concept of the Central Limit Theorem. For repeated sampling, the theorem provides us the behaviour of the population parameters estimates. When sample values are plotted on a graph, the theorem gives us the shape of the distribution formed by means. As the sample sizes get larger, the distribution of the means from the repeated sample tends to normalize and forms a normal distribution. Statistically, when sample size (n) is more than or equal to 30, the Central Limit Theorem works better. But in case, even though *n* is less than 30, the distribution of sample means may tend to normal if the source population is normally distributed.







From above, it is seen that

The averages of samples have approximately followed normal distribution.

Morever, as sample size increases, the Distribution of Averages tends to normal and the curve becomes narrow.

This Central Limit Theorem was first stated by great mathematician Laplace in 1812 and a regorous proof under general conditions was given by Liapounoff in 1901. Let us consider some particular cases of this general central limit theorem.

2.3.2 De-Moivre's Laplace Theorem

This theorem is a particular case of central milit theorem which is stated as follows :

"If $X_i = \begin{cases} 1, \text{ with probability } p \\ 0, \text{ with probability } q \end{cases}$

Then the distribution of the random variable $S_n = X_1 + X_2 + ... + X_n$, where X_i's are independent, is asymptotically normal as $n \to \propto$."

2.3.3 Lindeberg-Levy Theorem

This is an another particular case of central limit theorem which was jointly developed by Lindeberg and Levy. The statement of the theorem is given as :

"If $X_1, X_2, ..., X_n$ are independently and identically distributed random variable with

$$\left.\begin{matrix} E(X_i) = \mu_1 \\ V(X_i) = \sigma_1^{\ 2} \end{matrix}\right\}, \ i = 1,2,....,n$$

then the sum $S_n = X_1 + X_2 + ... + X_n$ is asymptotically normal with

$\mu = \frac{n^{\mu}}{1}$ and variance $\sigma^2 = n\sigma_1^2$".

2.3.4 Liapounoff's Central Limit Theorem

This particular case of generalised central limit theorem was developed by Liapounoff. The statement of the theorem is given as follow :

"Let $X_1, X_2, ..., X_n$ be independent random variable such that

$$\left.\begin{matrix} E(x_i) = \mu_i \\ V(X_i) = \sigma_i^{\ 2} \end{matrix}\right\} \ i = 1,2,....,n$$

Let us suppose that third absolute moment, say $\int_i^3$ of X_i about its mean exists, i.e.

$$\int_i^3 = E\{| X_i - \mu_i |^3\};\ i = 1, 2, \ldots\ldots, n \text{ is finite.}$$

Further let $\int^3 = \sum_{i=1}^{n} \int_i^3$.

If $\lim_{n \to \propto} \frac{\int}{\sigma} = 0$, the sum $X = X_1 + X_2 + \ldots\ldots + X_n$

is asymptotically $N(\mu,\ \sigma^2)$, where

$\mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^{\ 2}$ "

Check Your Progress

1. Define Standard Normal Variate.
2. What is the assertion of the statistic under Central Limit Theorem?
3. Name the particular cases of Central Limit Theorem.

Example: A certain group of people receives government welfare benefit of Rs. 110/- per week with a standard deviation of Rs. 20/-. If a random sample of size 25 people is drawn, what is that probability that their mean benefit will be greater than Rs. 120/- per week?

Solution : We are given,

$$\mu = 110$$
$$\overline{X} = 120$$
$$\sigma = 20$$

and $n = 25$

$$\therefore Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{120}{20 / \sqrt{25}} = \frac{10}{4} = 2.5$$

We are to find,

$$P(z > 2.5)$$

$$= 0.5 - P(0 < z < 2.5)$$
$$= 0.5 - 0.4938$$
$$= 0.0062$$

which is the required probability.

2.4 Estimations of the Mean and the Variance of the Sampling Distribution of the Sample Mean

Let $x_1, x_2 ...., x_n$ be a random sample of size n from a large population $X_1, X_2 ...., X_n$ of size N whose mean is μ and variance is σ^2.

The mean of the sampling distribution of the sample mean

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Now, $$E(\bar{x}) = E\left[\frac{x_1 + x_2 + ... + x_n}{n}\right]$$

$$= \frac{1}{n} E\left[x_1 + x_2 ... + x_n\right]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + ... + E(x_n)\right]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + ... + E(x_n)\right] \longrightarrow (1)$$

Since $x_i (i = 1,2,...n)$ is a sample observation from the population $X_i (i = 1,2,...N)$, hence it can take any one of the values $X_1, X_2 ..., X_N$ each with equal probability $\frac{1}{N}$.

$$\therefore E(x_i) = \frac{1}{N} X_1 + \frac{1}{N} X_2 + ... + \frac{1}{N} X_N$$

$$= \frac{1}{N}(X_1 + X_2 + ... + X_N)$$

$= \mu$, for each i from 1 to n.

Thus, $E(x_1) = E(x_2) = ... E(x_n) = \mu$

$$\therefore (1) \Rightarrow E(\bar{x}) = \frac{1}{n}\left[\mu + \mu + ... \text{ to n terms}\right]$$

$$= \frac{1}{n} . n\mu$$

$$= \mu$$

which shows that the mean of the sampling distribution of sample mean $\bar{x}$ is the population mean μ.

Again, $Var(\bar{x}) = Var\left(\frac{x_1 + x_2 ... + x_n}{n}\right)$

$$= \frac{1}{n^2}\left[\text{Var}(x_1) + \text{Var}(x_2) + ... + \text{Var}(x_n)\right] \longrightarrow (2)$$

the covariance terms vanish since the sample observations are independent of each other.

Now, $\text{Var}(x_i) = \text{E}\left[x_i - \text{E}(x_i)\right]^2$

$$= \text{E}\left[x_i - \mu\right]^2 \qquad \therefore \text{E}(x_i) = \mu$$

$$= \frac{1}{N}\left[(X_1 - \mu)^2 + (X_2 - \mu)^2 + ... + (X_N - \mu)^2\right]$$

$$= \sigma^2, \qquad \text{for each i from 1 to n}$$

$$\therefore (2) \Rightarrow \text{Var}(\bar{x}) = \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + ... \textit{to n terms}\right]$$

$$= \frac{1}{n^2} n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

$$\therefore \text{S.E.}(\bar{x}) = \sqrt{\text{Var}(\bar{x})} = \frac{\sigma}{\sqrt{n}}$$

2.5 Summing Up

The central limit theorem statis that for a random sample of size n drawn from a non-normal population with mean and variance, the sample mean approximately follows a noraml distribution with mean and variance. The larger the value of the size of the sample, the better will be the approximation to the normal.

The mean of the sampling distribution of sample mean $\bar{x}$ is the population mean μ and variance of the sample mean is $\frac{\sigma^2}{n}$. Further, we calculate S.E. $(\bar{x}) = \sqrt{V(\bar{x})}$.

2.6 Model Questions

1. Prove that the expectation of sample mean $\bar{x}$ is the population mean μ and the variance of sample mean is $\frac{\sigma^2}{n}$, where σ^2 is population variance and n is the sample size.

2. For a distribution with unknown mean μ has variance equal to 1.5. Use central limit theorem to find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample mean will be withing 0.5 of the population mean.

3. The life time of a certain brand of an electric bulb may be considered a random variable with mean 1200 hours and standard deviation 250 hours. Find the probability using central limit theorem, that the average life-time of 60 bulbs exceeds 1400 hours.

4. State the Lindberg-Levy Central Limit Theorem.

5. Define Central Limit Theorem. Write few applications of Central Limit Theorem.

2.7 References and Suggested Readings

1. Hogg, Tanis, Rao; Probability and Statistical Inference; Pearson.

2. Bhuyan K.C.; Probability Distribution Theory and Statistical Inference; New Central Book Agency (P) Ltd.

3. Gupta S.C. & Kapoor V.K.; Fundamentals of Mathematical Statistics; Sultan Chand & Sons.

4. Hazarika P.; Essential Statistics for Economics and Commerce; Akansha Publishing House.

5. Rao Radhakrishna C., Linear Statistical Inference and its Applications; Wiley Eastern Limited.

******

Unit-3
Point Estimation and Interval Estimation for Population Parameter

Unit Structure:

3.1 Introduction

The theory of statistical inference is based on sampling theory for making inferences about a population. The primary aim of sampling is to study the features of a population or to estimate the values of its parameter(s). It may be pointed out that it is possible to get reliable information about a population on the basis of sample information even if nothing is known about the population.

Estimation of population parameters by means of sample statistic is one of the important problems of statistical inference. This is often unavoidable for economic and business decisions and research studies. Thus, we can define the term estimation as follows -

"It is a procedure by which sample information is used to estimate the numerical magnitude of one or more parameters of the population. A function of sample values is called on estimator (or statistic) while its numerical value is called an estimate." For example $\overline{x}$ is an estimator of population mean μ. On the other hand, if $\overline{x}$=50 for a sample, the estimate of population mean is said to be 50.

3.2 Objectives

This unit is an attempt to have the basic ideas of Estimation. After going through this unit you will be able to –

- understand the concept of estimation

- understand point estimation and interval estimation
- explain the characteristics of a good estimator
- discuss the techniques of solving practical problems

3.3 Theory of Estimation

Let X be a random variable with probability density function or probability mass function $f(x; \theta_1, \theta_2, \ldots \theta_k)$, where $\theta_1, \theta_2, \ldots \theta_k$ are k parameters of the population.

Suppose, a random sample $(x_1, x_2, \ldots, x_n)$ of size *n* is drawn from the population and we are to estimate the k parameters $\theta_1, \theta_2, \ldots \theta_k$. In order to be specific, let x be a normal variate so that its probability density function can be written as $N(x : \mu, \sigma)$. Here, we may be interested to estimate the value of μ or σ.

It should be noted that, there may exist several estimators of a parameter, e.g., we can have any of the sample mean, median, mode, geometric mean, harmonic mean etc., as an estimator of population mean μ. Similarly, we can use either $S = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$ or $S = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$ as an estimator of population standard deviation σ.

This technique of estimation, where a single state like mean, median, Standard Deviation etc., is used as an estimator of population parameter, is known as Point Estimation. On the other hand, if an interval is estimated in which the value of the parameter is expected to lie, the procedure is termed as interval Estimation. The estimated interval is also termed as Confidence Interval.

3.4 Point Estimation

There can be more than one estimators of a population parameter. So, it is necessary to determine a good estimator out of a number of available estimators. We know that, a function of random variables $(x_1, x_2, \ldots, x_n)$, is a random variable. Therefore, a good estimator is one whose distribution is more concentrated around the population parameter. Thus, we may define point estimation as follows :

A particular value of a statistic which is used to estimate a given parameter is known as point estimate or estimator of the parameter.

According to R. A. Fisher, the founder of the theory, the following are some of the criteria of a good estimator :

(i) Unbaidsedness

(ii) Consistency

(iii) Efficiency

(iv) Sufficiency

(i) **Unbiasedness** :

A statistic $t = t\,(x_1, x_2, \ldots, x_n)$ is said to be an unbiased estimator of a parameter θ if $E(t) = \theta$. If $E(t) \neq \theta$, then it is said to be a biased estimator of θ. The magnitude of bias = $E(t) - \theta$.

We have seen that, $E(\overline{x}) = \mu$, where μ is the population mean, $\overline{x}$ is said to be an unbiased estimator of the population mean μ. But, since $E(S^2) \neq \sigma^2$, where $S^2 = \frac{1}{n}\sum_i (x_i - \overline{x})^2$ is not an unbiased estimator of the population variance σ^2. On the other hand, since $E(S^2) = \sigma^2$, where $s^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$ is an unbiased estimator of σ^2. However, since $S^2 = \frac{n-1}{n}s^2$ and $\frac{n-1}{n}$ approximates 1 when n is large, say, $n \geq 30$ for large sample, i.e., samples with size greater than or equal to 30, S can be taken as an estimator of σ.

It should be noted that, a statistic t is said to be positively or negatively biased according as $E(t) - \theta > 0$ or <0, i.e., $E(t) > \theta$ or $< \theta$

One should observe that the bias of an estimator usually decreases as the size of the simple increases.

(ii) **Consistency :**

A statistic $t_n = t_n(x_1, x_2, \ldots x_n)$ is said to be a consistent estimator of a parameter θ if t_n converges to θ in probability, i.e.,

$$\lim_{n \longrightarrow \infty} P\,(t_n \longrightarrow \theta) = 1$$

i.e., $P\{|t_n - \theta| > E\} \longrightarrow 0$ as $n \longrightarrow \infty$ for every $\in > 0$

It should be noted that, consistency is essentially a large sample property and strictly speaking it concerns not just one statistic, but a sequence of statistics.

We may note that $\overline{x}$ is a consistent estimator of population mean μ because $E(\bar{x}) = \mu$ and $Var(\bar{x}) = \frac{\sigma^2}{n} \longrightarrow 0$ as $n \longrightarrow \alpha$

Note : An unbiased estimator is necessarily a consistent estimator.

(iii) Efficiency :

It is possible to get many unbiased consistent estimators of a parameter. In such a situation efficiency is the criterion that decides the goodness of an estimator. If there exist several consistent estimators for a parameter θ, then the one whose sampling variance is minimum is known as the *most efficient estimator*.

Let us consider, t_1 and t_2 be two estimators of a population parameter θ such that both are either unbiased or consistent. Now, t_1 is said to be more efficient estimator than t_2 if Var (t_1) < Var (t_2).

For example, the sample mean $\bar{x}$ and sample median M_e both can be used as an estimator of the population mean μ. We have seen that, $E(\bar{x}) = \mu$ and $E(M_e) = \mu$, for large sample only. Again, $V(\bar{x}) = \frac{\sigma^2}{n}$ and $V(M_e) = \frac{\pi\sigma^2}{2n}$.

But, $V(\overline{x}) < V(M_e)$

So, sample mean $\bar{x}$ is more efficient than the sample median M_e.

(iv) Sufficeincy :

An estimator is said to be a sufficient estimator if it utilises all the information given in the sample about the parameter, i.e., a statistic $t = t(x_1, x_2, \ldots, x_n)$ based on a sample drawn from a population having probability density function (p.d.f.) $f(x, \theta)$ is said to be a sufficient estimator of θ if it contains all information about the parameter θ, i.e., if the conditional distribution of $x_1, x_2, \ldots, x_n$ for a given value of *t* is independent of θ, i.e., if $F\{(x_1, x_2, \ldots, x_n) | t = t_0\}$ does not depend on θ.

It is easy to observe that $\overline{x}$ is a sufficient estimator of μ.

Sufficient estimators are the most desirable but are not very commonly available. The following points must be noted about sufficient estimators :

1. A sufficient estimator is always consistent.
2. A sufficient estimator is most efficient if an efficient estimator exists.
3. A sufficient estimator may or may not be unbiased.

Methods of Points Estimation :

There are several methods of obtaining a point estimator of the population parameter. We shall, however, use the most popular method of maximum likelihood.

Let $x_1, x_2, \ldots x_n$ be a random sample of n independent observations from a population with probability density function (p.m.f) $f(x; \theta)$, where θ is unknown parameter for which we desire to find an estimator.

Since $x_1, x_2, \ldots x_n$ are independent random variables, their joint probability function or the probability of obtaining the given sample, termed as likelihood function, is given by

$$L = f(x_1;\theta).f(x_2;\theta).\,\ldots\ldots\ldots f(x_n;\theta)$$

$$= \prod_{i=1}^{n} f(x_i;\theta)$$

We have to find the value of θ for which L is maximum. The conditions for maxima of L are :

$$\frac{dL}{d\theta} = 0 \textit{ and } \frac{d^2L}{d\theta^2} < 0 .$$

The value of θ satisfying these conditions is known as Maximum Likelihood Estimator (MLE).

Generalising the above, if L is a function of k parameters $\theta_1, \theta_2, \ldots, \theta_k$, the first order conditions for maxima of L are :

$$\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial \theta_2} = \ldots = \frac{\partial L}{\partial \theta_k} = 0$$

From above, we will get k simultaneous equations in k parameters $\theta_1, \theta_2, \ldots, \theta_k$, and can be solved to get k maximum likelihood estimators.

In most cases, it is convenient to work using logarithm of L. Since log L is a monotonic transformation of L, the maxima of L and maxima of log L occur at the same value.

Example : Let a random sample of n observations $x_1, x_2, ..., x_n$ be drawn from a normally distributed population.

(i) If mean is unknown and the variance is known, find the maximum likelihood estimate of the mean,

(ii) if the mean is known but the variance is unknown, find the maximum likelihood estimate of the variance.

Solution : (i) Here $f(x_i, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

We have,

$$L = f(x_1, \mu) \bullet f(x_2, \mu) \bullet .... \bullet f(x_n, \mu)$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\sum\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow \log_e L = -\frac{n}{2}\log_e(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$$

Differentiating partially w.r.t. μ, we get

$$\frac{1}{L}\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2}\sum(x_i - \mu)$$

The likelihood equation for estimating μ is

$$\frac{\partial}{\partial \mu}(\log_e L) = 0$$

$$\Rightarrow \frac{1}{L}\frac{\partial L}{\partial \mu} = 0$$

$$\Rightarrow \frac{1}{\sigma^2}\sum(x_i - \mu) = 0$$

$$\Rightarrow \sum(x_i - \mu) = 0$$

$$\Rightarrow E x_i - n\mu = 0$$

$$\Rightarrow \mu = \frac{\sum x_i}{n}$$

$$\Rightarrow \mu = \overline{x}$$

Thus, the maximum likelihood estimate of μ is the sample mean.

(ii) $f(x_i, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Now, $L = f(x_1, \sigma^2).f(x_2, \sigma^2), \ldots, f(x_n, \sigma^2)$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow \log_e L = -\frac{n}{2}\log_e(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum (x_i - \mu)^2$$

Differentiating partially w.r.t. σ^2, we get

$$\frac{1}{L}\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum (x_i - \mu)^2$$

The likelihood equation for estimating σ^2 is

$$\frac{\partial}{\partial \sigma^2}(\log_e L) = 0$$

$$\Rightarrow \frac{1}{L}\frac{\partial L}{\partial \sigma^2} = 0$$

$$\Rightarrow -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum (x_i - \mu)^2 = 0$$

$$\Rightarrow \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

Thus, sample variance defined by $S^2 = \dfrac{\sum (x_i - \mu)^2}{n}$ is an estimator of σ^2.

Example : A random sample of size 5 is taken from a population containing 100 units. If the sample observations are 10, 12, 13, 7, 18, find

(i) an estimate of the population mean

(ii) an estimate of the standard error of sample mean

Solution : The estimate of the population mean (μ) is given by

$$\hat{\mu} = \overline{x}$$

The estimate of the standard error of sample mean is given by

$$\hat{SE}_{\bar{x}} = \frac{\sqrt{n}}{\sqrt{n-1}} \frac{s}{\sqrt{n}} \text{ for } SRSWR = \frac{\sqrt{n}}{\sqrt{n-1}} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \text{ for } SRSWOR$$

$$\text{i.e, } \hat{SE}_{\bar{x}} = \frac{S}{\sqrt{n-1}} \text{ for } SRSWR = \frac{S}{\sqrt{n-1}} \sqrt{\frac{N-n}{N-1}} \text{ for } SRSWOR$$

Now, let us prepare the following table

x	x^2
10	100
12	144
13	169
7	49
18	324
$\sum x - 60$	$\sum x^2 = 786$

$$\therefore \bar{x} = \frac{\sum x}{n} = \frac{60}{5} = 12$$

$$S^2 = \frac{1}{n} \sum x^2 - \bar{x}^2 = \frac{786}{5} - 12^2$$

$$= 157.20 - 144$$

$$= 13.20$$

$$= (3.633)^2$$

Hence we have

$$\hat{\mu} = 12$$

$$\hat{SE}_{\bar{x}} = \frac{3.633}{\sqrt{5-1}} \text{ for SRSWR}$$

$$= \frac{3.633}{\sqrt{5-1}} . \sqrt{\frac{100-5}{100-1}} \text{ for SRSWOR}$$

i.e. $\hat{SE}_{\bar{x}}$ =1.82 for SRSWR

= 1.78 for SRSWOR

Example : A random sample of size 65 was taken to estimate the man annual income of 1000 lower income families and the mean and standard deviation were found to be Rs. 6300 and Rs. 9.50 respectively. Find the standard error of the sample mean if sampling was done without replacement.

Solution : Since population is finite and sampling is drawn without replacement hence S.E. of $\bar{x}$ is given by

$$\hat{SE}_{\bar{x}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Here S = 9.5, N = 1000, n = 65

$$\therefore \hat{S.E.}_{\bar{x}} = \text{Rs.}\left[\frac{9.5}{\sqrt{65}} \times \sqrt{\frac{1000-65}{1000-1}}\right]$$

$$= \text{Rs.}\left[\frac{9.5}{8.06} \times \frac{30.58}{31.61}\right]$$

$$= \text{Rs.}1.14$$

3.5 Interval Estimation

Instead of estimating a parameter θ by a single value, we may consider an interval of values which is supposed to contain the parameter θ. An interval estimate is always expressed by a pair of unequal real values and the unknown parameter θ lies between these two values. Hence, an interval estimation may be defined as specifying two values that contains the unknown parameter θ on the basis of a random sample drawn from the population in all probability.

On the basis of random sample drawn from the population characterised by an unknown parameter θ, let us find two statistics t_1 and t_2 such that

$$p\,(t_1 < \theta) = \alpha_1$$

$$p\,(t_2 > \theta) = \alpha_2$$

for any two small positive quantities α_1 and α_2.

Combining these two conditions, we may write

$$p\,(t_1 \le \theta \le t_2) = 1 - \alpha \text{ where } \alpha = \alpha_1 + \alpha_2$$

where α is called the level of significance. The interval $[t_1, t_2]$ within which the unknown value of the parameter θ is expected to lie is called the confidence interval, the limits t_1 and t_2 so determined are known as confidence limits and $1-\alpha$ is called the confidence level of confidene coefficient. The term 'confidence interval' has its origin in the fact that if we select α=0.05 , then we feel confident that the interval [t_1, t_2], would contain the parameter θ in (1–α) % or (1–0.005)% or 95% of cases and the amount of confidence is 95%. This further means that if repeated samples of a fixed size are taken from the population with the unknown parameter θ, then in 95% of the cases, the interval [t_1, t_2] would contain θ and in the remaining 5% of the cases, it would fail to contain θ.

Check Your Progress
1. What do you mean by estimation?
2. What are the criteria of a good estimator?
3. Distinguish between point estimation and interval estimation.

Computation of Confidene Interval:

Let us assume that we have taken a random sample of size *n* from a normal population with mean μ and standard deviation σ. We assume further that the population standard deviation σ is known i.e. its value is specified. We know that the sample mean $\bar{x}$ is normally distributed with mean μ and standard deviation = $SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

If the assumption of normality is not tenable, then also the sample mean follows normal distribution approximately, statistically known as asymptotically, with population mean μ and standard diviation as $\frac{\sigma}{\sqrt{n}}$, provided the sample size *n* is sufficiently large. If the sample size *n*>30, then the asymptotic normality assumption holds. In order to select the appropriate confidence interval to the population mean, we need to determine a quantity *p*, say, such that

$$p(\bar{x} - p \times S.E._{\bar{x}} \leq \mu \leq \bar{x} + p \times S.E._{\bar{x}}) = 1 - \alpha$$

which finally leads to

$$\phi(p) = 1 - \frac{\alpha}{2}$$

Choosing as 0.05, we have

$$\phi(p) = 1 - \frac{0.05}{2} = 0.975 = \phi(1.96)$$

$$\Rightarrow p = 1.96$$

Hence 95% confidence interval to μ is given by

$$\left[\bar{x} - 1.96 \times S.E._{\bar{x}}, \bar{x} + 1.96 \times S.E._{\bar{x}}\right]$$

Similarly, 99% confidence interval to μ is given by

$$\left[\bar{x} - 2.58 \times S.E._{\bar{x}}, \bar{x} + 2.58 \times S.E._{\bar{x}}\right]$$

Below we mention the confidence limits of some important statistics for large random samples.

* Confience limits for population proportion P:

95% confidence limits are : p $\pm$ 1.96 $\times$ S.E.(p)

99% confidence limits are : p $\pm$ 2.58 $\times$ S.E.(p)

* Confidence limits for that difference $\mu_1 - \mu_2$ of two population means μ_1 and μ_2:

95% confidence limits are : $(\bar{x}_1 - \bar{x}_2) \pm 1.96 \times S.E.(\bar{x}_1 - \bar{x}_2)$

99% confidence limits are : $(\bar{x}_1 - \bar{x}_2) \pm 2.58 \times S.E.(\bar{x}_1 - \bar{x}_2)$

* Confience limits for the difference $P_1 - P_2$ of two population proportion :

95% confidence limits are : $(p_1 - p_2) \pm 1.96 \ S.E.(p_1 - p_2)$

99% confidence limits are : $(p_1 - p_2) \pm 2.58 \ S.E.(p_1 - p_2)$

Example : Construct 95% and 99% confidence intervals for mean of a normal population.

Solution : Let $x_1, x_2, \ldots x_n$ be a random sample of size *n* from a normal population with mean μ and standard deviation σ.

We know that sampling distribution of $\bar{x}$ is normal with mean μ and standard error $\frac{\sigma}{\sqrt{n}}$.

$$\therefore Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ will be a standard normal variate.

From the table of areas under standard normal curve, we can write

$$P(-1.96 \le z \le 1.96) = 0.95$$

$$\text{or } P\left(-1.96 \le \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \le 1.96\right) = 0.95 \text{ ———— (A)}$$

$$\text{or } P\left(-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Now, $-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x} - \mu$

$$\text{or } \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \text{ ———— (1)}$$

Similarly, $\bar{x} - \mu \le 1.96\frac{\sigma}{\sqrt{n}}$

$$\text{or } \mu \ge \bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \text{ ———— (2)}$$

Combining (1) and (2), we have

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

Thus, we can write equation (A) as

$$P\left(\bar{x} - 19.6\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

This gives us a 95% confidence interval for the parameter μ.
Similarly, we can construct a 99% confidence interval for μ as

$$P\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99$$

Example : A pharmaceutical company wants to estimate the mean life of a particular drug under typical weather conditions. A simple random sample of 81 bottles yields the following information :

Sample mean = 23 months

Population variance = 6.25 $(\text{months})^2$

Find an interval estimate with a confidence level of (i) 90% and (ii) 98%

Solution : Since the sample size $n = 81$ large, the mean life of the drug under consideration ($\bar{x}$) is asymptotically normal with population mean μ and Standard Error = Standard deviation

$$= \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{6.25}}{\sqrt{81}} = \frac{2.50}{9} = 0.2778$$

(i) Consulting Biometrika table, we find that

$$\phi(p) = 1 - \frac{\alpha}{2}$$
$$\Rightarrow \phi(p) = 1 - \frac{0.10}{2} = 0.95$$
$$\Rightarrow \phi(1.645) = 0.95$$
$$\Rightarrow p = 1.6450$$

$\therefore$ 90% confidence interval for μ is

$$\left[\bar{x} - p \times SE_{\bar{x}}, \bar{x} + p \times SE_{\bar{x}}\right]$$

$$= \left[23 - 1.645 \times 0.2778,\ 23 + 1.645 \times 0.2778\right]$$
$$= \left[22.5430, 23.4570\right]$$

(ii) In this case,

$$\phi(\rho) = 1 - \frac{0.02}{2} = 0.99$$
$$\Rightarrow \phi(2.325) = 0.99$$
$$\Rightarrow \rho = 2.325$$

Thus, 98% confidence interval to μ is
[23-2.3250×0.27778, 23+2.3250×0.27778]
= [22.35416, 23.6451]

Example : A random sample of 100 days shows an average daily sale of Rs. 1000 with a standard deviation of Rs. 250 in a particular shop. Assuming a normal distribution, find the limits which have a 95% chance of including the expected sales per day.

Solution : As given, n = 100

$\bar{x}$ = sample average sales = Rs. 1000

s = sample standard deviation = Rs. 250

$\therefore$ 95% confidence interval to the expected sales per day (μ) is given by

$$\text{Rs.}\left[\bar{x} \pm 1.96 \frac{s}{\sqrt{n-1}}\right]$$

$$= \text{Rs.}\left[1000 \pm 1.96 \times \frac{250}{\sqrt{100-1}}\right]$$

$$= \text{Rs.}[1000 \pm 49.25]$$

$$= [\text{Rs.}950.75, \text{Rs.}1049.25]$$

3.6 Summing Up

The theory of estimation is divided into two approaches namely point estimation and Interval estimation. In point estimation a single value of the statistic is used to provide an estimate of the parameter. On the other hand, in interval estimation, a range is specified within which the value of the parameter is most likely to lie with a known probability.

The characteristics of a good estimator under point estimation are unbiasedness, consistency, efficiency and sufficiency. There are several methods for obtaining the point estimates.

In interval estimation, we obtain the probable interval within which the unknown value of the parameter is expected to lie is called the confidence interval.

3.7 Model Questions

1. The following observations constitute a random sample from an unknown population. Estimate the mean and S.D. of the population. Also find the S.E. of sample means : 14,19,17,20,25.
2. A random sample of the heights of 100 students from a large population of students in a university having S.D. of 0.75 ft. has an average height of 5.6 ft. Find (i) 95% and (ii) 99% confidence limits for the average height of all the students of the university.
3. What do you understand by point Estimation? When would you say that estimate of a paraneter is good? Explain briefly.
4. State and explain the principle of maximum likelihood (M.L.) for estimation of population parameter.
5. Discuss the concept of interval estimation and provide suitable example.
6. Explain the following terms:
   (i) Sufficient estimator

(ii) Efficient estimator

(iii) Maximum Likelihood estimator

3.8 References and Suggested Readings

1. Hogg, Tanis, Rao; Probability and Statistical Inference; Pearson.
2. Bhuyan K.C.; Probability Distribution Theory and Statistical Inference; New Central Book Agency (P) Ltd.
3. Gupta S.C. & Kapoor V.K.; Fundamentals of Mathematical Statistics; Sultan Chand & Sons.
4. Hazarika P.; Essential Statistics for Economics and Commerce; Akansha Publishing House.
5. Rao Radhakrishna C., Linear Statistical Inference and its Applications; Wiley Eastern Limited.

******