# UNIT 6

# UNBIASEDNESS

## Structure



## 6.1 INTRODUCTION

In many real-life problems, the population parameter (characteristic of the population) is not known and someone is interested in obtaining the value of the parameter. But, if

- the whole population is too large to study,
- the units of the population are destructive in nature,
- there are limited resources and manpower available, etc.

then it is not practically convenient to examine each unit of the population to find the value of the parameter. For example, as you know many of us use Facebook and you are interested to know the average age of the people who use Facebook. However, the true value (average age) of Facebook users is not known. The only way to know the true average age of Facebook users is to survey each and every person in the world who uses Facebook. But it is not possible to survey everyone in the world. In such a situation, one can select randomly some persons who use Facebook and note their age. Suppose we randomly selected 20 Facebook users and obtained the following data of their age (in years):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 20 | 42 | 36 | 30 | 20 | 52 | 32 | 18 | 70 | 22 |
| 45 | 18 | 40 | 16 | 18 | 20 | 30 | 19 | 41 | 20 |

If we use the sample average age to estimate the unknown average age of the Facebook users, then we get an estimate of the same as

**Tools You Will Need**

The following terms are considered essential background material for this Unit. If you doubt your knowledge of any of these terms, you should review the appropriate Unit or section before proceeding:

- Sampling distributions (Units 2,3, 4 and 5).
- Probability distributions (MST-012).



**Unit Writer- Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi**

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{607}{20} = 30.25$$

Estimating is not something new to us. Every one of us uses an estimate in our day-to-day life. Some situations are as follows:


- At the metro station of Delhi, a guard may estimate the height of a child to be 3 feet or longer.
- Lavnik estimates the time to reach his school from home is about 20 minutes.
- A family estimates the monthly expenditure on the basis of particular needs.
- The distance between New Delhi and Gujrat is approximately 1112 km.

So, the question is, what is an estimation?

**The technique of finding an estimator to produce an estimate (approximate value) of the unknown parameter of the population on the basis of a sample is called estimation.**

Any statistic used to estimate an unknown population parameter is known as **estimator** and the particular value of the estimator is known as **estimate** of parameter. The estimated value of sample mean and sample variance are denoted by $\overline{X}$ and $S^2$, respectively.

As its name suggests, the objective of estimation is to determine the approximate value of a population parameter on the basis of a sample statistic.

There are two methods of estimation:

1. Point Estimation
2. Interval Estimation

In point estimation, we determine an appropriate single statistic whose value is used to estimate the unknown parameter whereas, in interval estimation, we determine an interval that contains the true value of the unknown parameter with a certain confidence. For example, in the case of Facebook users, we get the point estimate as 30.25 years because we estimated it by only one value (30.25 years) whereas if we estimate the same as the age group (18, 34) uses Facebook then it is an interval estimation because we estimated it by using an interval (18, 34) age. The point estimation and interval estimation are briefly described in Units 10-11 and Units 12-14, respectively.

Estimation admits two problems:

- The first is to select some criteria or properties such that if an estimator possesses these properties then it is called the best estimator among possible estimators, that is, properties of a good estimator, and
- The second is to derive some methods or techniques through which we obtain an estimator which possesses such properties, that is, methods of estimation.

Units 6, 7, 8 and 9 are devoted to describing the properties of a good estimator in detail, however, Units 10, 11, 12, 13 and 14 explain the methods of estimation.

This unit is divided into nine sections. Section 6.1 is introductory in nature. The basic terms used in estimation are defined in Section 6.2. Section 6.3 is devoted to explaining the criteria of a good estimator. Section 6.4 explores the concept of unbiasedness with examples. The properties of an unbiased estimator are described in Section 6.5. The unit ends by providing a summary of what we have discussed in this unit in Section 6.6. The terminal questions and the solution of the SAQs/TQs are given in Sections 6.7 and 6.8, respectively.

In the next unit, we shall discuss the second characteristic of a good estimator, that is, consistency.

## Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ define the parameter space;
- ❖ describe the properties of a good estimator;
- ❖ explain the unbiasedness characteristic of an estimator;
- ❖ check whether an estimator is unbiased or not; and
- ❖ describe the properties of an unbiased estimator.

# 6.2 BASIC TERMINOLOGY

Before discussing the properties of a good estimator, we discuss basic definitions of some important terms. These terms are very useful in understanding the fundamentals of the theory of estimation discussed.

### Discrete and Continuous Distributions

In Units 9 to 16 of MST-012, we have discussed standard discrete and continuous distributions as binomial, Poisson, normal, exponential, etc. We know that a population can be described with the help of a distribution, therefore, standard discrete and continuous distributions are also used in statistical inference. Here, we discuss some standard discrete and continuous distributions in brief as in tabular form and you have to learn at least the mean and variance of these distributions which will pay you to do the estimation questions easily.

| S. No. | Distribution | Parameter(s) | Mean | Variance |
|---|---|---|---|---|
| 1 | Bernoulli (discrete)<br>$P[X=x]=p^x(1-p)^{1-x};\ \ x=0,1$ | p | p | p(1 – p) |
| 2 | Binomial (discrete)<br>$P[X=x]={}^nC_x p^x(1-p)^{n-x};\ \ x=0,1,\ldots,n$ | n & p | np | np(1 – p) |
| 3 | Poisson (discrete)<br>$P[X=x]=\dfrac{e^{-\lambda}\lambda^x}{x!};\ \ x=0,1,\ldots\ \ ;\lambda>0$ | $\lambda$ | $\lambda$ | $\lambda$ |
| 4 | Uniform (discrete)<br>$P[X=x]=\dfrac{1}{n};\ \ x=1,2,\ldots,n$ | n | $\dfrac{n+1}{2}$ | $\dfrac{n^2-1}{12}$ |
| 5 | Hypergeometric (discrete)<br>$P[X=x]=\dfrac{{}^MC_x\,{}^{N-M}C_{n-x}}{{}^NC_n};\ x=0,1,\ldots,\min\{M,n\}$ | N, M & n | $\dfrac{nM}{N}$ | $\dfrac{NM(N-M)(N-n)}{N^2(N-1)}$ |

| | | | | |
|---|---|---|---|---|
| 6 | Geometric (discrete)<br>$P[X=x]=p(1-p)^x; \quad x=0,1,2,...$ | p | $\frac{p}{(1-p)}$ | $\frac{p}{(1-p)^2}$ |
| 7 | Negative Binomial (discrete)<br>$P[X=x]=\binom{x+r-1}{r-1}p^r(1-p)^x; \quad x=0,1,2,...$ | r & p | $\frac{rp}{(1-p)}$ | $\frac{rp}{(1-p)^2}$ |
| 8 | Normal (continuous)<br>$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; -\infty<x<\infty;$<br>$\sigma>0, -\infty<\mu<\infty$ | $\mu$ & $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 9 | Standard Normal (continuous)<br>$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}; \quad -\infty<x<\infty$ | -- | 0 | 1 |
| 10 | Uniform (continuous)<br>$f(x)=\frac{1}{b-a}; a<x<b, b>a$ | a & b | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| 11 | Exponential (continuous)<br>$f(x)=\theta e^{-\theta x}; \quad x\ge 0; \theta>0$<br>Negative Exponential or simply exponential (continuous)<br>$f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}; \quad x\ge 0; \theta>0$ | $\theta$<br>$\theta$ | $\frac{1}{\theta}$<br>$\theta$ | $\frac{1}{\theta^2}$<br>$\theta^2$ |
| 12 | Gamma (continuous)<br>$f(x)=\frac{b^a}{\Gamma a}e^{-bx}x^{a-1}; x>0; a,b>0$ | a & b | $\frac{b}{a}$ | $\frac{b}{a^2}$ |
| .13 | Beta First Kind (continuous)<br>$f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}; 0<x<1;$<br>$a>0, b>0$ | a & b | $\frac{a}{a+b}$ | $\frac{ab}{(a+b)^2(a+b+1)}$ |
| 14 | Beta Second Kind (continuous)<br>$f(x)=\frac{1}{B(a,b)}\frac{x^{a-1}}{(1+x)^{a+b}}; x>0; \ a,b>0$ | a & b | $\frac{a}{b-1}$ | $\frac{a(a+b+1)}{(b-1)^2(b-2)}$ |
| 15 | Standard Cauchy<br>$f(x)=\frac{1}{\pi(1+x^2)}; \ -\infty<x<\infty$ | --- | Does not exist | Does not exist |
| 16 | Laplace<br>$f(x)=\frac{1}{2b}e^{-\left\lvert\frac{x-\mu}{b}\right\rvert}; \ -\infty<x<\infty$ | $\mu$ & b | $\mu$ | $2b^2$ |

**Parameter Space**

The set of all possible values that the parameter $\theta$ or parameters $\theta_1, \theta_2, \ldots, \theta_k$ can assume is called the parameter space. It is denoted by $\Theta$ and is read as **"big theta"**. For finding the parameter space of a parameter, we have to think all possible values of the parameter yet the chance of these is very very small. For example, suppose the parameter $\theta$ represents the average life of electric bulbs manufactured by a company. Since the bulb can be fused at the initial time 0 or at 1, 2, 2.3, 3 hours, and so on, therefore, it lies from 0 to ∞. Hence, the parameter space of the average life of the bulbs, that is, $\theta$ is $\Theta = \{\theta : \theta \geq 0\}$.

It means that the parameter average life $\theta$ can take all possible values greater than or equal to 0, Similarly, in a normal distribution ($\mu$, $\sigma^2$), the parameter space of parameters $\mu$ and $\sigma^2$ is $\Theta = \left\{(\mu, \sigma^2) : -\infty < \mu < \infty;\ \ 0 < \sigma < \infty\right\}$.

**Mathematical Expectation**

If X is a continuous random variable having the probability density function $f(x)$, then the expected value of X (mean) is defined as

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx \text{ and } E\left(X^r\right) = \int_{-\infty}^{\infty} x^r\, f(x)\, dx$$

If X is a discrete random variable having the probability mass function $p(x)$, then the expected value of X is defined as

$$E(X) = \sum_{i=1}^{n} x_i\, p(x_i) \text{ and } E\left(X^r\right) = \sum_{i=1}^{n} x_i^r\, p(x_i)$$

Some properties of mathematical expectation are:

- $E(a) = a$ where ‘a’ is a constant
- $E(aX) = aE(X)$
- $E(aX \pm bY) = aE(X) \pm bE(Y)$

**Variance**

If X is a random variable then the variance of X in terms of expectation is defined as

$$Var(X) = E\left[X - E(X)\right]^2 = E\left(X^2\right) - \left[E(X)\right]^2$$

Some properties of variance are:

- $Var(a) = 0$
- $Var(aX) = a^2 Var(X)$
- If random variables X and Y are independent, then

  $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$

Now, try the following Self Assessment Question.

***SAQ 1***

If $\theta$ represents the average marks (out of 50) of the learner in the Term-End-Exam paper of the MST-016 course, then find the parameter space of $\theta$.

After understanding the basic definition and terminology which will help you to understand the properties of a good estimator. We now finally discuss the properties of a good estimator in the next section.

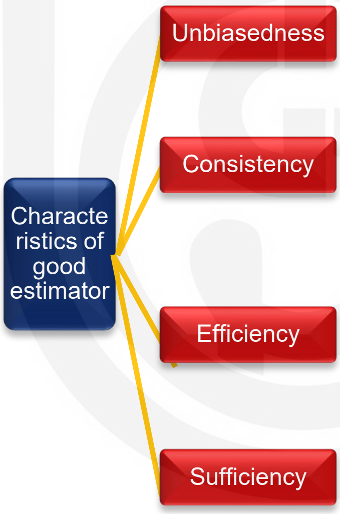
## 6.3 PROPERTIES OF GOOD ESTIMATOR

It is to be noted that a large number of estimators can be proposed for an unknown parameter. For example, in our case of estimating the average age of Facebook users, some possible estimators are:

- Sample mean $\overline{X} = \dfrac{607}{20} = 30.25$
- Sample median $\tilde{X} = \dfrac{22+30}{2} = 26$
- Sample mode $X_0 = 20$
- Average of extreme users $= \dfrac{\max + \min}{2} = \dfrac{70+14}{2} = 42$

Now, the questions arise,

- Which estimator should you use, that is, which is likely to give estimates closer to the true (but unknown) population value?
- Are some of the possible estimators better, in some sense, than the others?"




In general, an estimator whose sampling distribution concentrates as closely as possible near the true value of the parameter may be regarded as a good estimator. To give the answer to the above questions, Prof. Ronald A. Fisher gave some properties of a good estimator which are as follows:

- Unbiasedness
- Consistency
- Efficiency
- Sufficiency

We shall discuss these properties one by one in the subsequent Units.

Now, answer the following Self Assessment Question.

### *SAQ 2*

Write properties of a good estimator.

We now discuss the first characteristic of a good estimator in the next section.

## 6.4 UNBIASEDNESS

Any statistic which is used to estimate an unknown population parameter is known as **estimator.**

In the previous units, you have studied that any statistic such as sample mean, sample variance, sample proportion, etc. which is used to estimate an unknown population parameter is known as an estimator. You also saw that the value of any estimator changes from sample to sample, therefore, we consider the estimator as a random variable and we can find the mean and variance of the estimator. So we can define an estimator as an unbiased estimator as:

An estimator is said to be unbiased for a population parameter if and only if the average or mean of the sampling distribution of the estimator is equal to the true value of the parameter. This property of the estimator is called unbiasedness.

Let us see some examples,

- In Unit 1, you have seen that the mean of the sampling distribution of the sample mean of monthly salary of the employees is equal to the mean salary of all employees of the industry. So sample mean is an unbiased estimate of the population mean.
- Similarly, in Unit 3, we saw that the mean of the sample proportions of the children who like to dance is equal to the population proportion. Therefore, sample proportion is an unbiased estimate of the population proportion.

In general, we denote any population parameter such as a population mean, population standard deviation, population proportion, and so on by the Greek letter theta $\theta$, and its estimator such as the sample mean, sample standard deviation, and sample proportion by T or $\hat{\theta}$ (pronounced as **"theta-hat"**).

Mathematically,

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population whose probability density (mass) function is $f(x,\theta)$ where $\theta$ is the population parameter then an estimator $T = t(X_1, X_2, ..., X_n)$ is said to be an unbiased estimator of the parameter $\theta$ if and only if

$$E(T) = \theta$$

for all possible values of the parameter $\theta$.

An estimator is said to be unbiased if the expected value of the estimator is equal to the true value of the parameter being estimated.

However, if the expected value of the estimator does not equal to the true value of the parameter, then the estimator is said to be a **"biased estimator"**, that is, if

$$E(T) \neq \theta$$

then the estimator T is called the biased estimator of $\theta$.

We can also define bias as

The distance between the estimate obtained from a sample and the actual value of the population parameter from which the sample was taken is called bias.

The amount of biases is given by

$$b(\theta) = E(T) - \theta$$

- If $b(\theta) > 0$ or $E(T) > \theta$, then the estimator T is said to be positively biased for the parameter $\theta$.
- If $b(\theta) < 0$ or $E(T) < \theta$, then the estimator T is said to be negatively biased for the parameter $\theta$.
- If $E(T) \to \theta$ as $n \to \infty$, that is, if an estimator T is unbiased for a large sample only then the estimator T is said to be asymptotically unbiased for $\theta$. For example, suppose $E(T) = \theta + \frac{1}{n}$ then as $n \to \infty, E(T) \to \theta$.

An unbiased estimator is generally preferred in comparison to a biased estimator.

Now, we explain the procedure to show whether an estimator is unbiased or not for a parameter with the help of some examples.

**Example 1:** Show that the sample mean $(\bar{X})$ is an unbiased estimator of the population mean (μ) if it exists.

**Solution:** Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from any population with mean μ. We have to show that the sample mean $\bar{X}$ is an unbiased estimator for μ, therefore, we have to find $E(\bar{X})$ and check whether it is equal to μ or not. That is,

$$E\left(\bar{X}\right) = \mu$$

Consider,

$$E\left(\bar{X}\right) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \qquad \left[\text{By defination of the sample mean}\right]$$

$$= \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right] \left[\because E\left(aX + bY\right) = aE\left(X\right) + bE\left(Y\right)\right]$$

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population with mean μ and variance $\sigma^2$, therefore,

$$E(X_1) = E(X_2) = \ldots = E(X_n) = E(X) = \mu$$

Thus,

$$E(\bar{X}) = \frac{1}{n}\left(\underbrace{\mu + \mu + \ldots + \mu}_{n-\text{times}}\right) = \frac{1}{n}(n\mu) = \mu$$

Hence, the sample mean $(\bar{X})$ is an unbiased estimator of the population mean μ. Also if $x_1, x_2, \ldots, x_n$ are the observed values of the random sample $X_1, X_2, \ldots, X_n$ then $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an unbiased estimate of the population mean.

**Example 2:** Suppose the speed of lightweight vehicles on a particular stretch of roadway is normally distributed with a known standard deviation of 5 kph. A researcher measured the speed of 10 lightweight vehicles randomly and obtained the following results:

| **Vehicle** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Speed (in kph)** | 62 | 70 | 65 | 68 | 64 | 65 | 70 | 64 | 55 | 60 |

(i) Find the point estimate of the average speed.

(ii) Show that the sample average speed is an unbiased estimator of the average speed of all the lightweight vehicles on the roadway.

**Solution:** Generally, to draw the inference about the population mean, we use the sample mean, therefore, to find the point estimate of the average speed, we use the sample mean.

We can obtain the point estimate of average speed as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$= \frac{62 + 70 + 65 + 68 + 64 + 65 + 70 + 64 + 55 + 60}{10} = 64.3$$

Now, we have to show that the sample average speed is an unbiased estimate of the average speed of all lightweight vehicles on the roadway.

Since the speed of the vehicles is normally distributed and standard deviation ($\sigma$) is known, therefore, the sample average speed also follows a normal distribution with mean $\mu$ and variance $\sigma^2/n$.

Thus, $E\left(\overline{X}\right) = \mu$

Hence the sample average speed is an unbiased estimate of the average speed of all lightweight vehicles on the roadway.

**Example 3:** A machine produces a large number of water bottles. A quality inspector selected 40 water bottles randomly and found 2 defective water bottles. Find the point estimate of the proportion of all defective water bottles.

**Solution:** To draw the inference about the population proportion, we use the sample proportion, therefore, to find the point estimate of the proportion of all defective water bottles, we use sample proportion defectives.

Therefore, we can obtain the point estimate of the proportion of all defective water bottles as

$$p = \frac{X}{n} = \frac{2}{40} = 0.05$$

Hence, the point estimate of the proportion of all defective water bottles is 0.05.

**Example 4:** A furniture company manufacturing square tables of a side length $\mu$. Thus, the area of the table will be $\mu^2$ (unknown). Based on n independent measurements $X_1, \ldots, X_n$ of the length, estimate area of the table. Assume that the measurements of the length have mean $\mu$ and variance $\sigma^2$.

(i) Show that $\overline{X}^2$ is not an unbiased estimator for $\mu^2$.

(ii) For what value of k, is the estimator $\overline{X}^2 - kS^2$ unbiased for $\mu^2$?

**Solution:** Since the measurements of the length have mean $\mu$ and variance $\sigma^2$, therefore, the sampling distribution of mean has mean and variance as follows:

$$E\left(\overline{X}\right) = \mu \text{ and } Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$$

We have to show that $\overline{X}^2$ is not an unbiased estimator for $\mu^2$, therefore, we have to find $E(\overline{X}^2)$ and check whether it is equal to $\mu^2$ or not. But the question is how we find the $E(\overline{X}^2)$ without knowing the sampling distribution of $\overline{X}^2$. However, we know that

$Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$ and by the formula of variance, we have

$$\mathrm{Var}\left(\bar{X}\right) = E\left(\bar{X}^2\right) - \left[E\left(\bar{X}\right)\right]^2$$

Therefore,

$$E\left(\bar{X}^2\right) = \mathrm{Var}\left(\bar{X}\right) + \left[E\left(\bar{X}\right)\right]^2 = \frac{\sigma^2}{n} + \mu^2 \neq \mu^2$$

Hence, $\bar{X}^2$ is not an unbiased estimator of the area of the table, that is, $\mu^2$ and we can calculate the bias of the estimator as

$$E\left(\bar{X}^2\right) - \mu^2 = \frac{\sigma^2}{n} \Rightarrow E\left(\bar{X}^2 - \mu^2\right) = \frac{\sigma^2}{n} \qquad \left[\because E(a) = a\right]$$

We now find the value of k such that the estimator $\bar{X}^2 - kS^2$ is unbiased for $\mu^2$.

Since the estimator $\bar{X}^2 - kS^2$ is unbiased for $\mu^2$, therefore,

$$E\left(\bar{X}^2 - kS^2\right) = \mu^2$$

$$E\left(\bar{X}^2\right) - kE\left(S^2\right) = \mu^2 \qquad \left[\because E\left(aX \pm bY\right) = aE\left(X\right) \pm bE\left(Y\right)\right]$$

$$\mu^2 + \frac{\sigma^2}{n} - k\sigma^2 = \mu^2 \quad \text{[since } S^2 \text{ is an unbiased estimator of } \sigma^2\text{,i.e. } E(S^2)= \sigma^2\text{]}$$

Therefore,

$$k = \frac{1}{n}$$

Hence, for $k = \frac{1}{n}$, the estimator $\bar{X}^2 - kS^2$ is unbiased for $\mu^2$.

**Example 5:** If $X_1, X_2, \ldots, X_n$ is a random sample taken from a population with a mean $\mu$ and variance $\sigma^2$, then

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \text{ is a biased estimator of } \sigma^2.$$

whereas, $S^2 = \frac{n}{n-1}S'^2$ i.e. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is an unbiased estimator of $\sigma^2$.

**Solution:** We have to show that $S'^2$ is not an unbiased estimator for $\sigma^2$, therefore, we have to find $E(S'^2)$ and check whether it is equal to $\sigma^2$ or not. Therefore, we consider:

$$E\left(S'^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) \quad \left[\because E\left(aX\right) = aE\left(X\right)\right]$$

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i^2 + \bar{X}^2 - 2X_i\bar{X}\right)\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 + \sum_{i=1}^{n}\bar{X}^2 - -2\bar{X}\sum_{i=1}^{n}X_i\right)$$

$$= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2 + n\bar{X}^2 - 2\bar{X}n\bar{X}\right) \quad \left[\because \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i \Rightarrow \sum_{i=1}^{n}X_i = n\bar{X}\right]$$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n}\left[E\left(\sum_{i=1}^{n} X_i^2\right) - nE\left(\bar{X}^2\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} E\left(X_i^2\right) - nE\left(\bar{X}^2\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}\left(\mu^2 + \sigma^2\right) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right] \quad \text{[As discussed in Example 4]}$$

$$= \frac{1}{n}\left[n\left(\mu^2 + \sigma^2\right) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right]$$

$$= \mu^2 + \sigma^2 - \mu^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

$$E\left(S'^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Hence, $S'^2$ is not an unbiased estimator for $\sigma^2$.

Now, we check whether $S^2$ is unbiased of $\sigma^2$ or not, therefore, we consider:

$$E\left(S^2\right) = \frac{n}{n-1} E\left(S'^2\right) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Hence, estimator $S^2$ is an unbiased estimator for $\sigma^2$.

This is the reason why we consider $S^2$ in place of $S'^2$ for estimating the sample variance.

Now, you can assess your understanding by answering the following Self Assessment Question.

## *SAQ 3*

**a.** A company produces batteries for laptops and wants to estimate the average life of the battery. For that, the statistician of the company selected 5 batteries from the production and measured their lives. He suggests two estimators for estimating the average life of the battery:

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}, \; T_2 = \frac{X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5}{15}$$

where $X_1$, $X_2$, $X_3$, $X_4$ and $X_5$ repersent life of the selected batteries. It is known that the life of batteries has mean μ and variance $\sigma^2$. Are both estimators unbiased?

**b.** Show that the sample mean and sample median are both unbiased estimators for mean (μ) of a normal distribution.

## 6.5 PROPERTIES OF UNBIASED ESTIMATOR

After understanding the concept of unbiasedness and how to check whether an estimator is unbiased or not, we now discuss some properties of the unbiased estimator as follows:

1. Unbiased estimators may not be unique. For example, the sample mean and sample median are unbiased estimators of the population mean of a normal population.

2. Unbiased estimators do not always exist for all parameters. For example, for a Bernoulli distribution ($\theta$), there is no unbiased estimator for $\theta^2$. Similarly, for a Poisson distribution ($\lambda$), there exists no unbiased estimator for $1/\lambda$.

3. If an estimator is unbiased for all types of distribution, then it is called an **absolutely unbiased** estimator. For example, the sample mean is an absolutely unbiased estimator of the population mean, if the population mean exists.

4. If T and T* are two unbiased estimators of the parameter $\theta$ then

   $aT + (1-a)\,T^*$

   is also an unbiased estimator of $\theta$ where 'a' $(0 \le a \le 1)$ is any constant.

For a better understanding of the unbiasedness try Self Assessment Questions.

## *SAQ 4*

Some patients with high blood pressure are randomly assigned to a placebo group and a treatment group. The placebo patients receive an inactive pill, and the treatment patients receive a new drug that is expected to lower blood pressure. After the patients are treated for two months, the high blood pressures of the patients of both groups are measured and given as follows:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Placebo Group (X)** | 140 | 165 | 170 | 140 | 135 | 170 | 165 | 150 | 140 |
| **Treatment Group (Y)** | 130 | 135 | 140 | 130 | 120 | 120 | 132 | 118 | 120 |

If $E(X) = \mu_1, Var(X) = \sigma_1^2$ and $E(Y) = \mu_2, Var(Y) = \sigma_2^2$, then

(i) Show that the statistic $(\bar{X} - \bar{Y})$ is an unbiased estimator of the parameter $(\mu_1 - \mu_2)$. Also, find the estimate of the same using the given data.

(ii) Calculate the variance and standard deviation of the estimator in Part (i). Also, find the estimate of standard error.

(iii) Calculate an estimate of the ratio $\sigma_1 / \sigma_2$.

We now end this unit by giving a summary of what we have covered in it.

## 6.6 SUMMARY

In this unit, we have covered the following points:

- If we estimate an unknown parameter by a single statistic then this technique is known as point estimation whereas if we determine an interval (using sample values) that contains the true value of the unknown parameter with a certain confidence then it is known as interval estimation.

- The set of all possible values that the parameter $\theta$ or parameters $\theta_1, \theta_2, \ldots, \theta_k$ can assume is called the parameter space. It is denoted by $\Theta$.
- The properties of a good estimator are unbiasedness, consistency, efficiency and sufficiency.
- An estimator is said to be unbiased if the expected value of the estimator is equal to the true value of the parameter being estimated.
- The properties of an unbiased estimator.

## 6.7 TERMINAL QUESTIONS

**1.** A random sample $X_1, X_2, \ldots, X_n$ of size n taken from a population whose pdf is given by

$$f(x,\theta)=\frac{1}{\theta}e^{-x/\theta}; \quad x>0, \theta>0$$

Show that sample mean $(\bar{X})$ is an unbiased estimator of parameter $\theta$.

**2.** Consider a population comprising three LED televisions of a certain company. If the lives of the LED televisions are 8, 6 and 10 years then construct the sampling distribution of the average life of the LED televisions by taking samples of size 2 and show that the sample mean is an unbiased estimator of the population mean life. Also, show that $S'^2$ is not an unbiased estimator of population variance whereas $S^2$ is an unbiased estimator of population variance where

$$S'^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \text{ and } S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

## 6.8 SOLUTIONS / ANSWERS

### Self Assessment Questions (SAQs)

**1.** Since parameter $\theta$ represents the average marks in the paper of the MST-016 course of 50 marks and there is no negative marking, therefore, a learner can take a minimum 0 marks and a maximum 50 marks. Thus, the parameter space of $\theta$ is $\Theta=\{\theta: 0\le\theta\le 50\}$.

**2.** Refer to Section 6.3.

**3(a)** We have to check whether estimator $T_1$ and $T_2$ are unbiased or not. Therefore, we have to find $E(T_1)$ and $E(T_2)$ and check whether it is equal to $\mu$ or not. Since $X_1, X_2, X_3, X_4$ and $X_5$ are independent and taken from the same population with a mean $\mu$ and variance $\sigma^2$, therefore,

$$E\left(X_i\right)=\mu \text{ and } Var\left(X_i\right)=\sigma^2 \quad \text{for all } i=1,2,\ldots,5$$

We now consider

$$E\left(T_1\right)=E\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right]$$

$$= \frac{1}{5}\left[E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)\right]$$

$$= \frac{1}{5}\left[\mu + \mu + \mu + \mu + \mu\right]$$

$$E(T_1) = \mu$$

Similarly,

$$E(T_2) = E\left[\frac{X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5}{15}\right]$$

$$= \frac{1}{15}\left[E(X_1) + 2E(X_2) + 3E(X_3) + 4E(X_4) + 5E(X_5)\right]$$

$$= \frac{1}{15}\left[\mu + 2\mu + 3\mu + 4\mu + 5\mu\right] = \frac{1}{15}(15\mu)$$

$$E(T_2) = \mu$$

Hence, both estimators $T_1$ and $T_2$ are unbiased estimators of μ.

**3(b)** Let $\overline{X}$ and $\tilde{X}$ be the sample mean and sample median respectively. We have seen in Unit 2 that if we draw samples from the population whose mean is μ and variance $\sigma^2$ then the sampling distribution of mean has mean μ and variance $\sigma^2/n$.

Therefore,

$$E(\overline{X}) = \mu \text{ and } Var(\overline{X}) = \frac{\sigma^2}{n}$$

Hence, the sample mean is an unbiased estimator of the population mean.

Similarly, the sampling distribution of the median has mean μ and variance $\frac{\pi}{2}\frac{\sigma^2}{n}$. Therefore,

$$E(\tilde{X}) = \mu \text{ and } Var(\tilde{X}) = \frac{\pi\sigma^2}{2n}$$

Hence, the sample median is also an unbiased estimator of the population mean.

**4.** We have to show that $(\overline{X} - \overline{Y})$ is an unbiased estimator for $(\mu_1 - \mu_2)$, therefore, we have to find $E(\overline{X} - \overline{Y})$ and check whether it is equal to $(\mu_1 - \mu_2)$ or not. Thus, we consider

$$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_1 - \mu_2 \quad \left[\because E(aX - bY) = aE(X) - bE(Y)\right]$$

Hence, the estimator $(\overline{X} - \overline{Y})$ is an unbiased estimator for $(\mu_1 - \mu_2)$.

We now find the estimate of the same using the given data. Since estimate is the value of the estimator, therefore, we find the mean of each group as

| Placebo Group (X) | Treatment Group (Y) | $X^2$ | $Y^2$ |
|---|---|---|---|
| 140 | 130 | 19600 | 16900 |
| 165 | 135 | 27225 | 18225 |
| 170 | 140 | 28900 | 19600 |
| 140 | 130 | 19600 | 16900 |
| 135 | 120 | 18225 | 14400 |
| 170 | 120 | 28900 | 14400 |
| 165 | 132 | 27225 | 17424 |
| 150 | 118 | 22500 | 13924 |
| 140 | 120 | 19600 | 14400 |
| **1375** | **1145** | **211775** | **146173** |

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i = \frac{1375}{9} = 152.78$$

$$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i = \frac{1145}{9} = 127.22$$

Thus, we find the estimate of the parameter $(\mu_1 - \mu_2)$ as

$$\bar{X} - \bar{Y} = 152.78 - 127.22 = 25.56.$$

We now find the variance and standard deviation of $(\bar{X} - \bar{Y})$ as

$$\text{Var}(\bar{X} - \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

And standard deviation

$$\text{SD}(\bar{X} - \bar{Y}) = \sqrt{\text{Var}(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

By the definition of standard error, we can find it as

$$\text{SE}(\bar{X} - \bar{Y}) = \text{SD}(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

We can compute the estimate of standard error as

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Therefore, we first calculate $S_1^2$ and $S_2^2$ as

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 = \frac{1}{n_1 - 1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

$$= \frac{1}{8}(211775 - 9 \times 152.78 \times 152.78) = 212.43$$

$$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2 = \frac{1}{n_2 - 1}\left(\sum_{i=1}^{n_2} Y_i^2 - n\bar{Y}^2\right)$$

$$= \frac{1}{8}\left(146173 - 9 \times 127.22 \times 127.22\right) = 63.58$$

Finally, we can compute the estimate of the standard error as

$$SE\left(\bar{X} - \bar{Y}\right) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{212.43}{9} + \frac{63.58}{9}} = \sqrt{30.67} = 5.54$$

Note that $S^2$ is an unbiased estimator of $\sigma^2$ however, S is not an unbiased estimator of σ. Similarly, $S_1/S_2$ is not an unbiased estimator of $\sigma_1/\sigma_2$.

An estimate of $\sigma_1/\sigma_2$ is $S_1/S_2$ (this is a biased estimate)

$$\frac{S_1}{S_2} = \sqrt{\frac{S_1^2}{S_2^2}} = \sqrt{\frac{212.43}{63.58}} = 1.83$$

## Terminal Questions (TQs)

**1.** We have to show that $\bar{X}$ is an unbiased estimator for θ, therefore, we have to find $E(\bar{X})$ and check whether it is equal to θ or not. Here, we are given that

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}; \quad x > 0, \theta > 0$$

It is negative exponential distribution with parameter θ and know the mean of this distribution as

$$E\left(X\right) = \theta$$

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from the same population having mean θ, therefore,

$$E\left(X_1\right) = E\left(X_2\right) = \ldots = E\left(X_n\right) = E\left(X\right) = \theta$$

Therefore, we consider

$$E\left(\bar{X}\right) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \qquad \left[\text{By defination of sample mean}\right]$$

$$= \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right] \quad \left[\begin{array}{l}\because E\left(aX + bY\right) \\ = aE\left(X\right) + bE\left(Y\right)\end{array}\right]$$

$$= \frac{1}{n}(\underbrace{\theta + \theta + \ldots + \theta}_{\text{n-times}}) = \frac{1}{n}\left(n\theta\right) = \theta$$

Thus, the sample mean is an unbiased estimator of the parameter θ.

2. Here, the population consists of three LED televisions whose lives are 8, 6 and 10 years so we can find the population mean and variance as

$$\mu = \frac{8 + 6 + 10}{3} = 8$$

$$\sigma^2 = \frac{1}{3}\left[\left(8-8\right)^2 + \left(6-8\right)^2 + \left(10-8\right)^2\right] = \frac{8}{3} = 2.67$$

Here, we are given that

N = 3 and n = 2

Therefore, the possible numbers of samples (with replacement) that can be drawn from this population are $N^n = 3^2 = 9$. For each of these 9 samples, we will calculate the values of $\bar{X}, S'^2$ and $S^2$ by the formulae given below:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ S'^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \text{and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

and the necessary calculations for these results are shown in the following table:

**Calculation for $\bar{X}$, $S'^2$ and $S^2$**

| Sample | Sample Observations | $\bar{X}$ | $\sum_{i=1}^{2}(X_i - \bar{X})^2$ | $S'^2$ | $S^2$ |
|---|---|---|---|---|---|
| 1 | 8, 8 | 8 | 0 | 0 | 0 |
| 2 | 8, 6 | 7 | 2 | 1 | 2 |
| 3 | 8,10 | 9 | 2 | 1 | 2 |
| 4 | 6, 8 | 7 | 2 | 1 | 2 |
| 5 | 6, 6 | 6 | 0 | 0 | 0 |
| 6 | 6, 10 | 8 | 8 | 4 | 8 |
| 7 | 10, 8 | 9 | 2 | 1 | 2 |
| 8 | 10, 6 | 8 | 8 | 4 | 8 |
| 9 | 10, 10 | 10 | 0 | 0 | 0 |
| **Total** | | **72** | | **12** | **24** |

We calculate $\bar{X}, S'^2$ and $S^2$ as

$$\bar{X}_1 = \frac{1}{2}(8+8) = 8,$$

$$\bar{X}_2 = \frac{1}{2}(8+6) = 7,\ldots,$$

$$\bar{X}_9 = \frac{1}{2}(10+10) = 10$$

$$S_1'^2 = \frac{1}{2}\left[(8-8)^2 + (8-8)^2\right] = 0,$$

$$S_2'^2 = \frac{1}{2}\left[(8-7)^2 + (6-7)^2\right] = 1,\ldots,$$

$$S_9'^2 = \frac{1}{2}\left[(10-10)^2 + (10-10)^2\right] = 0$$

$$S_1^2 = \frac{1}{2-1}\left[(8-8)^2 + (8-8)^2\right] = 0,$$

$$S_2^2 = \frac{1}{2-1}\left[(8-7)^2 + (6-7)^2\right] = 2,\ldots,$$

$$S_9^2 = \frac{1}{2-1}\left[(10-10)^2 + (10-10)^2\right] = 0$$

From the above table, we have

$$E\left(\bar{X}\right) = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_i = \frac{1}{9}\times 72 = 8 = \mu$$

Hence, the sample mean is an unbiased estimator of the population mean.

Also

$$E\left(S'^2\right) = \frac{1}{k}\sum_{i=1}^{k}S_i'^2 = \frac{1}{9}\times 12 = 1.33 \neq \sigma^2$$

Therefore, $S'^2$ is not an unbiased estimator of $\sigma^2$ whereas,

$$E\left(S^2\right) = \frac{1}{k}\sum_{i=1}^{k}S_i^2 = \frac{1}{9}\times 24 = 2.67 = \sigma^2$$

Hence, the estimator $S^2$ is an unbiased estimator of parameter $\sigma^2$.