QUESTIONS

1. What is analysis of variance technique ? Explain its basic assumptions and uses.
2. (a) Discuss the assumptions of Analysis of Variance test (or techniques)
   (b) Distinguish between one-way and two-way ANOVA technique.
3. Discuss the technique of analysis of variance with an illustration for one-way classifcation.
4. Describe the technique of ANOVA for two-way classification.
5. What do you understand by analysis of variance? Explain the assumptions in an analysis of variance.
6. What is analysis of variance problem? Comment on the variance between the samples and within the samples.
7. What are the objectives, assumptions and uses of analysis of variance?
8. What is analysis of variance? Mention its applications.
9. Explain the meaning and significance of ANOVA. How is an ANOVA table set up and how a test is performed.

lysis of Variance

uses.

classifcation.

n an analysis

amples and

et up and

7

Statistical Estimation Theory

INTRODUCTION :

Very often, we will need to make an estimate of the population parameter from the sample statistic. For example, suppose we are to find out the average amount of Pepsi Cola drunk by the students in Kurukshetra University, Kurukshetra, per day. It is difficult to find out the average of all the students and hence what usually done is, a sample taken and the average amount of Pepsi Cola drunk is found out. This sample mean is then used to find the average of the population. In fact, we estimate the population average on the basis of sample average. The theory of estimation deals with the estimation of the unknown population parameters (such as population mean and variance) from the corresponding sample statistics (such as sample mean and variance). **Statistical estimation in a procedure of estimating the unknown population parameters from the correspondiong sample statistics.**

SOME IMPORTANT TERMS

The following terms are used in the study of statistical estimation :

(1) Estimators and Estimates : Generally for the purpose of estimating a population parameter we can use various sample statistics. Those sample statistics (such as sample mean $\overline{X}$, sample median M, sample variance σ^2, etc.) which are used to estimate the unknown population parameters (such as population mean μ, population variance σ^2, etc.) are called estimators and the actual value taken by the estimators are called estimates. If θ (read as theta) denotes the parameters to be estimated, then its estimator will be denoted by $\hat{\theta}$ (read as theta hat). Thus, $\hat{\theta}$ is an estimator of the population parameter θ.

(2) Point Estimate and Interval Estimate : An estimate of the population parameter can be done in two ways :

(i) Point Estimate : A single value of a statistic that is used to estimate the unknown population parameter is called a point estimate. For example, the sample mean $\overline{X}$ which we use for estimating the population mean μ is a point estimator of μ. Similarly, the statistic s^2 is a point estimator of σ^2, where the value of s^2 is computed from a random sample. The point estimate is a single point on the real number scale and hence the name point estimator.

(ii) Interval Estimate : An interval estimate refers to the probable range within which the real value of a parameter is expected to lie. The two extreme limits of such a range are called **fiducial or confidence limits** and the range is called a **confidence interval.** These are determined on the basis of sample studies of a population. Thus, on the basis of sample studies when we estimate that the average monthly expenditure of students staying in a certain hostel is between Rs. 1000 and Rs. 2000, it will be a case of interval estimate and the figures of 1000 and 2000 will be the two estreme limits within which the actual expenditure of the students would lie.

PROPERTIES OF A GOOD ESTIMATOR

There can be more than one estimators of a population parameter. For example, the population mean (μ) may be estimated either by sample mean ($\overline{X}$) or by sample median (M) or by sample mode (Z), etc. Similarly, the population variance (σ^2) may be estimated either by the sample variance (s^2), sample S.D. (s), sample mean deviation, etc. Therefore, it becomes necessary to determine a good estimator out of a number of available estimators. **A good estimator is one which is as close to the true value of the parameter as possible.** A good estimator possess the following characteristics or properties:

(1) Unbiasedness
(2) Consistency
(3) Efficiency
(4) Sufficiency

Let us consider them in detail :

(1) Unbiased Estimator : An estimator $\hat{\theta}$ is said be unbiased estimator of the population parameter θ if the mean of the sampling distribution of the estimator $\hat{\theta}$ is equal to the corresponding population parameter θ. Symbolically,

$$\mu_{\hat{\theta}} = \theta$$

In terms of mathematical expectation, $\hat{\theta}$ is an unbiased estimator of θ if the expected value of the estimator is equal to the parameter being estimated. Symbolically,

$$E(\hat{\theta}) = \theta$$

Example 1. **Sample mean $\overline{X}$ is an unbiased estimate of the population mean μ because the mean of the sampling distribution of the means $\mu_{\overline{X}}$ or $E(\overline{X})$ is equal to the pouplation mean μ. Symbolically,**

$$\mu_{\overline{X}} = \mu \quad \text{or} \quad E(\overline{X}) = \mu$$

Example 2. **Sample variance s^2 is a biased estimate of the population variance σ^2 because the mean of the sampling distribution of variance is not equal to the population variance. Symbolically,**

$$\mu_s^2 \neq \sigma^2 \quad \text{or} \quad E(s^2) \neq \sigma^2$$

However, the modified sample variance ($\hat{s}^2$) is unbiased estimate of the population variance σ^2 because

$$E(\hat{s}^2) = \sigma^2 \qquad \text{where, } \hat{s}^2 = \frac{n}{n-1} \times s^2$$

Example 3. **Sample proportion p is an unbiased estimate of the population proportion P because the mean of the sampling distribution of proportion is equal to the population proportion. Symbolically,**

$$\mu_p = P \quad \text{or} \quad E(P) = P$$

(2) Consistent Estimator : An estimator is said to be consistent if the estimator approaches the population parameter as the sample size increases. In other words, an estimator $\hat{\theta}$ is said to be consistent estimator of the population parameter θ, if the probability that $\hat{\theta}$ approaches θ is 1 as n becomes larger and larger. Symbolically,

$$P(\hat{\theta} \to \theta) \to 1 \quad \text{as} \quad n \to \infty$$

Note : A consistent estimator need not to be unbiased

A sufficient condition for the consistency of an estimator is that

(i) $E(\hat{\theta}) \to \theta$

(ii) $\mathrm{Var}(\hat{\theta}) \to 0$ as $n \to \infty$

Example 1 : **Sample mean $\overline{X}$ is a consistent estimator of the population mean μ because the expected value of the sample mean approaches the population mean and the variance of the sample mean approaches zero as the size of the sample is sufficiently increased. Symbolically,**

(i) $E(\overline{X}) \to \mu$

(ii) $\mathrm{Var}(\overline{X}) = \dfrac{\sigma^2}{n} \to 0$ as $n \to \infty$

Example 2 : **Sample median is also consistent estimator of the population mean because :**

(i) $E(M) \to \mu$

(ii) $\mathrm{Var}(M) \to 0$ as $n \to \infty$

(3) Efficient Estimator : Efficiency is a relative term. Efficiency of an estimator is generally defined by comparing it with another estimator. Let us to take two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$. The estimator $\hat{\theta}_1$ is called an efficient estimator of θ if the variance of $\hat{\theta}_1$ is less than the variance of $\hat{\theta}_2$. Symbolically,

$$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$$

Then, $\hat{\theta}_1$ is called an efficient estimator.

Example : **Sample mean $\overline{X}$ is an unbiased and efficient estimator of the population mean (or true mean) than the sample median M because the variance of the sampling distribution of the means is less than the variance of the sampling distribution of the medians.**

The relative efficiency of the two unbiased estimators is given below :

We know that, $\mathrm{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$, $\mathrm{Var}(M) = \dfrac{\pi}{2} \cdot \dfrac{\sigma^2}{n}$

$$\text{Efficiency} = \frac{\mathrm{Var}(\overline{X})}{\mathrm{Var}(M)} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = \frac{2}{\pi} = \frac{14}{22} = \frac{7}{11} = 0 \cdot 64 \quad \left[\because \pi = \frac{22}{7}\right]$$

$$\therefore \quad \mathrm{Var}(\overline{X}) = 0 \cdot 64\, \mathrm{Var}(M)$$

Therefore, sample mean $\overline{X}$ is 64% more efficiency than the sample median. Hence, the sample mean is more efficient estimator of the population mean as compared to sample median.

(4) Sufficient Estimator : The last property that a good estimator should possess is sufficiency. An estimator $\hat{\theta}$ is said to be a 'sufficient estimator' of a parameter θ if it contains all the informations in the sample regarding the parameter. In other words, a sufficient estimator utilises all informations that the given sample can furnish about the population. Sample means $\overline{X}$ is said to be a sufficient estimator of the population mean.

4. APPLICATION OF POINT ESTIMATION

The applications relating to point estimation are studied under two headings :

(1) Point Estimation in case of Single Sampling

(2) Point Estimation in case of Repeated Sampling.

(1) Point Estimation in case of Single Sampling : When a single independent random sample is drawn from a unknown population, the point estimate of the population parameter can be illustrated by the following examples :

Example 1. **A sample of 10 measurements of the diameter of a sphere gave a mean $\overline{X} = 4{\cdot}38$ inches and a standard deviation $= {\cdot}06$ inches. Determine the unbiased and efficient estimates of (a) the true mean (*i.e.,* population mean) and (b) the true variance (*i.e.,* population variance).**

Solution. We are given : $n = 10, \overline{X} = 4 \cdot 38, s = {\cdot}06$

(a) The unbiased and efficient estimate of the true mean μ is given by :

$$\overline{X} = 4 \cdot 38$$

(b) The unbiased and efficient estimate of the true variance σ^2 is :

$$\hat{s}^2 = \frac{n}{n-1} \cdot s^2$$

Putting the values, we get

$$\hat{s}^2 = \frac{10}{10-1} \times {\cdot}06 = 1 \cdot 11 \times 0 \cdot 06 = {\cdot}066$$

Thus, $\mu = 4 \cdot 38, \sigma^2 = 0 \cdot 666$

Example 2. **The following five observations constitute a random sample from an unknown population :**

6·33, 6·37, 6·36, 6·32 and 6·37 centimeters.

Find out unbiased and efficient estimates of (a) true mean, and (b) true variance.

Solution. (a) The unbiased and efficient estimate of the true mean (*i.e.,* population mean) is given by the value of

$$\overline{X} = \frac{\Sigma X}{n} = \frac{6 \cdot 33 + 6 \cdot 37 + 6 \cdot 36 + 6 \cdot 32 + 6 \cdot 37}{5} = \frac{31 \cdot 75}{5} = 6 \cdot 35$$

(b) The unbiased and efficient estimate of the true variance (*i.e.,* population variance) is :

$$\hat{s}^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} \qquad \text{where, } \hat{s}^2 = \text{modified sample variance.}$$

$$= \frac{(6.33-6.35)^2+(6.37-6.35)^2+(6.36-6.35)^2+(6.32-6.35)^2+(6.37-6.35)^2}{5-1}$$

$$= \frac{.0022}{4} = .00055 \text{ cm}^2$$

Example 4. **The following data relate to a random sample of 100 students in Kurukshetra University classified by their weights (kg) :**

Weight (kg) :	60–62	63–65	66–68	69–71	72–74
No. of Students :	5	18	42	27	8

Determine unbiased and efficient estimates of (a) population mean and (b) population variance.

Solution.

Calculation of Mean and Variance

Weight	No. of Students (f)	M.V. (m)	$A = 67$ $d = m - A$	$d' = d/3$	fd'	fd'^2
60–62	5	61	− 6	− 2	−10	20
63–65	18	64	− 3	− 1	− 18	18
66–68	42	67	0	0	0	0
69–71	27	70	+ 3	+ 1	+ 27	27
72–74	8	73	+ 6	+ 2	+ 16	32
	$n = 100$				$\Sigma fd' = 15$	$\Sigma fd'^2 = 97$

(a) The unbiased and efficient estimate of the population mean is given by the value :

$$\bar{X} = A + \frac{\Sigma fd'}{n} \times i$$

$$= 67 + \frac{15}{100} \times 3 = 67 + (0.45) = 67.45$$

(b) The unbiased and efficient estimate of the population variance is :

$$\hat{s}^2 = \frac{n}{n-1} \cdot s^2$$

where,

$$s^2 = \frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2 \times i^2$$

$$= \left[\frac{97}{100} - \left(\frac{15}{100}\right)^2\right] \times 3^2$$

$$= [0.97 - .0225] \times 9 = 8.5275$$

Now, $$\hat{s}^2 = \frac{n}{n-1}s^2 = \frac{100}{99} \times 8 \cdot 5275 = 8 \cdot 6136$$

Thus, $\mu = 67 \cdot 45, \sigma^2 = 8 \cdot 6136$

(2) Point Estimation in Case of Repeated Sampling : When large number of random samples of same size are drawn from the population with or without replacement, then the point estimates of the population parameter can be illustrated by the following examples :

Example 4. **A population consists of five values : 3, 4, 5, 6 and 7. List all possible samples of size 3 without replacement from this population and calculate the mean $\overline{X}$ of each sample. Verify that sample mean $\overline{X}$ is an unbiased estimate of the population mean.**

Solution. The population consists of the five values : 3, 4, 5, 6, 7. The total number of possible samples of size 3 without replacement are $5_{c_3} = 10$ which are shown in the following table :

Sample No. (1)	Sample Values (2)	Sample Mean ($\overline{X}$) (3)
1	(3, 4, 5)	$\frac{1}{3}(3+4+5) = \frac{12}{3} = 4$
2	(3, 4, 6)	$\frac{1}{3}(3+4+6) = \frac{13}{3} = 4 \cdot 33$
3	(3, 4, 7)	$\frac{1}{3}(3+4+7) = \frac{14}{3} = 4 \cdot 67$
4	(3, 5, 6)	$\frac{1}{3}(3+5+6) = \frac{14}{3} = 4 \cdot 67$
5	(3, 5, 7)	$\frac{1}{3}(3+5+7) = \frac{15}{3} = 5 \cdot 0$
6	(3, 6, 7)	$\frac{1}{3}(3+6+7) = \frac{16}{3} = 5 \cdot 33$
7	(4, 5, 6)	$\frac{1}{3}(4+5+6) = \frac{15}{3} = 5 \cdot 00$
8	(4, 5, 7)	$\frac{1}{3}(4+5+7) = \frac{16}{3} = 5 \cdot 33$
9	(4, 6, 7)	$\frac{1}{3}(4+6+7) = \frac{17}{3} = 5 \cdot 67$
10	(5, 6, 7)	$\frac{1}{3}(5+6+7) = \frac{18}{3} = 6 \cdot 00$
Total	$k = 10$	$\Sigma\overline{X} = 50$

$$\text{Mean of Sampling Distribution of Means} = \mu_{\overline{X}} = \frac{\Sigma\overline{X}}{k} = \frac{50}{10} = 5.$$

$$\text{Population Mean} = \mu = \frac{3+4+5+6+7}{5} = 5$$

Since, $\mu_{\overline{X}} = \mu$, sample mean $\overline{X}$ is an unbiased estimate of the population mean μ.

Example 5. **Consider a hypothetical population comprising three values : 1, 2, 3. Draw all possible samples of size 2 with replacement. Calculate the mean $\overline{X}$ and variance**

s^2 for each sample. Examine whether the two statistics ($\overline{X}$ and s^2) are unbiased and efficient for the corresponding parameters.

Solution. The population consists of three values : 1, 2 and 3. The total number of possible samples of size 2 with replacement are $N^n = 3^2 = 9$ which are given by

Sample No.	Sample Values	Sample Mean ($\overline{X}$)	Sample Variance $s^2 = \frac{1}{2}[(x_1-\bar{x})^2+(x_2-\bar{x})^2]$	Modified Sample Variance $\left(\hat{s}^2 = \frac{n}{n-1}s^2\right)$
1.	(1, 1)	$\frac{1}{2}(1+1)=1\cdot0$	$\frac{1}{2}\cdot[(1-1)^2+(1-1)^2]=0\cdot00$	0·00
2.	(1, 2)	$\frac{1}{2}(1+2)=1\cdot5$	$\frac{1}{2}\cdot[(1-1\cdot5)^2+(2-1\cdot5)^2]=0\cdot25$	0·50
3.	(1, 3)	$\frac{1}{2}(1+3)=2\cdot0$	$\frac{1}{2}\cdot[(1-2)^2+(3-2)^2]=1\cdot00$	2·00
4.	(2, 1)	$\frac{1}{2}(2+1)=1\cdot5$	$\frac{1}{2}\cdot[(2-1\cdot5)^2+(1-1\cdot5)^2]=0\cdot25$	0·5
5.	(2, 2)	$\frac{1}{2}(2+2)=2\cdot0$	$\frac{1}{2}\cdot[(2-2)^2+(2-2)^2]=0\cdot00$	0·00
6.	(2, 3)	$\frac{1}{2}(2+2)=2\cdot5$	$\frac{1}{2}[(2-2\cdot5)^2+(3-2\cdot5)^2]=0\cdot25$	0·50
7.	(3, 1)	$\frac{1}{2}(3+1)=2\cdot0$	$\frac{1}{2}\cdot[(3-2)^2+(1-2)^2]=1\cdot00$	2·00
8.	(3, 2)	$\frac{1}{2}(3+2)=2\cdot5$	$\frac{1}{2}[(3-2\cdot5)^2+(2-2\cdot5)^2]=0\cdot25$	0·50
9.	(3, 3)	$\frac{1}{2}(3+3)=3\cdot0$	$\frac{1}{2}[(3-3)^2+(3-3)^2]=0\cdot00$	0·00
Total	$k=9$	$\Sigma\overline{X}=18$		$\Sigma\hat{s}^2=6$

(a) Mean of Sampling Distribution of Means $= \mu_{\bar{x}} = \dfrac{\Sigma \overline{X}}{k} = \dfrac{18}{9} = 2$. Here, K = No. of samples.

$$\text{Population Mean } \mu = \frac{1+2+3}{3} = 2.$$

Since, $\mu_{\overline{X}} = \mu$, sample mean $\overline{X}$ is an unbiased estimate of the population mean μ.

(b) Mean of the Sampling Distribution of Variance $= \mu_{s^2} = \dfrac{\Sigma s^2}{k} = \dfrac{3}{9} = \dfrac{1}{3}$

$$\text{Population Variance } \sigma^2 = \frac{(1-2)^2+(2-2)^2\ (3-2)^2}{3} = \frac{2}{3}$$

Since, $\mu_{s^2} \neq \sigma^2$, sample variance s^2 is not an unbiased estimate of the population variance (σ^2).

But the modified sample variance defined as $\hat{s}^2 = \dfrac{n}{n-1}s^2$ will be unbiased estimate of the population variance σ^2 because :

$$\mu_{\hat{s}^2} = \frac{\Sigma \hat{s}^2}{k} = \frac{6}{9} = \frac{2}{3}$$

$$\sigma^2 = \frac{2}{3}$$

$$\therefore \quad \mu_{\hat{s}^2} = \sigma^2$$

Since, $\mu_{\hat{s}^2} = \sigma^2$, the modified sample variation is an unbiased estimate of the population variance.

Example 6. **Show that the sample mean $(\overline{X})$ is an unbiased estimate of the population mean.**

or

An independent random sample $x_1, x_2, x_3, \ldots, x_n$ is drawn from a population with mean μ. Prove that the expected value of the sample mean $\overline{X}$ equals the population mean μ.

Solution. A random sampling is one where each sample has an equal chance of being selected. We draw a random sample of size 'n'.

Then,

$$E(\bar{x}) = E\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right] \quad \text{Where } x_1 \text{ is the sample observation.}$$

$$= \frac{1}{n} \cdot [E(x_1) + E(x_2) + \ldots E(x_n)]$$

Now the expected values of x_i (a member of the population) is population mean μ.

$$\therefore \quad E(\bar{x}) = \frac{1}{n}[\mu + \mu + \ldots + \mu] \qquad [\because E(x_1) = E(x_2) = \ldots E(x_n) = \mu]$$

$$= \frac{1}{n} \cdot [n\,\mu] = \mu \qquad [\because \Sigma C = C_1 + C_2 + \ldots C_n = nC]$$

Thus, sample mean $\overline{X}$ is an unbiased estimate of population mean.

EXERCISE – 1

1. Measurements of a sample of masses were determined to be $8\cdot3, 10\cdot6, 9\cdot7, 8\cdot8, 10\cdot2$ and $9\cdot4$ kilograms (kg) respectively. Determine unbiased and efficient estimates of (a) the population mean and (b) the population variance, and (c) compare the sample standard deviation and estimated population S.D. **[Ans.** (a) 9·5 (b) ·736 (c) $\hat{s} = \sigma = 0\cdot86$, $s = \cdot78$]
2. A random samples of 9 individuals has the following heights in inches : 45, 47, 50, 52, 48, 47, 49, 53 and 51. Find the unbiased and efficient estimate of (a) true mean. (b) true variance. **[Ans.** (a) 49·11 (b) 6·91]
3. A population consists of four numbers : 3, 4, 2, 5. List all possible distinct samples of size two which can be drawn without replacement and verify that the population mean is equal to the mean of sample means.
4. A population consists of three numbers : 2, 5 and 8. List all possible distinct samples of size two which can be drawn without replacement from this population. Calculate the mean $\bar{x}$ for each sample. Verify that sample mean $\bar{x}$ is an unbiased estimate of the population mean.

5. A population consists of five numbers : 2, 3, 6, 8, 11. List all possible samples of size 2 which can be drawn with replacement from this population. Calculate the mean $\bar{X}$ and variance $s^2 = \frac{1}{2}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2]$ for each sample. Examine whether the two statistics are unbiased for the corresponding population parameters. What is the sampling variance of $\bar{X}$?

6. A sample of 10 television tubes produced by a company showed a mean life of 1200 hr. and a standard deviation of 10 hr. Find the unbiased and efficient estimates of the (a) population mean and (b) population variance. **[Ans.** $\mu = 1200$ h. $\hat{s}^2 = 111 \cdot 11$]

INTERVAL ESTIMATION (OR CONFIDENCE INTERVAL)

A point estimator, however, good it may be, cannot be expected to coincide with the true value of the parameter and in some cases may differ widely from it. In the theory of interval estimation, we find an interval or two numbers within which the value of unknown population parameter is expected to lie with a specified probability. The method of interval estimation consists in the determination of two constant t_1 and t_2 such that P [$t_1 < \theta < t_2$ for given value of t] $= 1 - \alpha$, where α is the level of significance. The interval $[t_1, t_2]$ within which the unknown value of parameter θ is expected to lie is known as confidence interval and the limits t_1 and t_2 so determined are known as confidence limits and $1 - \alpha$ is called the confidence coefficient, depending upon the desired precision of the estimate. For example, $\alpha = 0 \cdot 05$ (or $0 \cdot 01$) gives 95% (or 99%) confidence limits.

Procedure for Setting up Confidence Interval (or Interval Estimation) or Limits for a Population Parameter

The following steps enable us to compute the confidence interval or confidence limits for the population parameter θ in terms of the sample statistic t :

(1) Compute or take the appropriate sample statistic t.

(2) Obtain the S.E. (t), the standard error of the sample statistic t.

(3) Select the confidence level and corresponding to that specified level of confidence, we note down the critical value of the statistic t.

Applications of Interval Estimation (or Confidence Interval)

The applications relating to interval estimation (or confidence interval) are studied under the following heads :

(A) Interval Estimation for Large Samples ($n > 30$)

(B) Inteval Estimation for Small Samples ($n \leq 30$)

Let us discuss them :

(A) Interval Estimation (or Confidence Interval) for Large Samples ($n > 30$) : In large sample ($n > 30$), the interval estimation is further studied under the following heads :

(1) Confidence Interval or Limits for Population Mean

(2) Confidence Interval or Limits for Population Proportion

(3) Confidence Interval or Limits for Population Standard Deviation

(4) Determination of a Proper Sample Size for Estimating μ or P.

(i) Confidence Interval or Limits for Population Mean μ (when $n > 30$) : The determination of the confidence interval or limits for the population mean μ in case of large sample ($n > 30$) requires the use of normal distribution.

(i) $(1-\alpha)$ 100% Confidence limits for μ are given by :

$$\overline{X} \pm Z_{\alpha/2} \cdot S.E_{\overline{X}}$$

$$\overline{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{where, } \sigma \text{ is known.}$$

or

$$\overline{X} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{when } \sigma \text{ is not known. [For large sample, } \sigma \approx s]$$

or

(ii) $(1-\alpha)$ 100% confidence interval for μ is given by :

$$\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\overline{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{where, } \sigma \text{ is not known.}$$

or [For large sample, $\sigma = s$]

In particular, 95% confidence limits for μ are :

$$\overline{X} \pm 1{\cdot}96 \frac{\sigma}{\sqrt{n}}$$

Similary, 99% confidence limits for μ are

$$\overline{X} \pm 2{\cdot}58 \frac{\sigma}{\sqrt{n}}$$

Procedure: The construction of confidence interval for population mean μ involves the following steps :

(i) Compute $\overline{X}$ or take $\overline{X}$

(ii) Compute the $S.E_{\overline{X}}$ by using the following formula :

(a) $S.E_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, when σ is known.

(b) $S.E_{\overline{X}} = \frac{s}{\sqrt{n}}$, when σ is not known.

(iii) Select the desired confidence level and corresponding to that level of confidence, we find that value of $Z_{\alpha/2}$.

(iv) Substituting the value of $\overline{X}$, $S.E_{\overline{X}}$ and $Z_{\alpha/2}$ in the above stated formula.

Note :

1. If the population S.D. is not known, the sample S.D. (s) is used for large samples.
2. The values of $Z_{\alpha/2}$ (for large samples) corresponding to various level of confidence are given below :

Confidence Level $(1-\alpha)$ 100%	90%	95%	96%	98%	99%	Without any reference to the confidence level
Z-Value	± 1·64	± 1·96	± 2·06	± 2·33	± 2·58	± 3

For other confidence level, the values of $Z_{\alpha/2}$ can be found from the tables of area under the normal curve given at the end of the book.

Note : Where no reference to the confidence interval is given, then we always $Z_{\alpha/2} = 3$. This e corresponds to 99·73% level of confidence.

The following examples illustrate the procedure for setting up confidence limits for μ :

Example 7. **A random sample of 100 observations yields sample mean $\overline{X} = 150$ and sample variance $s^2 = 400$. Compute 95% and 99% confidence interval for the population mean.**

Solution. We are given : $n = 100, \overline{X} = 150, s^2 = 400 \Rightarrow s = 20$

$$S.E._{\overline{X}} = \frac{s}{\sqrt{n}} \qquad \text{[For large sample, } \sigma = s]$$

$$= \frac{20}{\sqrt{100}} = 2$$

At 95% confidence level, the value of $Z_{\alpha/2} = 1\cdot96$

At 99% confidence level, the value of $Z_{\alpha/2} = 2\cdot58$

(a) 95% confidence Interval or Limits for μ are :

$$\overline{X} \pm 1\cdot96\, S.E._{\overline{X}}$$

Putting the values, we get

$$150 \pm 1.96 \times 2 = 150 \pm 3.92 = 153.92 \text{ or } 146.08$$

Thus, $146.08 < \mu < 153.92$

(b) 99% confidence interval or Limits for μ are :

$$\bar{x} \pm 258.\, S.E._{\overline{X}}$$

$$= 150 \pm 2\cdot58 \times 2$$

$$= 150 \pm 5\cdot16$$

$$= 155\cdot16 \text{ or } 144\cdot84$$

Thus, $144\cdot84 < \mu < 155\cdot16$

Example 8. **A random sample of 900 workers in a steel plant showed an average height of 67 inches with a standard deviation of 5 inches.**

(a) Establish a 95% confidence interval estimate of the mean height of all the workers at the steel plant.

(b) Establish a 99% confidence interval estimate of the mean height of all the workers at the steel plant.

Solution. We are given : $n = 900, \overline{X} = 67, s = 5$

$$S.E.(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{900}} = 0\cdot167 \qquad \text{[For large sample, } s = \sigma]$$

At 95% confidence level, the value of $Z_{\alpha/2} = 1\cdot96$

At 99% confidence level, the value of $Z_{\alpha/2} = 2\cdot58$

(a) 95% confidence interval for μ is :

$$\overline{X} \pm 1\cdot96, S.E._{\overline{X}}$$

Putting the values, we get

$$67 \pm 1\cdot96 \times (0\cdot167)$$

$$= 67 \pm 0\cdot327 = 67\cdot327 \text{ to } 66\cdot673$$

Thus, $66\cdot673 < \mu < 67\cdot327$

(b) 99% confidence interval for μ are :

$$\overline{X} \pm 2{\cdot}58.\text{S.E}_{\overline{X}}$$

Putting the values, we get

$$= 67 \pm 2{\cdot}58 \cdot (0{\cdot}167)$$
$$= 67 \pm 0{\cdot}43$$
$$= 67{\cdot}43 \text{ to } 66{\cdot}57$$

Thus, $66{\cdot}57 < \mu < 67{\cdot}43$

Example 9. **Upon collecting a sample of 100 from a population with known standard deviation of Rs. 50, the mean is found to be Rs. 500.**

(i) Find 90% confidence interval for the population mean.

(ii) Find 98% confidence interval for the population mean.

Solution. We are given : $n = 100, \overline{X} = 500, \sigma = 50$

$$\text{S.E}_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$$ [Here, σ is known]

At 90% confidence level, the value of $Z_{\alpha/2} = 1{\cdot}64$

At 98% confidence level, the value of $Z_{\alpha/2} = 2{\cdot}33$.

(a) 90% confidence interval for μ is :

$$\overline{X} \pm 1{\cdot}64\,.\,\text{S.E}_{\overline{X}}$$

Putting the values, we get

$$= 500 \pm 1{\cdot}64 \cdot (5)$$
$$= 500 \pm 8{\cdot}2$$
$$= 508{\cdot}2 \text{ to } 491{\cdot}8$$

Thus, $491{\cdot}8 < \mu < 508{\cdot}2$

(b) 98% confidence interval for μ is

$$\overline{X} \pm 2{\cdot}33\,.\,\text{S.E}_{\overline{X}}$$

Putting the values, we get

$$= 500 \pm 2{\cdot}33 \cdot (5)$$
$$= 500 \pm 11{\cdot}65$$
$$= 511{\cdot}65 \text{ to } 488{\cdot}35$$

Thus, $488{\cdot}35 < \mu < 511{\cdot}65$

Confidence Interval or Limits for population mean when sample is drawn without replacement from a finite population. In this case $(1-\alpha)$ 100% confidence interval or limits are given by :

$$\overline{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \text{when } \sigma \text{ is known}$$

or

$$\overline{X} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \text{when } \sigma \text{ is not known.}$$

Where, $\sqrt{\frac{N-n}{N-1}}$ = Finite Population Correction Factor

Example 10. **A random sample of 100 articles selected from a batch of 2000 articles shows that the average diameter of the articles = 0·354 with a standard deviation = 0·048. Find 95% confidence limits for the average of this batch of 2000 articles. [Given Z value with 95% confidence is 1·96]**

Solution. We are given : $N = 2000, n = 100, \overline{X} = 0\cdot354$, and $s = 0\cdot048$

$$S.E._{\overline{X}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \text{[For Large Sample, } \sigma = s\text{]}$$

$$= \frac{0\cdot048}{\sqrt{100}} \times \sqrt{\frac{2000-100}{2000-1}}$$

$$= \frac{0\cdot048}{10} \times \sqrt{\frac{1900}{1999}}$$

$$= 0\cdot0048 \times \sqrt{0\cdot95048} = 0\cdot0048 \times 0\cdot97493$$

$$= 0\cdot00468$$

At 95% confidence level, the value of $Z_{\alpha/2} = 1\cdot96$.

95% confidence limits for μ are :

$$\overline{X} \pm 1\cdot96 \cdot S.E._{\overline{X}}$$

Putting the values, we get

$$= 0\cdot354 \pm 1\cdot96 \times (0\cdot00468)$$

$$= 0\cdot354 \pm 0\cdot009173$$

$$= 0\cdot3448 \text{ to } 0\cdot3632$$

Example 11. **A manager wants an estimate of average sales of salesmen in his company. A random sample of 121 out of 600 salesmen is selected and average sales is found to be Rs. 760. If the population standard deviation is Rs. 150, manager specifies a 99% level of confidence. What is the interval estimate for population mean μ ?**

Solution. We are given : $N = 600, n = 121, \overline{X} = 760, \sigma = 150$

$$S.E._{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \text{[Here, } \sigma \text{ is given]}$$

$$= \frac{150}{\sqrt{121}} \times \sqrt{\frac{600-121}{600-1}}$$

$$= 12\cdot2$$

At 99% confidence level, $Z_{\alpha/2} = 2\cdot58$

99% confidence limits for μ are given by :

$$\overline{X} \pm 2\cdot58 \cdot S.E._{\overline{X}} \qquad \text{[Here, } \sigma \text{ is given]}$$

Putting the values, we get

$$= 760 \pm 2\cdot58 \times 12\cdot2$$

$$= 760 \pm 31\cdot48 = 728\cdot52 \text{ to } 791\cdot48$$

Thus, $728\cdot52 < \mu < 791\cdot48.$

EXERCISE – 2

1. A random sample of 144 observations yields sample mean $\overline{X} = 160$ and sample variance $s^2 = 100$. Compute a 95% confidence interval for population mean.

   [Ans. $158 \cdot 37 < \mu < 161 \cdot 03$]

2. From a random sample of 64 farms are found to have a mean area of 45 hectares with a standard deviation of 12. What are the 95% and 99% confidence limits for the mean area?

   [Ans. (a) $47 \cdot 94$, $42 \cdot 06$ (b) $48 \cdot 87$, $41 \cdot 63$]

3. From a random sample of 100 farms are found to have a mean area of 250 hectares with a standard deviation of 50. Compute 99% confidence interval of the mean area. How does the width of the confidence interval change if the size of the sample were increased to 400?

   [Ans. (a) $237 \cdot 1 < \mu < 262 \cdot 9$ (b) reduced to half]

   [Hint : The width of the confidence interval is inversely related with the size of the sample]

4. Upon collecting a sample of 200 from a population with known standard deviation of $5 \cdot 23$, the mean is found to be $76 \cdot 3$.
   (i) Find 90% confidence interval for the mean.
   (ii) Find 98% confidence interval for the mean.

   [Ans. (i) $75 \cdot 695 < \mu < 76 \cdot 905$ (ii) $75 \cdot 441 < \mu < 77 \cdot 159$]

5. A simple random sample of size 100 has mean 15, the population variance being 25. Find an interval estimate of the population mean with confidence level of (i) 99% and (ii) 95%. If the population variance is not given, then what should be done to find out the required interval estimates.

   [Ans. (i) $14 \cdot 02 < \mu < 15 \cdot 98$, (ii) $13 \cdot 71 < \mu < 16 \cdot 29$; $\sigma^2 = \hat{s}^2 = \frac{n}{n-1} \cdot s^2$]

6. A sample of size 64 was drawn from a population consisting of 128 units. The sample mean of the measurements of a certain characteristics was found to be 28. Set up a 96% confidence limits for the population mean, if it is known that the population. S.D. for the characteristic is 4. [Hints : $S.E_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ For 96% , $Z = 2 \cdot 05$]

   [Ans. $28 \cdot 7267$ and $27 \cdot 272$]

---

(2) Confidence Interval or Limits for Population Proportion P : Though the sampling distribution associated with proportions is the binomial distribution, the normal distribution can be used as an approximation provided the sample is large (*i.e.*, $n > 30$) and both np and $nq \geq 5$ (when n is the size of the sample, p is the proportion of success and $q = 1 - p$).

(1) $(1-a)$ 100% Confidence limits for P are given by :

$$p \pm Z_{\alpha/2} \cdot S.E.(p)$$

or

$$p \pm Z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} \qquad \text{when } P \text{ is known.}$$

or

$$p \pm Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \qquad \text{when } P \text{ is not known.}$$

$\therefore (1-\alpha)$ 100% confidence interval for P is given by

$$p - Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} < P < p + Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

In particular, 95% confidence limits for P are :

$$p \pm 1 \cdot 96 \cdot \sqrt{\frac{pq}{n}}$$

Similarly, 99% confidence limits for P are :

$$p \pm 2 \cdot 58 \cdot \sqrt{\frac{pq}{n}}$$

Procedure : The construction of confidence limits or interval for population proportion involves the following steps :

(i) Compute p or take p.

(ii) Compute the S.E. (p) by using the following formula :

$$\text{S.E. } (p) = \sqrt{\frac{PQ}{n}} \quad \text{when } P \text{ is known.}$$

$$\text{S.E. } (p) = \sqrt{\frac{pq}{n}} \quad \text{when } P \text{ is not known.}$$

(iii) Select the desired confidence level and corresponding to that level, we find the value of $Z_{\alpha/2}$.

(iv) Substituting the values of p, S.E. (p) and $Z_{\alpha/2}$, in the above stated formula.

Note :

1. If the population proportion (P) is not known, then sample proportion (p) is used for large samples.
2. When no reference to the confidence level is given, then always take $Z_{\alpha/2} = 3$ for 99·73% confidence level.

Example 12. **Out of 1,200 tosses of a coin, it gave 480 heads and 720 tails. Find the 95 percent confidence interval for the heads.**

Solution. We are given : $n = 1200$, X = Total heads $(np) = 480$

$$p = \text{Sample proportion heads} = \frac{480}{1200} = 0 \cdot 4$$

Also, the population proportion of heads $= P = 0 \cdot 50$

$$\therefore Q = 1 - P = 1 - 0 \cdot 50 = 0 \cdot 50$$

$$\text{S. E.}_{(p)} = \sqrt{\frac{PQ}{n}} \quad \text{[For large sample, } p = P\text{]}$$

$$= \sqrt{\frac{0 \cdot 5 \times 0 \cdot 5}{1200}} = 0 \cdot 0144$$

For 95% confidence level, the value of $Z_{\alpha/2} = 1 \cdot 96$

95% confidence interval for P is given by :

$$p \pm 1 \cdot 96 \text{ S. E.}_p$$

Putting the values, we get

$$= 0.4 \pm 1.96 \times 0.0144$$
$$= 0.4 \pm 0.028$$
$$= 0.372 \text{ to } 0.428$$

Thus,
$$0.372 < P < 0.428$$

Example 13. **A random sample of 1000 households in a city revealed that 500 of these had Gita. Find 95% and 99% confidence limits for the proportion of households in the city with Gita.**

Solution. We are given : $n = 1000$, $x =$ No. of households having Gita = 500

$$p = \frac{500}{1000} = 0.50$$
$$q = 1 - p = 1 - 0.50 = 0.50$$
$$\text{S.E.}(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.50 \times 0.50}{1000}} = 0.0158$$

For 95% confidence level, the value of $Z_{\alpha/2} = 1.96$

For 99% confidence level, the value of $Z_{\alpha/2} = 2.58$

(a) 95% confidence limits for P are given by :

$$p \pm 1.96 \text{ S.E}_{\overline{X}}$$

Putting the values, we get

$$= 0.50 \pm 1.96 \times 0.0158$$
$$= 0.50 \pm 0.031$$
$$= 0.531 \text{ to } 0.469$$

(b) 99% confidence limits for P are given by :

$$p \pm 2.58 \times \text{S.E}_{\overline{X}}$$

Putting the values, we get

$$= 0.50 \pm 2.58 \times 0.0158$$
$$= 0.50 \pm 0.040$$
$$= 0.054 \text{ to } 0.46$$

Example 14. **A random sample of 600 pineapples was taken from a large consignment and 75 of them were found to be bad. Estimate the proportion of bad apples in the consignment and obtain the standard error of the estimate. Assign the limits within which the percentage of bad pineapples in the consignment lies.**

Solution. We are given : $n = 600$ $x =$ No. of bad pineapples = 75

Sample proportion, $p = \frac{75}{600} = 0.125 = 12.5\%$

$$q = 1 - 0.125 = 0.875$$
$$\text{S.E.}(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.125 \times 0.875}{600}} = 0.013$$

Since, the level of confidence is not specified, we assume it as 99.73%.

For, 99.73% confidence level, the value of $Z_{\alpha/2} = 3$.

99.73% confidence limits for P are given by

$$p \pm 3 \times \text{S.E.}_{\overline{X}}$$

Putting the values, we get

$$= 0.125 \pm 3 \times 0.013$$
$$= 0.125 \pm 0.039$$
$$= 0.164 \pm 0.086$$

Hence, the required percentage lies between 16·4% and 8·6%.

Confidence Interval or Limits for Population Proportion P when the sample is drawn without replacement from a finite population : In this case, $(1-\alpha)$ 100% confidence interval or limits are given by :

$$p \pm Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

where, $\sqrt{\dfrac{N-n}{N-1}}$ = Finite Population Correction Factor

Note : If N is sufficiently large as compared to the sample size n, the finite population correction factor may be ignored.

...ample 15. **Out of 20,000 customers ledger accounts, a sample of 600 accounts was taken to test the accuracy of posting and balancing where in 45 mistakes were found. Assign limits within which the number of defective cases can be expected at 95% level.**

Solution. We are given : $n = 600$, $N = 20{,}000$, x = No. of mistakes in the sample ledger accounts = 45

Sample proportion $\quad p = \dfrac{x}{n} = \dfrac{45}{600} = 0.075$

$$q = 1 - p = 1 - 0.075 - 0.925$$

Since, N is sufficiently large as compared to the sample size n, the finite population correction factor $\sqrt{\dfrac{N-n}{N-1}}$ may be ignored. Hence, assuming it as a sample from finite (large) population, the standard error of p is given by

$$\text{S.E.}\,(p) = \sqrt{\frac{pq}{n}}$$

$$= \sqrt{\frac{0.075 \times 0.925}{600}} = \sqrt{0.0001156}$$

$$= 0.011 \text{ (approx)}$$

For 95% confidence level, the value of $Z_{\alpha/2} = 1.96$.

95% confidence limits for population P are given by :

$$p \pm 1.96\, \text{S.E.}_{\overline{X}}$$

Putting the values, we get

$$= 0.075 \pm 1.96 \times 0.011$$
$$= 0.075 \pm 0.022 = (0.053,\ 0.097)$$

Hence, the number of defective cases in a lot of 20,000 are expected to lie between 20,000 × 0·053 and 20,000 × 0·097 *i.e.*, 1060 and 1940.

Note : If the finite population correction factor is not ignored then; 95% confidence limits for P are :

$$p \pm 1{\cdot}96 \cdot \sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$= 0{\cdot}075 \pm 1{\cdot}96\sqrt{\frac{0{\cdot}075 \times 0{\cdot}925}{600}} \times \sqrt{\frac{20{,}000-600}{20{,}000-1}}$$

$$= 0{\cdot}075 \pm 1{\cdot}96 \times 0{\cdot}0108$$

$$= 0{\cdot}075 \pm 0{\cdot}021168$$

$$= (0{\cdot}0538,\ 0.096168)$$

Hence, the required number of defective cases in the lot lies between 20,000 (0·0538, 0·096168) *i.e.*, 1076 and 1924.

EXERCISE – 3

1. A random sample 300 households in a city revealed that 123 of these houses had Gita. Find a 95 percent confidence interval for the proportion of hoseholds in the city with Gita. [**Ans.** $0{\cdot}355 < P < 0{\cdot}465$]
2. A random sample of 500 houses in a city disclosed that 125 of these houses had colour T.V. sets. Find a 98 percent confidence interval for the proportion of houses in the city with colour T.V. sets. (Table value of Z for 98% confidence level is 2·33). [**Ans.** $0{\cdot}205 < P < 0{\cdot}295$]
3. In a market survey for the introduction of a new product given in a town, a sample of 400 persons was drawn. When they were approached for sale, 80 of them purchased the product. Find a 95% confidence limits for the purchase of persons who would buy the product in the town. [**Ans.** 0·1608, 0·2392]
4. In a large consignment of oranges, a random sample of 100 organs revealed that 5 oranges were bad. Set up 96% confidence limits for the proportion of defective oranges in the whole consignment. [**Ans.** $0{\cdot}005 < P < 0{\cdot}095$]
5. A sample of 500 screws is taken from a large consignment and 65 are found to be defective. Estimate the percentage of defectives in the consignment and assign limits within which the percentage lies. [**Ans.** (a) $0{\cdot}085 < P < 0{\cdot}175$ (b) 8·5% to 17·5%]
6. In a random sample of 1000 civil servants, the proportion of those having favourable reaction to the newly introduced income tax structure was observed to be 0·45. Construct a 95 percent confidence interval of the proportions of all civil servants having favourable section. [**Ans.** $0{\cdot}419 < P < 0{\cdot}48$]
7. A random sample of 700 units from a large consignment showed that 200 were damaged. Find (i) 95 percent and (ii) 99 percent confidence limits for the proportion of damaged units. [**Ans.** (i) $0{\cdot}253 < P < 0{\cdot}319$ (ii) $0{\cdot}242 < P < 0{\cdot}330$]
8. Out of 10,000 customers ledger accounts, a sample of 400 accounts was selected to judge and accuracy of posting and balancing. It contained 40 mistakes. Assign limits within which the number of defective cases could be expected at 95 percent level. [**Ans.** (a) $0{\cdot}071 < P < 0{\cdot}129$ (b) $710 < x < 1290$]

(i) Confidence Interval or Limits for Population Standard Deviation : The determination of confidence interval or limits for population S.D. σ in case of large sample ($n > 30$) requires the use of normal distribution.

(i) $(1-\alpha)$ 100% confidence limits for σ are given by

$$s \pm Z_{\alpha/2} \cdot S.E._s$$

$$s \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{2n}} \qquad \text{when } \sigma \text{ is known.}$$

or

$$s \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{2n}} \qquad \text{when } \sigma \text{ is not known.}$$

or

(ii) **$(1-\alpha)$ 100% confidence interval for σ is given by :**

$$s - Z_{\alpha/2} \cdot \frac{s}{\sqrt{2n}} < \sigma < s + Z_{\alpha/2} \cdot \frac{s}{\sqrt{2n}}$$

In particular, 95% confidence limits for σ are :

$$s \pm 1 \cdot 96 \cdot \frac{s}{\sqrt{2n}} \qquad \text{[For large sample, } s = \sigma]$$

Similarly, 99% confidence limits for σ are :

$$s \pm 2 \cdot 5 \pm \frac{s}{\sqrt{2n}}$$

Procedure : The construction of confidence limits for σ involves the following steps :

(i) Compute s or take s

(ii) Compute S.E. (s) by using the following formula :

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

or

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

(iii) Select the desired confidence level and corresponding to that confidence level, the value of $Z_{\alpha/2}$.

(iv) Substituting the values of s, $Z_{\alpha/2}$ and n in the above stated formula.

Example 16. **A random sample of 50 observations gave a value of its standard deviation equal to 24·5. Construct a 95% confidence interval for population standard deviation σ.**

Solution. We are given : $n = 50$, $s = 24 \cdot 5$

$$S.E.(s) = \frac{s}{\sqrt{2n}} = \frac{24 \cdot 5}{\sqrt{100}} = 2 \cdot 45$$

For 95% confidence level, the values of $Z_{\alpha/2} = 1 \cdot 96$.

95% confidence interval for σ is given by :

$$s \pm 1 \cdot 96 \cdot S.E._s$$

Putting the values we get

$$= 24 \cdot 5 \pm 1 \cdot 96 \times 2 \cdot 45$$

$$= 24 \cdot 5 \pm 4 \cdot 802$$

$$= 29 \cdot 032 \text{ to } 19 \cdot 698$$

Thus, $19 \cdot 698 < \sigma < 29 \cdot 302$

EXERCISE – 4

1. A sample of 100 items gives a standard deviation of 25. Set up the limits for the population standard deviation at 95% level of confidence. [Ans. 21·55, 28·46]
2. A sample of 100 items gives a standard deviation of 4700. Set up the limits for the population standard deviation at 99% confidence level of confidence. [Ans. 5032·4, 4367·60]

(4) Determination of a Proper Sample Size for Estimating μ or P:

So far we have calculated the confidence intervals based on the assumption that the sample size n is known. In most of the practical situation, generally, sample size is not known. The method of determining a proper sample size is studied under two headings :

(a) Sample Size for Estimating a Population Mean

(b) Sample Size for Estimating a Population Proportion

(a) Sample Size for Estimating a Population Mean : In order to determine the sample size for estimating a population mean, the following three factors must be known :

(i) the desired confidence level and the corresponding values of Z.

(ii) the permissible sampling error E.

(iii) the standard deviation σ or an estimate of σ (*i.e.*, $\bar{s}$)

After having known the above mentioned factors, the sample size n is given by :

$$n=\left(\frac{Z\,.\sigma}{E}\right)^2$$

Note :

1. The values of Z and E are predetermined.
2. The population S.D. (σ) may be actual or estimated.

Example 17. **A cigarette manufacturer wishes to use a random sample to estimate the average nicotin content. The sampling error should not be more than one milligram above or below the true mean, with 99 percent confidence level. The population standard deviation is 4 milligram. What sample size should the company use in order to satisfy these requirements ?**

Solution. We are given : $E=1$, $Z_{\alpha/2}=2{\cdot}58$ for 99% confidence level and $\sigma=4$.

Sample size formula is :

$$n=\frac{Z^2\,.\sigma^2}{E^2}$$

Substituting the values, we get

$$n=\frac{(2{\cdot}58)^2\,(4)^2}{1^2}=106{\cdot}50 \text{ or } 107$$

Hence, the required sample size $n=107$ which the company should use for their requirements to be fulfilled.

(b) Sample size for Estimating a Population Proportion : In order to determine the sample size for estimating population proportion, the following three factors must be known :

(i) the desired level of confidence and the corresponding value of Z.

(ii) the permissible sampling error E.

(iii) the actual or estimated true proportion of success P.

The sample size n is given by :

$$n=\frac{Z^2\times PQ}{E^2} \qquad \text{where, } Q=1-P$$

Note :

1. The values of Z and E are predetermined.
2. The value of the population proportion P may be actual or estimated.

Example 18. A firm wishes to determine with a maximum allowable error of 0·05 and a 98 percent level of confidence the proportion of consumer who prefer its product. How large a sample will be required in order to make such an estimate if the preliminary sales reports indicate that 25 percent of all the consumers prefer the firm's product ?

Solution. We are given : $E=0\cdot05$, $P=0\cdot25$, $Q=1-0\cdot25=0\cdot75$, $Z=2\cdot33$ for 98% confidence level.

Sample size formula is

$$n=\frac{Z^2\times PQ}{E^2}$$

Substituting the values, we get

$$n=\frac{(2\cdot33)^2}{(0\cdot05)^2}(0\cdot25)(0\cdot75)$$

$$=\frac{5\cdot4289}{0\cdot0025}(0\cdot1875)=\frac{1\cdot0179}{0\cdot0025}=407\cdot16 \text{ or } 408$$

Hence, the required sample size $n=408$.

EXERCISE – 5

1. A firm wishes to estimate with an error of not more than 0·03 and a level of confidence of 98%, the proportion of consumers that prefers its brand of household detergent. Sales reports indicate that about 0·20 of all consumers prefer the firm's brand. What is the requisite sample size ? **[Ans. $n=965$]**
2. Mr. X wants to determine the average time to complete a certain job. The past records show that population standard deviation is 10 days. Determine the sample size so that Mr. X may be 95% confident that the sample average remains ± 2 days of the average. **[Ans. $n=96$]**
3. In measuring reaction time, a psychologist estimates that the standard deviation is 0·05 seconds. How large a sample of measurements must be taken in order to be 95% confident that the error of his estimate will not be exceeded 0·01 seconds. **[Ans. $n=96$]**

B. INTERVAL ESTIMATION (OR CONFIDENCE INTERVAL) FOR SMALL SAMPLES ($n \le 30$)

The determination of confidence intervals in case of small sized sample ($n \le 30$) is studied under two headings :

(1) **Confidence Interval or Limits for Population Mean μ**

(2) **Confidence Interval or Limits for Population Variance σ^2**

(1) **Confidence Interval or Limits for Population Mean ($n \le 30$) :** When the samples size is small (*i.e.*, $n \le 30$) and σ (the population S.D.) is unknown the desired confidence interval or limits for population mean μ can be found by making use of *t*-distribution. In case of small samples, *t*-values are used in place of Z-values.

(i) $(1-\alpha)$ 100% confidence limits for population Mean μ are given by :

$$\overline{X} \pm t_{\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}} \quad \text{where, } \hat{s} = \text{modified sample S.D.} = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}} \text{ or } \hat{s} = \sqrt{\frac{n}{n-1}}\, s^2$$

(ii) $(1-\alpha)$ 100% confidence interval for μ is given by :

$$\overline{X} \pm t_{\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}} < \mu < \overline{X} \pm t_{\alpha/2} \cdot \frac{\hat{s}}{\sqrt{n}}$$

In particular, 95% confidence limits for μ are given by

$$\overline{X} \pm t_{0.025} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Similarly, 99% confidence limits for μ are given by

$$\overline{X} \pm t_{0.005} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Procedure : The construction of the confidence interval or limits in case of small sample ($n \le 30$) involves the following steps :

(i) Compute $\overline{X}$ or take $\overline{X}$.

(ii) Compute modified sample S.D. using the following formula.

$$\hat{s} = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}}$$

or

$$\hat{s} = \sqrt{\frac{n}{n-1} \cdot s^2} \qquad \text{when } s \text{ is given.}$$

(iii) Compute the degree of freedom (*d.f.*) using the formula :

$$d.f. = v = n-1$$

(iv) Select the desired confidence level and corresponding to that specified level of confidence and for given degrees of freedom, we note the value of the $t_{\alpha/2}$ from the *t*-table.

(v) Substituting the values of $\overline{X}$, $\hat{s}$ and $t_{\alpha/2}$ in the above stated formula.

Example 19. **A random sample of size 16 has 50 as mean with standard deviation of 3. Obtain 98 percent confidence limits of the mean of the population.**

Solution. We are given : $n = 16, \overline{X} = 50, s = 3 \Rightarrow s^2 = 9$

$$\hat{s} = \sqrt{\frac{n}{n-1} \cdot s^2} = \sqrt{\frac{16}{16-1} \times 9} = 3 \cdot 098$$

Degrees of freedom $= v = n = 16 - 1 = 15$

For a 98% confidence level, $\alpha = 0 \cdot 02$ so that $\frac{\alpha}{2} = \frac{0 \cdot 02}{2} = 0 \cdot 01$

Using t-table the value of $t_{.01}$ for 15 d.f. $= 2 \cdot 602$.

98% confidence limits for μ are given by

$$\overline{X} \pm t_{.01} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Putting the values, we get

$$= 50 \pm 2 \cdot 602 \times \frac{3 \cdot 098}{\sqrt{16}}$$

$$= 50 \pm 2 \cdot 015$$

$$= 52 \cdot 015 \text{ to } 47 \cdot 985$$

Example 20. **A random sample of 16 items from a normal population showed a mean of 53 and the sum of squares of deviations from this mean is equal to 150. Obtain 95% and 99% confidence limits for the mean of the population.**

Solution. We are given : $n = 16, \overline{X} = 53, \Sigma(X - \overline{X})^2 = 150$

$$\hat{s} = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}}$$

$$= \sqrt{\frac{150}{16-1}} = \sqrt{\frac{150}{15}} = \sqrt{10} = 3 \cdot 162$$

Degrees of freedom $= v = n - 1 = 16 - 1 = 15$

For a 95% confidence level, $\alpha = 0 \cdot 05$ so that $\frac{\alpha}{2} = \frac{0 \cdot 05}{2} = 0 \cdot 025$

For a 99% confidence level, $\alpha = 0 \cdot 01$ so that $\frac{\alpha}{2} = \frac{0 \cdot 01}{2} = 0 \cdot 005$

The table value of $t_{.025}$ for 15 d.f. $= 2 \cdot 131$

The table vale of $t_{.005}$ for 15 d.f. $= 2 \cdot 947$

(a) 95% confidence limits for population mean μ are :

$$\overline{X} \pm t_{0.025} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Putting the values, we get

$$= 53 \pm 2 \cdot 131 \times \frac{3 \cdot 162}{\sqrt{16}}$$

$$= 53 \pm 2 \cdot 131 \times \frac{3 \cdot 162}{4}$$

$$= 53 \pm 1 \cdot 684$$

$$= 51 \cdot 316 \text{ to } 54 \cdot 684$$

Thus, $51 \cdot 316 < \mu < 54 \cdot 684$

(b) 99% confidence limits for population mean μ are

$$\bar{X} \pm t_{0 \cdot 005} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Putting the values, we get

$$= 53 \pm 2 \cdot 947 \times \frac{3 \cdot 162}{\sqrt{16}}$$

$$= 53 \pm 2 \cdot 947 \times \frac{3 \cdot 162}{4}$$

$$= 53 \pm 2 \cdot 947 \times 0 \cdot 7905$$

$$= 53 \pm 2 \cdot 33$$

$$= 50 \cdot 67 \text{ and } 55 \cdot 33$$

Thus, $50 \cdot 67 < \mu < 55 \cdot 33.$

Example 22. **A sample of 5 individuals had the following heights in centimeters: 6·33, 6·37, 6·36, 6·32 and 6·37. Find out the unbiased and efficient estimates of (a) true mean and (b) the variance. Also find 95% confidence interval for true mean (*i.e.,* population mean).**

Solution. (a) Unbiased and efficient estimate of the true mean (*i.e.,* population mean) is given by :

$$\bar{X} = \frac{\Sigma X}{n} = \frac{6 \cdot 33 + 6 \cdot 37 + 6 \cdot 36 + 6 \cdot 32 + 6 \cdot 37}{5} = 6 \cdot 35 \text{ cm}$$

(b) Unbiased and efficient estimate of the true variance (*i.e.,* population variance is given by :

$$\hat{s}^2 = \frac{n}{n-1} \cdot s^2 = \frac{\Sigma (X - \bar{X})^2}{n-1}$$

$$= \frac{(6 \cdot 33 - 6 \cdot 35)^2 + (6 \cdot 37 - 6 \cdot 35)^2 + (6 \cdot 36 - 6 \cdot 35)^2 + (6 \cdot 32 - 6 \cdot 35)^2 + (6 \cdot 37 - 6 \cdot 35)^2}{5-1}$$

$$= 0 \cdot 00055 \text{ cm}^2$$

$$\hat{s} = \sqrt{0 \cdot 00055} = 0 \cdot 023 \text{ cm.}$$

Part B : We are given : $n = 5, \bar{X} = 6 \cdot 35, \hat{s} = 0 \cdot 023$

Degrees of freedom $= v = n - 1 = 5 - 1 = 4$

For a 95% confidence level $\alpha = 0 \cdot 05$, so that $\frac{\alpha}{2} = \frac{0 \cdot 05}{2} = 0 \cdot 025.$

The table value of $0 \cdot 025$ for 4 d.f. $= 2 \cdot 776.$

95% confidence limits for population mean μ are given by

$$\bar{X} \pm t_{0 \cdot 05} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Putting the values, we get

$$= 6.35 \pm 2.776 \times \frac{0.023}{\sqrt{5}}$$

$$= 6.35 \pm 2.776 \times \frac{0.023}{2.236}$$

$$= 6.35 \pm 0.0285$$

$$= 6.3785 \text{ to } 6.3215$$

Thus, $6.3215 < \mu < 6.3785$

EXERCISE – 6

1. A sample of 9 cigarettes of a certain brand was observed for nicotine content. It showed an average nicotine of 25 milligrams and a standard deviation of 2·8 milligrams. Construct a 99 percent confidence interval for the true average nicotine content of this particular brand of cigarettes. **[Ans.** $21.67 < \mu < 28.33$**]**
2. A random sample of 15 ladies from a colony in Chandigarh shows that their monthly expenditure on cosemetics is Rs. 120 with a standard deviation of Rs. 40. Construct 95 percent confidence interval for the true monthly average expenditure on cosemetics by all the ladies in Chandigarh. **[Ans.** $97.01 < \mu < 142.99$**]**
3. A random sample of size 9 has 49 as mean. The sum of squares of deviations taken from mean is 52. Obtain 95% and 99% confidence limits for the mean. **[Ans.** (a) $47.04 < \mu < 50.96$ (b) $46.14 < \mu < 51.96$**]**
4. A sample of 10 measurements of the diameter of a sphere gave a mean $\overline{X} = 4.38$ and standard deviation $s = 0.06$ inches. Find (a) 95% and (b) 99% confidence limits for the actual diameter. **[Ans.** (a) 4·425 to 4·335 (b) 4·444 to 4·316**]**
5. A random sample of 10 families had the following percentage expenses on food : 68, 60, 75, 70, 73, 69, 59, 60, 49 and 44. Obtain 95% and 99% confidence limits for the population mean. **[Ans.** (a) 70·72 to 55·38 (b) 73·23 and 52·17**]**
6. A survey of 17 agricultural labourers reveals an income of Rs. 40 per week with a standard deviation of Rs. 8. Find out the limits of mean weekly wages in the population with a confidence of 95%. (Given $t = 2.131$ for 16 df). **[Ans.** $35.74 < \mu < 44.26$**]**
7. A sample of 6 persons in an office revealed an average daily smoking of 10, 12, 8, 9, 16, 5 cigarettes. Determine unbiased and efficient estimates of (a) the true mean and (b) the true variance. Also find 95% confidence interval for the true mean. **[Ans.** (a) $\overline{X} = 10$ (b) $\hat{s}^2 = 14$ (c) $6.92 < \mu < 13.08$**]**

(2) Confidence Interval or Limits for Population Variance (When $n < 30$). The determination of confidence interval or limits for population variance σ^2 requires the use of χ^2 (Chi-square) distribution. Here χ^2-values are used in place t-values.

$(1-\alpha)$ 100% confidence interval for population variance σ^2 is given by :

$$\frac{(n-1)\,\hat{s}^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)\,\hat{s}^2}{\chi^2_{1-\alpha/2}}$$

In particular, **95% confidence interval for the population variance σ^2 is**

$$\frac{(n-1)\,\hat{s}^2}{\chi^2_{0.025}} < \sigma^2 < \frac{(n-1)\,\hat{s}^2}{\chi^2_{0.975}}$$

Similarly, 99% confidence interval for the population variance σ^2 is

$$\frac{(n-1)\,\hat{s}^2}{\chi^2_{0.005}} < \sigma^2 < \frac{(n-1)\,\hat{s}^2}{\chi^2_{0.995}}$$

Procedure : The construction of the confidence interval for the variance σ^2 involves the following steps :

(i) Calculate modified sample variance $(\hat{s})^2$ by using the formula

$$\hat{s}^2 = \frac{n}{n-1}\,s^2 = \frac{\Sigma (X-\overline{X})^2}{n-1}$$

(ii) Select the desired confidence level and corresponding to that specified level of confidence, we note the value of the confidence coefficient $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ from the χ^2-table for certain degrees of freedom

(iii) Construct the confidence interval for σ^2 by putting the values of $\hat{s}^2$, $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ in the above stated formula.

Example 22. **A random sample of size 15 selected from a normal population has a standard deviation $s = 2\cdot5$. Construct a 95 percent confidence interval for variance σ^2 and standard deviation σ.**

Solution. We are given : $n = 15$, $s = 2\cdot5 \Rightarrow s^2 = 6\cdot25$

$$\hat{s}^2 = \left(\frac{n}{n-1}\right)s^2$$

$$= \frac{15}{15-1} \times 6\cdot25 = 6\cdot696$$

For a 95% confidence level, $\alpha = 0\cdot05$ so that $\frac{\alpha}{2} = 0\cdot025$ and $1-\alpha = 1-0\cdot025 = 0\cdot975$.

Degrees of freedom $(v) = n-1 = 15-1 = 14$

The table value of $\chi^2_{.025}$ for 14 d.f. $= 26\cdot1$

The table value of $\chi^2_{0.0975}$ for 14 d.f. $= 5\cdot63$

(a) 95% confidence interval for σ^2 is

$$\frac{(n-1)\,\hat{s}^2}{\chi^2_{.025}} < \sigma^2 < \frac{(n-1)\,\hat{s}^2}{\chi^2_{.975}}$$

Putting the values, we get

$$\frac{(15-1)\times 6\cdot696}{26\cdot1} < \sigma^2 < \frac{(15-1)\times 6\cdot696}{5\cdot63}$$

or $$3\cdot59 < \sigma^2 < 16\cdot65$$

(b) 95% confidence interval for σ is :

$$\sqrt{3\cdot59} < \sigma < \sqrt{16\cdot65}$$

or $$1\cdot89 < \sigma < 4\cdot08$$

Example 23. **A sample of 5 individuals had the following heights in inches 63·3, 63·7, 63·6, 63·2 and 3.8. Construct 95% confidence interval for population variance.**

Solution.

$$\overline{X} = \frac{63 \cdot 3 + 63 \cdot 7 + 63 \cdot 6 + 63 \cdot 2 + 63 \cdot 8}{5} = 63 \cdot 52$$

$$\hat{s}^2 = \frac{\Sigma (X - \overline{X})^2}{n-1}$$

$$= \frac{(63 \cdot 3 - 63 \cdot 52)^2 + (63 \cdot 7 + 63 \cdot 52)^2 + (63 \cdot 6 - 63 \cdot 52)^2 + (63 \cdot 2 - 63 \cdot 52)^2 + (63 \cdot 8 - 63 \cdot 52)^2}{5-1}$$

$$= \frac{\cdot 0484 + 0 \cdot 0324 + 0 \cdot 0064 + 0 \cdot 1024 + 0 \cdot 0784}{4}$$

$$= \frac{0 \cdot 268}{4} = 0 \cdot 067$$

95% confidence interval for σ^2 is

$$\frac{(n-1)\,\hat{s}^2}{\chi^2_{0 \cdot 025}} < \sigma^2 < \frac{(n-1)\,\hat{s}^2}{\chi^2_{0 \cdot 975}}$$

Degrees of freedom $(v) = n - 1 = 5 - 1 = 4$

The table value of $\chi^2_{0 \cdot 025}$ for 4 d.f. $= 11 \cdot 14$

The table value of $\chi^2_{0 \cdot 0975}$ for 4 d.f. $= \cdot 484$

95% confidence interval for σ^2 is

$$\frac{(5-1) \times (\cdot 067)}{11 \cdot 14} < \sigma^2 < \frac{(5-1)(\cdot 067)}{\cdot 484}$$

$$\Rightarrow \quad \frac{0 \cdot 268}{11 \cdot 14} < \sigma^2 < \frac{0 \cdot 268}{\cdot 484}$$

$$\Rightarrow \quad 0 \cdot 0240 < \sigma^2 < 0 \cdot 5537$$

EXERCISE – 7

1. A random sample of size 12 selected from a normal population has a standard deviation $s = 2 \cdot 4$. Construct 95 percent confidence interval for (a) variance σ^2 and (b) standard deviation σ. **[Ans.** (a) $3 \cdot 15 < \sigma^2 < 18 \cdot 08$ (b) $1 \cdot 77 < \sigma < 4 \cdot 25$]
2. A random sample of size 25 selected from normal population has a standard deviation $s = 7$. Construct 95% confidence interval for (a) variance σ^2, (b) standard deviation σ. **[Ans.** (a) $31 \cdot 12 < \sigma^2 < 98 \cdot 79$ (b) $5 \cdot 578 < \sigma < 9.939$]
3. A random sample of 15 ladies of a posh locality shows that their monthly expenditure on cosmetics is Rs. 120 with a standard deviation of Rs. 40. Construct 99 percent confidence interval for (a) variance σ^2 and (b) standard deviation σ. **[Ans.** (a) $766 \cdot 8 < \sigma^2 < 5896 \cdot 8$ (b) $27 \cdot 7 < \sigma^2 < 76 \cdot 7$]

QUESTIONS

1. What is statistical estimation ? Distinguish between point estimation and interval estimation. Describe the desirable properties of a good estimator.
2. What is an estimator ? Discuss the important properties of a good estimator. Show that the sample mean is a good estimate of population mean.
3. Differentiate between :
   (a) Estimator and Estimate
   (b) Statistic and Parameter
   (c) Point Estimator and Interval Estimate.
4. Explain the (i) consistency (ii) Unbiasedness (iii) Efficiency and (iv) Sufficiency properties of an estimator.
5. Define unbiased and efficient estimates of (a) true mean and (b) true variance.

   OR

   Define efficiency and unbiasedness of an estimator.
6. Define unbiased and consistency properties of an estimator.
7. Explain the concept of confidence interval or interval estimation. Outline the procedure for setting up a confidence interval for the population parameter.
8. Explain the procedure for setting up a confidence interval for (a) population mean (b) population proportions and (c) population variance.
9. Define the following terms and given an example of each :
   (i) Unbiased statistic (ii) Consistent statistic
   (iii) Efficient statistic (iv) Sufficient statistic
10. Show that the sample mean ($\overline{X}$) is an unbiased estimate of the population mean (μ).
11. Explain why a random sample of size 25 is to be preferred to a random sample of 20 to estimate population mean.