14

Estimation of Parameters

INTRODUCTION

One important problem of statistical inference is the estimation of population parameters (such as population mean, variance, etc.) from the corresponding sample statistics (*i.e.*, sample mean, variance, etc.). There are several occasions, on which, we have to estimate population values from sample data in order to make a business decision. For example, a firm may wish to estimate the average amount of time its salesmen spend on each sales call; the telephone department may be interested in estimating the average length of a conversation for a long distance telephone call; a company may be interested in estimating the share of the population aware of its products. If all these estimates are obtained on a census basis, it would be very costly and a time-consuming proposition. Hence, quite often, sampling theory is employed to obtain information about samples drawn at random from a known population and an attempt is made to infer information about a population by use of samples drawn from it.

Statistical Estimation is the procedure of using a sample statistic to estimate a population parameter. A statistic used to estimate a parameter is called an *estimator* and the value taken by the estimator is called an *estimate*. Statistical estimation is divided into two main categories: Point Estimation and Interval Estimation.

Point Estimation. An estimate of a population parameter given by a single number is called a *point estimate* of the parameter. For example, if a firm takes a sample of 50 salesmen and finds out that the average amount of time each salesman spends with his customers is 80 minutes. If this figure is used for an estimate for all the salesmen employed by the firm, it is referred to as a "Point estimate" because we are using one value to obtain the population value. If one must rely on a single value as an estimate of a parameter, it is desirable to select the random variable that is expected to provide the most dependable estimate. Now, suppose there are several alternative estimators which might be used for estimating the same parameter. For example, the population parameter may be a measure of central tendency of the population values; then sample mean, sample median and sample mode may be considered as the possible estimators of the population parameter. Which sample statistics should be used as the estimator of the population parameter ? We need to establish "criteria" for choosing a satisfactory estimator which tell us which statistic does the "best" job of estimating it. The best estimator is the one that is more suitable to a given problem, most likely to give the desired result, is the least risky, and has such desirable properties as of unbiasedness, consistency, efficiency and sufficiency which are discussed in detail in this chapter.

Properties of a Good Estimator

A good estimator, is one which is "close" to the population parameter being estimated. Some of the desirable properties of an estimator are :

(1) Unbiasedness,

(2) Consistency,

(3) Efficiency, and
(4) Sufficiency.

Unbiasedness. An estimator is a random variable as it is always a function of sample values. Then, if the average of these sample values is equal to the population parameter, then, it is unbiased estimate. Thus, an estimator is said to be unbiased *if the expected value of the estimator is equal to the population parameter being estimated.* If θ is the parameter being estimated and (read "theta hat") is an unbiased estimator of θ, then

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

For example, the sample mean is an unbiased estimator of the population mean, since

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

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

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

Similarly, consider a problem of estimating *p*, the population proportion of successes in a binomial distribution. It can be shown that, if a sample yields *x* successes is *n* trials, then the ratio *x/n* is an unbiased estimate of *p*. Since

$$E\left[\frac{x}{n}\right] = \frac{1}{n} E[x] = \frac{1}{n} np = p$$

Therefore, the sample proportion is an unbiased estimator of population proportion.

If the sampling distribution of $\hat{\theta}$ is such that

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

Then the estimator is said to be *biased.* The bias is the quantity by which $E(\hat{\theta})$ and θ differ. The sample variance s^2 computed with the division constant $1/n$ is a biased estimator of σ^2, because

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

Here, the bias is the quantity $-\frac{\sigma^2}{n}$.

Hence, sample variance is not an unbiased estimator of population variance. But, if sample variance is computed with a division constant $1/(n-1)$, then it can be shown that $E(s)^2 = \sigma^2$, and therefore, s^2 is an unbiased estimator of σ^2, howsoever small may be the sample size. It may be pointed out that although s^2 is an unbiased estimator of σ^2, *s* is not an unbiased estimator of σ. The bias, however, diminishes rapidly as *n* increases.

Consistency. As the sample size increases, the difference between the sample statistic and the population parameter should become smaller and smaller. If the difference continues to become smaller and smaller as the sample size becomes larger, the sample statistic is said to converge in probability to a parameter is said to be *consistent* estimator of that parameter. Symbolically, if $\hat{\theta}$ is a sample statistic computed from a sample of size *n* and θ is the parameter being estimated. If

$$Pr\left[\left|\hat{\theta} - \theta\right| \leq d\right] \rightarrow 1 \qquad \text{as } n \rightarrow \infty$$

for any positive arbitrary *d*, then $\hat{\theta}$ is said to be a consistent estimator of θ.

This is true for $\bar{x}$ and s^2 which are consistent estimators of μ and σ^2 respectively. The sample median is a consistent estimator of the population mean only if the population distribution is symmetrical.

Efficiency. If the variance of the estimator is small, the distribution of the estimator will be better in that its value will be closer to the parameter value. This is the notion of efficiency. Efficiency can be treated as a relative term. In a sense all estimators are efficient; however, some estimators are more efficient than others. The efficiency of an estimator depends on its variance. A measure of relative efficiency can be computed by taking the ratio of the variances of two estimators of interest. If $\hat{\theta}_1$ is an unbiased estimator of $\hat{\theta}$, and $\hat{\theta}_2$ represents another unbiased estimator of θ, then the relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is given by

$$\text{Relative efficiency} = \frac{\text{Var}\,[\hat{\theta}_2]}{\text{Var}\,[\hat{\theta}_1]}$$

For a symmetrical distribution, both the sample mean and sample median are unbiased and consistent estimators of the population mean. We choose between them on the basis of relative efficiency, we select that one which has the smaller variance. It can be proved that sample mean $\bar{x}$ is preferred to the sample median as an estimator of μ because $\bar{x}$ is more efficient estimator.

Sufficiency. The fourth and last property of a good estimator that was developed by a famous statistician, Sir R.A. Fisher, is sufficiency.

A sufficient estimator is one that uses all information about the population parameter contained in the sample. For example, the sample mean, $\bar{x}$, is a sufficient estimator of the population mean since all the information in the sample is used in its computation. On the other hand, sample mid-range is not a sufficient estimator since it is computed by averaging only the highest and lowest values in the sample.

A sufficient estimator ensures that all information that a sample can furnish with respect to the estimation of a parameter is being utilized. It may be noted that $\bar{x}$, p, $(\bar{x}_1 - \bar{x}_2)$ and $(p_1 - p_2)$ are sufficient estimators of the corresponding parameters μ, π, $(\mu_1 - \mu_2)$ and $(\pi_1 - \pi_2)$ respectively. A primary importance of the property of sufficiency is that it is a necessary condition for efficiency.

It is desirable that an estimator has all the properties discussed above. However, in practice, it is not always possible to determine such an estimator. It may be noted that though the point estimates are easy to use and understand, they cannot tell us how close we are to the true population value when used by themselves. Even the best point estimate may deviate enough from the parameter value to make the estimate unsatisfactory.

We have discussed above the properties desirable of an estimator to possess. We shall now discuss one of the most important methods called method of maximum likelihood that may provide estimators satisfying these properties.

Method of Maximum Likelihood

A general method for determining good estimators is called the *Method of Maximum Likelihood.* If a parameter θ is viewed as a variable, the method of maximum likelihood leads to the selection of a value of θ such that the likelihood (probability) of randomly obtaining a set of sample values is a maximum.

If x is a discrete random variable having a probability function $f(x\,;\theta)$ with only one parameter, θ, the likelihood function of the random sample $x_1, x_2, \ldots\ldots, x_n$ is

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

The general rule followed in finding that value which maximizes the likelihood function is to differentiate the likelihood function with respect to the parameter θ, set the resultant function equal to zero, and solve. Hence, the value of the parameter θ that maximizes the likelihood function is the maximum likelihood estimate.

The logic of this method can more easily be grasped by an example. Random sampling is based on the assumption that a sample, as a small scale replica of the population, will tend to reflect the properties of the population. For example, if a random sample of 35 students is selected from a class of 100 and average weight is calculated, we might get average weight as 120 lbs. One might ask what is the most likely value of μ in view of the sample result ? The conclusion is that the most likely value of μ is $\bar{x}$ = 120 lbs. and not 140 or 110 lbs. or some other value. The value of μ is indeed a definite, fixed value; but the method of maximum likelihood views μ, as if, it were a variable such that the most likely value of μ is, its *maximum likelihood estimator.*

The method of maximum likelihood provides estimators which are consistent, efficient and sufficient but it does not always provide estimators that are unbiased.

Illustration 1. A market organisation wants to introduce lottery systems to promote sales. The manager surveyed a sample of ten consumers for introduction of this system. Out of ten, six are in favour of this system. What is reasonable estimate of the population proportion π ?

Solution : Sample proportion $p = \frac{x}{n} = \frac{6}{10} = 0.6$. We would like to know, given, $p = 0.6$, how likely is it that true proportion $\pi = 0.1, 0.2, \ldots\ldots 0.9$.

By the binomial distribution, the probability of x successes is given by π is

$$B(x\,;\pi) = {}^{n}C_{x}\,\pi^{x}\,(1-\pi)^{n-x}$$

Suppose, the true population proportion is π = 0.3, this probability of obtaining sample with 6 successes in 10 is

$$B(6\,;0.3) = {}^{10}C_{6}\,(0.3)^{6}\,(0.7)^{4} = 0.0368$$

That is, if π = 0.3, the chances are less than 4 in a hundred that this sample result would occur.

If π = 0.5, we obtain a likelihood of

$$B(6\,;0.5) = {}^{10}C_{6}\,(0.5)^{6}\,(0.5)^{4} = 0.2051$$

Similarly, we can consider all other possible values of the parameter to determine how likely it is that the parameter considered would yield the sample actually observed.

We find that for π = 6.0, it has maximum probability 0.2508 which is the maximum likelihood estimate.

Interval Estimation. An estimate of a population parameter given by two numbers between which the parameter may be considered to lie is called as *interval estimate* of the parameter. Interval estimates indicate the precision or accuracy of an estimate and are, therefore, preferable to point estimates. The interval estimate or a "confidence interval" consists of an upper confidence limit and lower confidence limit, and we assign a probability that this interval contains the true population value. The first step in constructing a confidence interval is to decide how much confidence we want that this interval will contain the population value. Let us say, that we want 95 per cent confidence. This is known as a "95 per cent confidence level."

The previous chapter showed that when sampling is from a normally distributed population, the sampling distribution of sample mean is normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. Knowing that $\bar{x}$ is normally distributed allows us to make additional statements about the distribution of $\bar{x}$. Thus, the sampling distribution of $\bar{x}$ can be transformed into the standard normal distribution by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Suppose we take the point $\mu \pm 1.96\ \sigma/\sqrt{n}$

Let

$$\bar{x} = \mu + 1.96\ \sigma/\sqrt{n}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1.96\sigma/\sqrt{n}}{\sigma/\sqrt{n}} = 1.96$$

So, we can actually view z as the number of standard deviations a point is from μ . Since a z value of 1.96 corresponds to the point $\mu + 1.96\ \sigma/\sqrt{n}$, and z of -1.96 corresponds to $\mu - 1.96\ \sigma/\sqrt{n}$, clearly the area of the sampling distribution of $\bar{x}$ in the interval

$$\mu - 1.96\ \sigma/\sqrt{n} \text{ to } \mu + 1.96\ \sigma/\sqrt{n}$$

is 0.95. Therefore, 95 per cent of all sample means $\bar{x}$ are contained in this interval. Thus, the probability of drawing a random sample of size n and obtaining an $\bar{x}$ in this interval is 0.95.

$$P[\mu - 1.96\ \sigma/\sqrt{n} \le \bar{x} \le \mu + 1.96\ \sigma/\sqrt{n}\,] = 0.95$$

Similarly, it can be shown that

$$P[\mu - 2.578\ \sigma/\sqrt{n} \le \bar{x} \le \mu + 2.578\ \sigma/\sqrt{n}\,] = 0.99$$

The numbers 1.96, 2.578, etc., in the confidence limits are called *confidence coefficients* or *critical values*. It may be noted that the true mean may be expected to be no farther away than $3\,\sigma_{\bar{x}}$ from the sample mean that is a range of $\bar{x} \pm 3\,\sigma_{\bar{x}}$ will include the unknown true mean. Thus, the procedure in interval estimate comprises 3 steps :

1. The particular statistic, say, the mean of the sample or standard deviation of the sample is determined.
2. The confidence level is decided, *i.e.*, 95%, 99%, etc.
3. The standard error of the particular statistic is calculated.

Finally, we state, with a known degree of confidence, that the parameter is included in this interval.

Confidence Limits for Population Mean

Confidence limits for estimation of population mean μ are given by sample statistic $\pm\ z_c$ (S.E.) where z_c is critical value of z. 95% confidence limit for estimation of the population mean μ are given by $\bar{x} \pm 1.96\sigma_{\bar{x}}$; where lower confidence limit $= L = \bar{x} - 1.96\,\sigma_{\bar{x}}$ and upper confidence limit $= U = \bar{x} + 1.96\,\sigma_{\bar{x}}$.

Similarly, 99% confidence limits will be given by $\bar{x} \pm 2.58\,\sigma_{\bar{x}}$. For all practical purposes, a range of plus and minus three standard errors attached to sample mean, that is $\bar{x} \pm 3\,\sigma_{\bar{x}}$, will include the unknown true mean. More generally, the confidence limits are given by $\bar{x} \pm z_c\sigma_{\bar{x}}$ where z_c depends on the particular level of confidence desired. However, this is true in case sampling is from an infinite population or, if sampling is with replacement from a population of finite size N then confidence limits are given by :

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Illustration 2. The quality department of a wire manufacturing company periodically selects a sample of wire specimens in order to test for breaking strength. Past experience has shown that the breaking strengths of a certain type of wire are normally distributed with standard deviation of 200 kg. A random sample of 64 specimens gave a mean of 6,200 kg. The quality control supervisor wanted a 95 per cent confidence interval for the mean breaking strength of the population.

Solution. The z_c value corresponding to a confidence coefficient of 0.95 is 1.96. Therefore, the limits are :

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 6{,}200 \pm 1.96 \frac{200}{\sqrt{64}}$$

$$= 6{,}200 \pm 1.96 \times 25 = 6151 \text{ to } 6249$$

Hence, the 95 per cent confidence limits are 6151 to 6249.

Illustration 3. A manager wants an estimate of average sales of salesman in his company. A random sample of 100 out of 500 salesmen is selected and average sales is found to be Rs. 750 (thousand). If population standard deviation is Rs. 150 (thousand), manager specifies a 98% level of confidence. What is the interval estimate for average sales of salesman ?

Solution. Here $N = 500$, $n = 100$, $\bar{x} = 750$ and $\sigma = 150$.

The confidence limits are given by

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

where $$z_c = 2.33, \text{ and } \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{150}{100} \sqrt{\frac{400}{499}} = 15\ (0.895) = 13.425$$

The required confidence interval is

$$750 \pm 2.33(13.425)$$

Thus, it can be stated that for 0.98 level of confidence, the population mean falls within the interval Rs. 718720 to Rs. 781280.

Confidence Limits for Population Proportion

If sampling is from an infinite population, or if sampling is with replacement from a finite population, the confidence limits for the population proportion are given by :

$$p \pm z_c \sqrt{\frac{pq}{n}}$$

where p is the proportion of success in the sample of size n.

If we are interested in calculating the 95% confidence limits for the population proportion, they would be given by :

$$p \pm 1.96 \sqrt{\frac{pq}{n}}$$

Illustration 4. The Human Resource director of a large organisation wanted to know what proportion of all persons who had ever been interviewed for a job with his organisation had been hired. He was willing to settle for 95 per cent confidence interval. A random sample of 500 interview records revealed that 76 or 0.152 of the persons in the sample, had been hired.

Solution. The 95 per cent confidence interval for the population proportion is given by $p \pm 1.96 \sqrt{\frac{pq}{n}}$

$$= 0.152 \pm 1.96 \sqrt{\frac{0.152 \times 0.848}{500}} = 0.152 \pm 0.032 = 0.12 \text{ to } 0.184.$$

Hence, the required proportion varies between 0.121 and 0.183.

Illustration 5. Out of 20,000 customers' ledger accounts, a sample of 600 accounts was taken to test the accuracy of posting and balancing wherein 45 mistakes were found. Assign limits within which the number of defective cases can be expected at 5% level of confidence.

Solution. We are given that

$$n = 660,\ p = \text{proportion of mistakes} = \frac{45}{600} = 0.075$$

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

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.075 \times 0.925}{600}} = 0.011$$

Therefore, 95% confidence limits are given by

$$p \pm z_c \sqrt{\frac{pq}{n}}$$

$$= 0.075 \pm 1.96\ (0.011) = 0.075 \pm 0.022 = 0.053 \text{ to } 0.097$$

Hence, it is expected that the number of mistakes would vary between 5.3% and 9.7% at 5% level of significance. Here, the number of defective cases in a lot of 20,000 are expected to be between 20,000 × 0.053 and 20,000 × 0.097 or 1060 and 1940.

Confidence Limits for Difference of Two Means

When two independent random samples of $n_1 > 30$ and $n_2 > 30$ are taken then the sampling distribution of the difference of the two sample means $\bar{x}_1 - \bar{x}_2$ is approximately normal with mean $(\mu_1 - \mu_2)$ and $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. Given the confidence coefficient, the approximate confidence limits for $(\mu_1 - \mu_2)$ may be expressed as follows :

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_c\, \sigma_{\bar{x}_1 - \bar{x}_2}$$

Illustration 6. A sample of 150 items from machine *A* had an average life of 1,400 hrs. A similar sample of 100 items from machine *B* had a mean life of 1,200 hrs. Past records indicate that the standard deviation of the items produced by machine *A* is 120 hrs. and by machine *B* is 80 hours. Find 95 per cent confidence limits on the difference in the average lifetimes of the populations of the items produced by the two machines.

Solution. The 95 per cent confidence limits are given by

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_c\, \sigma_{\bar{x}_1 - \bar{x}_2}$$

where $\bar{x}_1 - \bar{x}_2 = 1400 - 1200 = 200$, and $z_c = 1.96$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{100}} = 12.6$$

Therefore, the required 95 per cent confidence limits are :

$$200 \pm 1.96\ (12.6) = 175.304 \text{ to } 224.696$$

Hence, the 95 per cent confidence limits are 175.304 to 224.696 for the difference in the average lifetime of the items produced by the two machines *A* and *B*.

Confidence Limits for Difference of Two Proportions

The confidence limits for the difference of two population proportions, where the populations are infinite, are given by :

$$(p_1 - p_2) \pm z_c \sigma_{p_1 - p_2}$$

where $(p_1 - p_2)$ = difference of proportions, and

$\sigma_{p_1 - p_2}$ = Standard error of the difference of two proportions given by :

$$\sigma_{p_1 - p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

We have discussed above the procedure of estimating a population measure for sample measure. The discussion has been confined to estimation for mean, proportion, difference of means, difference of proportions. These few parameters have been singled out for discussion because they are by far the most important decision parameters in so far as univariate data are concerned. However, it may be noted that the same general procedure of interval estimation can be applied to other parameters provided *n* is large enough for the central limit theorem to operate.

It may be noted that the procedure of estimation discussed in this chapter is applicable only in case of large samples (sample size greater than 30). Small samples require special treatment.

Determination of a Proper Sample Size

Thus far we have calculated the confidence intervals based on the assumption that the sample size n is known. In most of the practical situations, generally, it is not known. Instead, one may prefer to specify the width of the interval and use this information to solve for n. The method of determining a proper sample size, n is given for the following two cases :

(a) Sample size for estimating a population mean

The confidence interval formula is given by

$$\bar{x} \pm z_c \sigma / \sqrt{n}$$

or

$$\bar{x} \pm E$$

where

$$E = z_c \sigma / \sqrt{n}$$

is the maximum allowable sampling error, *i.e.*, difference between the population mean and the sample mean.

or

$$\sqrt{n} = \frac{z_c \sigma}{E}$$

or

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

Here, both the values of z_c and E must be specified by the researcher, the value of the population σ may be actual or estimated.

Illustration 7. A cigarette manufacturer wishes to use a random sample to estimate the average nicotine content. The sampling error should not be more than one milligram above or below the true mean, with a 99 per cent confidence coefficient. The population standard deviation is 4 milligrams. What sample size should the company use in order to satisfy these requirements?

Solution. Here $E = 1$, $z_c = 2.58$, and $\sigma = 4$

Sample size formula is

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

Substituting the values, we get

$$n = \frac{(2.58)^2 (4)^2}{1^2} = 106.50 \text{ or } 107$$

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

(b) Sample size for estimating a population proportion

The confidence interval formula for proportion is given by

$$p \pm z_c \sqrt{\frac{pq}{n}}$$

Using E to represent the maximum allowable sampling error, we may write the above equation as

$$p \pm E$$

where E is the difference between the sample proportion and the population proportion.

Now

$$E = z_c \sqrt{\frac{pq}{n}}$$

Solving for n, we get

$$n = \frac{z_c^2 pq}{E^2}$$

where the values of z_c and E are predetermined. The value of population proportion p may be actual or estimated from the past experience.

Illustration 8. A firm wishes to estimate with a maximum allowable error of 0.05 and a 95 per cent level of confidence, the proportion of consumers 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 per cent of all consumers prefer the firm's product ?

Solution. Here $E = 0.05$, $p = 0.25$, and $z_c = 2.33$.

Substituting these values in the formula

$$n = \frac{z_c^2}{E^2} pq, \text{ we get}$$

$$n = \frac{(2.33)^2}{(0.05)^2}(0.25)(0.75)$$

$$= \frac{5.4289}{0.0025}(0.1875) = \frac{1.0179}{0.0025} = 407.16 \text{ or } 407$$

Hence the required sample size $n = 407$.

MISCELLANEOUS ILLUSTRATIONS

Illustration 9. A machine is producing ball bearings with diameter of 0.5 inches. It is known that the standard deviation of the ball bearings is 0.005 inches. A sample of 100 ball bearings is selected and their average diameter is found to be 0.498 inches. Determine the 99 per cent confidence interval.

Solution. Using the formula for confidence interval

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

where $\bar{x} = 0.498$, $z_c = 2.58$, $\sigma = 0.005$ and $n = 100$

We get $0.498 \pm 2.58 \dfrac{0.005}{\sqrt{100}} = 0.498 \pm 2.58\ (0.0005)$

$$= 0.498 \pm 0.00129 = 0.4967 \text{ to } 0.4993$$

Hence, the 99 per cent confidence interval is 0.4967 to 0.4993 inches.

Illustration 10. Suppose we want to estimate the proportion of families in a town which has two or more children. A random sample of 144 families shows that 48 families have two or more children. Construct a 95 per cent confidence interval. *(MBA, HPU, 2007)*

Solution. Sample proportion $p = \dfrac{x}{n} = \dfrac{48}{144} = \dfrac{1}{3} = 0.333$

The confidence interval formula for proportion p is given by

$$p \pm z_c \sqrt{\frac{pq}{n}}$$

where $p = \dfrac{1}{3}$, $q = \dfrac{2}{3}$, $z_c = 1.96$, $n = 144$

and $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(1/3)(2/3)}{144}} = 0.0393$

Therefore, the required confidence interval is

$$0.333 \pm 1.96\ (0.0393) = 0.333 \pm 0.077 = 0.256 \text{ to } 0.410.$$

Hence, the population proportion of families who have two or more children is between 25.6 and 41.0 per cent.

Illustration 11. A ball pen manufacturer makes a lot of 10,000 refills. The procedure desires some control over these lots so that no lot will contain an excess number of defective refills. He decides to take a random sample of 400 refills for inspection from a lot of 10,000 and finds 9 defectives. Obtain a 90% confidence interval for the number of defectives in the entire lot.

Solution. The formula to be used for confidence interval is

$$p \pm z_c \sqrt{\frac{pq}{n}}$$

where $p = \dfrac{x}{n} = \dfrac{9}{400} = 0.0225$, $z_c = 1.645$

and $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{0.0225 \times 0.9775}{400}} = 0.0074$

Therefore, the required confidence interval is given by

$$0.0225 \pm 1.645(0.0074) = 0.0225 \pm 0.0122 = 0.0103 \text{ to } 0.0347$$

Hence, we may conclude with 90 per cent confidence that the population contains between 1.03 and 3.47 per cent defective in the entire lot.

Illustration 12. In a large consignment of oranges, a random sample of 500 oranges revealed that 65 oranges were bad. Prove that 99.73% of bad oranges in the consignment certainly lie between 8.5% and 17.5%.

Solution. Given that $n = 500$

$$p = \text{number of bad oranges in the consignment } \frac{65}{500} = 0.13,\ q = 1 - p = 1 - 0.13 = 0.87,\ z_c = 3$$

and
$$\sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13 \times 0.87}{500}} = 0.015$$

The 99.73% confidence limits for the population proportion of bad oranges in the consignment are given by

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3 \times 0.015 = 0.13 \pm 0.045 = 0.085 \text{ and } 0.175$$

Hence, the percentage of bad oranges in the consignment certainly lies between 8.5% and 17.5%.

Illustration 13. 400 labourers were selected at random from a certain city. Their mean income was Rs. 1700 per month with a standard deviation of Rs. 140. Set up 95% confidence limits within which the income of the labour community of the district is expected of lie.

Solution. Given $\bar{x} = 1700,\ \sigma = 140\ n = 400 \text{ and } z_c = 1.96$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{140}{20} = 7$$

Therefore, 95% confidence limits are given by

$$\bar{x} \pm z_c \sigma_{\bar{x}} = 1700 \pm 1.96\,(7) = 1700 \pm 13.72$$
$$= 1686.28 \text{ to } 1713.72.$$

Illustration 14. In an attempt to control the quality of output for a manufactured part, a sample of parts is chosen and examined in order to estimate the population proportion of parts that are defective. The manufacturing process operates continuously unless it must be stopped for inspection or adjustment. In the latest sample of 90 parts, 15 defectives are found. Determine the following estimates of π the population proportion defective (*a*) a point estimate (*b*) 98 per cent interval estimate.

Solution. (*a*) Point estimate : $p = \dfrac{15}{90} = 0.167$

(*b*) Interval estimate : $\sigma_p = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{(0.167)(0.833)}{90}} = 0.0393$

98 per cent interval estimate shall be given by

$$p \pm z_c \sigma_p = 0.167 \pm 2.33 \times 0.0393$$
$$= 0.167 \pm 0.092 \text{ or } 0.075 \text{ to } 0.259.$$

Illustration 15. A random sample of 200 consumer accounts at a large brokerage firm is selected for the purpose of estimating the mean number of transactions per year for each customer. The sample mean is 12. Determine 99% confidence interval for the mean number of transactions of all consumer accounts of the firm.

Solution. Using the formula

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

where, $\bar{x} = 43,\ z_c = 2.58,\ \sigma = 2.5,\ n = 200$

$$43 \pm 2.58\,\frac{2.5}{\sqrt{200}} = 43 \pm 0.456 = 42.544 \text{ to } 43.456.$$

PROBLEMS

1-A : Answer the following questions, each question carries **one** mark:

(*i*) What is statistical estimation ?

(*ii*) Distinguish between point estimate and interval estimate.

(*iii*) What is point estimation ? (*MA, Eco. M.K. Univ.*)

(*iv*) What are confidence limits for population mean ?

(*v*) What are confidence limits for population proportion ?

(*vi*) Name the important properties of a good estimator.

(*vii*) Give the formula for the method of maximum likelihood of a good estimator.

(viii) Which formula is used for determining confidence limits for difference of two means ?
(ix) Give the formula for determining sample size for estimating a population mean.
(x) What are confidence limits for difference of two proportions ?
(xi) Differentiate between confidence limits and confidence level.

1-B : Answer the following questions, each question carries **four** marks:
(i) Explain the concept of confidence interval with suitable example.
(ii) Briefly explain any two properties of a good estimator.
(iii) Describe the desirable properties of a good estimator.
(iv) What are confidence limits ? How are they determined ?
(v) How sample size is determined ? Explain with the help of an example.

2. What do you understand by estimation? In what sense, do we consider estimation as a procedure of decision-making ?

3. (a) What do you mean by 'Statistical Estimation'? Briefly explain the methodology used for estimating the mean of the population from the mean of the sample.
(b) Distinguish clearly between the point estimation and interval estimation. In what way, do we say that an interval estimate is better than a point estimate ?

4. (a) Explain clearly the desirable properties of a point estimate.
(b) What information and assumptions must be given to compute the sample size for an interval estimate of the universe mean ?

5. What is the difference between 'Statistic' and 'Parameter'? Explain, with examples, the methods employed for the estimation of population parameters based on sample means, difference of two means, sample proportion and difference of the sample proportions.

6. What is meant by confidence interval of a population parameter ?

7. With the help of an example, explain the method of maximum likelihood and point out its significance.

8. Comment on the statement, "Theoretically speaking, it is possible to have an estimate which is identical with the parameter being estimated. In practice, however, such an estimate is often unnecessary and physically impossible."

9. (a) Explain clearly the procedure involved in interval estimation.
(b) Describe briefly the problems of estimation of population parameters.

10. Explain the following terms with the help of an example :
(i) Confidence limits, (ii) Confidence interval,
(iii) Interval estimate, (iv) Confidence coefficients or critical values.

11. What are the properties of a good estimator ? Prove that the mean of a simple random sample from a given population is an unbiased estimator of the population mean.
(a) Explain briefly the properties of a good estimator.
(b) Explain the concepts of (i) the power of statistical test, (ii) reliability and validity of measurements.

12. In a consignment of 1,00,000 tennis balls, 400 were drawn at random and examined. It was found that 20 of these were defective. How many defective balls can you expect in the whole consignment at 95% confidence level ?

13. A statistics consultant with the association of personnel director was asked to determine what proportion of electrical personnel who change jobs do so because they are bored with their work. A random sample of 400 electrical personnel who had recently changed jobs were enquired, and 200 stated that they changed jobs because they were bored. The statistician prepared a 95 per cent confidence interval for the true proportion changing jobs because of boredom. What are the lower and upper limits of this interval ?

14. A bank official is interested in knowing the difference between the average amount of money or deposit by customers in two branch banks. A random sample of 35 customers was selected from each branch. The sample means were as follows : Branch A : Rs. 4500; Branch B : Rs. 3250. The two populations are normally distributed with variances $\sigma_A^2 = 760$ and $\sigma_B^2 = 850$. Construct the 95 and 99 per cent confidence interval for $\mu_A - \mu_B$.

15. A random sample of 50 persons was interviewed to find their preference between two brands of tea. 35 of the interviewed persons preferred brand A to brand B. Find the 95 per cent confidence interval for the proportion of persons who prefer brand A.

16. After an intensive advertisement campaign of polish, the manufacturers wanted to know how many of the possible customers had read the advertisement. They selected a random sample of 50 customers and found that only 15 of them had read the advertisement. Find 95 per cent confidence interval for the proportion of customers who had not read the advertisement.

17. A manufacturer of television picture tubes tested 75 tubes to determine their mean lifetime. The sample yield an average of 4,200 hours with a standard deviation of 430 hours. Use a 95 per cent level of confidence for the interval estimate of the value below which the mean of the population should not fall.

18. A new drug has been developed for the treatment of a certain disease. A group of 400 patients suffering from the disease were treated with the new drug. Another group of 400 patients were treated with an alternative drug. At the end of two weeks, 320 of the patients receiving the new drug recovered, while 240 of those taking the alternative drug recovered. Construct the 95 per cent confidence interval for the difference in the true proportion of patients who might be expected to respond to the two drugs.

19. A sample of 16 observations has been taken from a population in which the random variable is normally distributed. The sample is 50 and the sample standard deviation is 10. Determine a 95 per cent confidence interval for the population mean.

20. A statistician is asked to conduct a survey to determine an estimate of the proportion of the people who favour the recall of the local politician. He is told that his estimate should not differ from the true proportion by more than 2 per cent with 95 per cent confidence. How large should his random sample be to produce an estimate of the proportion satisfying this condition?

21. The wearing quality of a certain type of truck tyre is to be estimated by road testing a sample of the tyres. It is estimated that the standard deviation of wearing quality is 200 km.

(*a*) If the maximum allowable sampling error is 600 km, at a 95 per cent level of confidence, what should be the sample size ?

(*b*) If the level of confidence were 99 per cent, what would be the appropriate sample size?

(*c*) If the maximum allowable error were 300 km, what would be the appropriate sample size for a 95 per cent level of confidence ?

22. 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 exceed 0.01 seconds ?
[96]

23. A factory is producing 50,000 pairs of shoes daily. From a sample of 500 pairs, 20% were found to be of substandard quality. Estimate the number of pairs that can be reasonably expected to be spoiled in the daily production and assign limits at 5% level of significance.
[Between 385 and 1,615]

24. The guaranteed average life of a certain type of electric light bulbs is 1,000 hours with a standard deviation of 125 hours. It is proposed to sample the output so as to ensure that 90% of the bulbs do not fall short of the guaranteed average by more than 2.5 per cent. What should be the minimum size of the sample ?

25. A random sample of six castings drawn from a universe of 75 castings shows the following weight for each. Compute an interval estimate for μ at 2% level of confidence.

Casting No. :	1	2	3	4	5	6
Weight (kg) :	82.9	83.5	84.1	83.6	82.5	84.4

26. In a random sample of 81 items taken from a large consignment, some were found to be defective. If the standard error of the proportion of defective items in the sample is 1/16, find 95% confidence limits of the percentage of defective items in the consignment.

27. From previous studies, the population standard deviation for a placement test has been determined to be 12.4. The test is scored on a scale of 0 – 100. A placement agency wants to be 90% confident that the average test score of a sample falls within plus or minus 3 points of the population average score. How large a sample should be selected ?

28. The foreman of a mining company has estimated the average quantity of ore extracted per shift to be 34.6 tons and the sample standard deviation to be 2.8 tons per shift based upon a random sample of six shifts. Construct 95% and 90% confidence around sample average estimate.

29. The life (in hours) of a 100 watt bulb is known to be normally distributed with standard deviation of 36 hours. A random sample of 15 bulbs yielded the following results :

Life in hours

2216 2237 2249 2204 2225 2301 2281 2263 2318 2255 2275 2295 2250 2238 2300

Construct a 95% two sided confidence interval so that the actual mean of the life of bulbs fall within this interval.

30. A machine fills cans with a soft drink beverage and the manufacturer is interested in obtaining a confidence interval estimate of the variance of its fill volume. A random sample of 20 cans yields a sample variance of 0.0225. Construct a two-sided 95% confidence interval for variance.

31. A manufacturer is interested in estimating the proportion (p) of acceptable products. Find an upper limit of sample size that would ensure that this estimate does not deviate from the true value by more than 0.04 at 99% level of confidence.

32. In order to test the durability of a new paint, a highway department had test strips painted across heavily travelled roads in 15 different locations. If on the average, the test strips disappeared after they had been crossed by 1,46, 692 cars and with standard deviation 14,380 cars, construct 99% confidence interval for the true average number of cars it will take to wear off.

33. **A machine is supposed to drill holes with a diameter of 1 inch. In fact, the diameters are normally distributed with a mean of 1.01 inches and a standard deviation of 0.02 inch. If there is a tolerance of 0.02 inch, the holes should be between 0.91 and 1.02 inches. What percentage of the holes drilled are within clearance ?**

34. **A machine is producing ball bearings with diameter of 0.5 inches. It is known that the standard deviation of the ball bearings is 0.05 inches. A sample of 100 ball bearings is selected and their average diameter is found to be 0.498 inches. Determine the 99 per cent confidence interval.**

35. **With a sample size of 400, the calculated standard error of mean is 2 with a mean of 120. What sample size would be required so that we could be 95% confident that the population mean is within ± 3.5 of the sample mean ?**

36. **A random sample of 160 people is taken and 120 were in favour of liberalising licensing regulations. With 95% confidence, what proportion of all people are in favour ?**

37. **A sample of 64 men from a population with known standard deviation of height of 2.4 inches gives a mean height of 69.8 inches. Find a 90% confidence interval of μ, the mean height of the men in the population.**
[69.3065 to 70.2935]

38. **Given a population with a standard deviation of 8.6, What sample size is needed to estimate the mean of the population within ±0.5 of the sample mean with 99 per cent confidence ?**
(MBA, KU, 2003)

*****