UNIT 7 INTERVAL ESTIMATION FOR ONE POPULATION

Structure

7.1 INTRODUCTION

In the previous unit, we have discussed the point estimation, under which we learn how one can obtain point estimate(s) of the unknown parameter(s) of the population using sample observations. Everything is fine with point estimation but it has one major drawback that it does not specify how confident we can be that the estimated close to the true value of the parameter.

Hence, point estimate may have some possible error of the estimation and it does not give us an idea of how these estimates deviate from the true value of the parameter being estimated. This limitation of point estimation is over come by the technique of interval estimation. Therefore, instead of making the inference of estimating the true value of the parameter through a point estimate one should make the inference of estimating the true value of parameter by a pair of estimate values which are constituted an interval in which true value of parameter expected to lie with certain confidence. The technique of finding such interval is known as **"Interval Estimation"**.

For example, suppose that we want to estimate the average income of persons living in a colony. If 50 persons are selected at random from that colony and the annual average income is found to be Rs. 84240 then the statement that the average annual income of the persons in the colony is between Rs. 80000 and Rs. 90000 definitely more likely to be correct than the statement that the annual average income is Rs. 84240.

This unit is divided into eleven sections. Section 7.1 is introductory in nature. The confidence interval and confidence coefficient are defined in Section 7.2. The general method of finding the confidence interval is explored in Section

7.3. Confidence interval for population mean in different cases as population variance is known and unknown are described in Section 7.4 whereas in Section 7.5, the confidence interval for population proportion is explained. The confidence interval for population variance in different cases when population mean is known and unknown are described in Section 7.6. Section 7.7 is devoted to explain the confidence interval for non-normal populations. The concept of shortest confidence interval and determination of sample size are explored in Sections 7.8 and 7.9 respectively. Unit ends by providing summary of what we have discussed in this unit in Section 7.10 and solution of exercises in Section 7.11.

Objectives

After studying this unit, you should be able to:

- need of interval estimation;
- define the interval estimation;
- describe the method of obtaining the confidence interval;
- obtain the confidence interval for population mean of a normal population when population variance is known and unknown;
- obtain the confidence interval for population proportion;
- obtain the confidence interval for population variance of a normal population when population mean is known and unknown;
- obtain the confidence intervals for population parameters of a non-normal populations;
- explain the concept of the shortest interval; and
- determination of sample size.

7.2 INTERVAL ESTIMATION

In previous section, we introduced you with the interval estimation and after that we can conclude that if we find two values with the help of sample observations and constitute an interval such that it contain the true value of parameter with certain probability then it is known as interval estimate of the parameter. This technique of estimation is known as "Interval Estimation".

In this section, we will formally define:

- Confidence Interval and Confidence Coefficient
- One-sided Confidence Intervals

in the following two sub-sections.

7.2.1 Confidence Interval and Confidence Coefficient

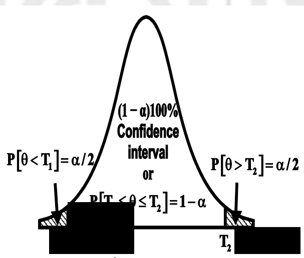

Fig. 7.1

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a population whose probability density (mass) function is f(x, θ). Let $T_1 = t_1(X_1, X_2, ..., X_n)$ and $T_2 = t_2(X_1, X_2, ..., X_n)$ (where $T_1 \le T_2$) be two statistics such that the probability that the random interval $[T_1, T_2]$ includes the true value of population parameter θ is $(1-\alpha)$, that is,

$$P\left[T_1 \le \theta \le T_2\right] = 1 - \alpha$$

as shown in the Fig.7.1, where, α does not depend on θ.

Then the random interval $[T_1, T_2]$ is known as $(1 - \alpha)$ 100% confidence interval for unknown population parameter θ and $(1 - \alpha)$ is known as confidence coefficient or confidence level. The above probability statement may be explained with the help of an example:

Suppose we say that the probability that the random interval contains the true value of parameter θ is 0.95 then by this statement we simple mean that if 100 samples of same size, say, 'n' are drawn from the given population $f(x,\theta)$ and the random interval $[T_1, T_2]$ is computed for each sample, then 95 times out of 100 intervals, the random interval $[T_1, T_2]$ contains the true value of parameter θ. Hence, higher the probability $(1 - \alpha)$, more the confident we will have that the random interval $[T_1, T_2]$ will actually includes the true value of θ. The statistics T_1 and T_2 are known as lower and upper confidence limits or confidence bounds or fiducial limits, respectively for θ. And the interval is known as two-sided confidence interval. The length of confidence interval is defined as

L = Upper confidence limit − Lower confidence limit

i.e. $L = T_2 - T_1$

The confidence interval may also be one-sided so in the next sub-section we define one-sided confidence intervals.

7.2.2 One-Sided Confidence Intervals

In some situations, we may be interested in finding an upper bound or lower bound but not both for population parameter with a given confidence. For example, one may be interested to obtain a bound such that he /she is 95% confident that the average life of the electric bulbs of a company is no less than one years. In such cases, we construct one-sided confidence intervals.

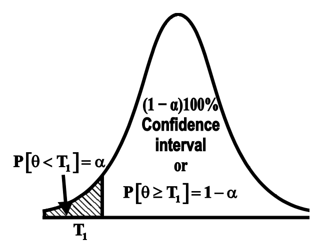

Fig. 7.2

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from the population having probability density(mass) function $f(x, \theta)$ and also let T_1 be a statistic such that

$$P[\theta \ge T_1] = 1 - \alpha$$

as shown in Fig. 7.2, then statistic T_1 is called a lower confidence bound for parameter θ with confidence coefficient $(1 - \alpha)$ and $[T_1, \infty)$ is called a lower one sided $(1 - \alpha)$ 100% confidence interval for parameter θ.

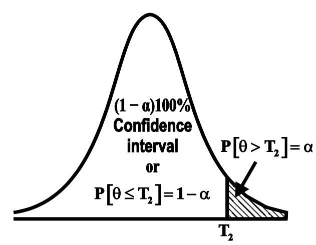

Fig. 7.3

Similarly, let T_2 be a statistic such that

$$P[\theta \le T_2] = 1 - \alpha$$

as shown in Fig. 7.3, then statistic T_2 is called an upper confidence bound for parameter θ with confidence coefficient $(1 - \alpha)$ and $(-\infty, T_2]$ is called a upper one sided $(1 - \alpha)$ 100% confidence interval for parameter θ.

Note 1: Generally, one-sided confidence intervals are rarely used so we focus on two-sided confidence interval in this course. **Generally, confidence interval means two-sided confidence interval unless or otherwise it is stated as one-sided.**

Now, you can try the following exercises.

E1) Find the length of the following confidence intervals:

(i) $P[-1.65 \le \mu \le 3.0] = 0.95$ (ii) $P[-1.68 \le \mu \le 2.70] = 0.95$

(iii) $P[-1.70 \le \mu \le 2.54] = 0.95$ (iv) $P[-1.96 \le \mu \le 1.96] = 0.95$

E2) Find the lower and upper confidence limits and also confidence coefficient of the following confidence intervals:

(i) $P\left[0 \le \theta \le 1.5\right] = 0.90$ (ii) $P\left[-1 \le \theta \le 2\right] = 0.95$

(iii) $P\left[-2 \le \theta \le 2\right] = 0.98$ (iv) $P\left[-2.5 \le \mu \le 2.5\right] = 0.99$

7.3 METHOD OF OBTAINING CONFIDENCE INTERVAL

After knowing about the confidence interval, the question may arise in your mind that "how confidence intervals are obtained?" There are following two methods are generally used for obtaining confidence intervals:

1. Pivotal quantity method
2. Statistical method

The statistical method is beyond of scope of this course so we will keep our focus only on pivotal-quantity method which is also known as general method of interval estimation. Before going to describe this method, we first define the pivotal quantity as:

Pivotal Quantity

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a population having probability density (mass) function f (x, θ). If quantity $Q = q(X_1, X_2, ..., X_n, \theta)$ is a function of $X_1, X_2, ..., X_n$ and parameter θ such that its distribution does not dependent on unknown parameter θ then the quantity Q is known as a pivotal quantity.

For example, if $X_1, X_2, ..., X_n$ is a random sample taken from normal population with mean μ and variance 4, i.e. N(μ, 4) where, the parameter μ is unknown then we know that the sampling distribution of sample mean is also normal with mean μ and variance 4/n, that is, $\bar{X} \sim N\left(\mu, 4/n\right)$ and the sampling distribution of variate $Z = \dfrac{\bar{X} - \mu}{\sqrt{4/n}}$ is $N\left(0, 1\right)$. Since distribution of $\bar{X}$ depends on the parameter μ to be estimated, therefore, it is not a pivotal quantity whereas the distribution of variate Z is independent of parameter μ so it is a pivotal quantity.

Pivotal Quantity Method

The pivotal quantity method for confidence interval has following steps:

Step 1: First of all, we search the statistic for unknown parameter which can be used to estimate the parameter, say, θ, preferably a sufficient statistic, whose distribution is completely known. After that we find the function based on that statistic whose distribution does not dependent on parameter θ which is to be estimated i.e. we find the pivotal quantity Q.

Step 2: Introduce two constants, say, 'a' and 'b', depending on α but not on unknown parameter θ, such that

$$P\left[a \le Q \le b\right] = 1 - \alpha$$

Step 3: Since pivotal quantity is a function of parameter, therefore, we convert above interval for parameter θ as

$$P\left[T_1 \le \theta \le T_2\right] = 1-\alpha$$

where, T_1 and T_2 are functions of sample values and a & b.

Step 4: Determine constants 'a' and 'b' by minimizing the length of the interval

$$L = T_2 - T_1$$

With the help of the pivotal quantity method, we will find the confidence intervals for population mean, proportion, variance which will describe one by one in subsequent sections.

Now, you can try the following exercise.

E3) Describe the general method of constructing confidence interval for population parameter.

7.4 CONFIDENCE INTERVAL FOR POPULATION MEAN

There are so many problems in real life where it becomes necessary to obtain the confidence interval of population mean. For example, an investigator may interested to find the interval estimate of average income of the people living in a particular geographical area, a product manager may want to find the interval estimate of average life of electric bulbs manufactured by a company, a pathologist may want to obtain the interval estimate of the mean time required to complete a certain analysis, etc.

For describing confidence interval for population mean, let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population having mean μ and variance σ^2. We can determine confidence interval for population mean μ under following two cases:

1. When population variance σ^2 is known
2. When population variance σ^2 is unknown.

These two cases are discussed one by one in subsequent Sub-sections 7.4.1 and 7.4.2 respectively.

7.4.1 Confidence Interval for Population Mean when Population Variance is Known

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population $N(\mu, \sigma^2)$ when σ^2 is known, that is, σ^2 has a specify value, say, σ_0^2.

To find out the confidence interval for population mean, first of all we search the statistic for estimating μ whose distribution is completely known. Generally, we use the value of statistic (sample mean) $\bar{X}$ to estimate the population mean μ and also it is a sufficient statistic for parameter θ. Therefore, we use $\bar{X}$ to make the pivotal quantity.

We know that when parent population is normal $N(\mu, \sigma^2)$ then sampling distribution of sample mean $\bar{X}$ is normally distributed with mean μ and variance σ^2/n, that is, if

$$X_i \sim N\left(\mu,\, \sigma^2\right)$$

then

$$\bar{X} \sim N\left(\mu, \sigma^2 / n\right)$$

and the variate

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,\, 1)$$

follows the normal distribution with mean 0 and variance unity. Therefore, the probability density function of Z is

$$f\left(z\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2};\ -\infty < z < \infty$$

Since distribution of Z is independent of the parameter to be estimated i.e. μ, therefore, Z can be taken as pivotal quantity. So we introduce two constants, say, $z_{\alpha/2}$ and $z_{1-\alpha/2} = -z_{\alpha/2}$(since distribution of Z is symmetrical about Z = 0 line see Fig. 7.4) such that

$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha$$

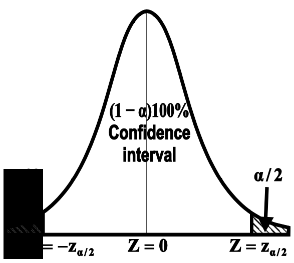

Fig. 7.4

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z as shown in Fig. 7.4.

By putting the value of Z in above equation / probability statement, we get

$$P\left[-z_{\alpha/2} \le \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \le z_{\alpha/2}\right] = 1 - \alpha$$

Now, for converting this interval for parameter μ, we multiply each term in above inequality by $\sigma / \sqrt{n}$, we get

$$P\left[-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \bar{X} - \mu \le +z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Now, subtracting $\bar{X}$ from each term in above inequality then we get

$$P\left[-\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Multiplying each term by (-1) in above inequality, we get

$$P\left[\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \ge \mu \ge \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \quad \left[\begin{array}{l}\because \text{by multiplying} \left(-1\right) \\ \text{the inequality is reversed}\end{array}\right]$$

This can be rewritten as

$$P\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Hence, (1 - α) 100% confidence interval of population mean is given by

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right] \qquad \ldots (1)$$

and corresponding limits are given by

$$\bar{X} \mp z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad \ldots (2)$$

Note 2: The value of $z_{\alpha/2}$ can be obtained by the method described in Unit 14 of MST-003. In interval estimation generally, we have to find the 90%, 95%, 98% and 99% confidence intervals therefore, the corresponding values of $z_{\alpha/2}$ are summaries in the Table 7.1 given below and we will use these values directly when needed:

Table 7.1: Communally Used Values of Standard Normal Variate Z

1 – α	0.90	0.95	0.98	0.99
α	0.10	0.05	0.02	0.01
α/2	0.05	0.025	0.01	0.005
$z_{\alpha/2}$	1.645	1.96	2.33	2.58

For example, if we want to find the 99% confidence interval (two-sided) for μ then

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01$$

For α = 0.01, the value of $z_{\alpha/2} = z_{0.005}$ is 2.58 therefore, 99% confidence interval for μ is given by

$$\left[\bar{X} - 2.58\frac{\sigma}{\sqrt{n}},\ \bar{X} + 2.58\frac{\sigma}{\sqrt{n}}\right]$$

Application of the above discussion can be seen in the following example.

Example 1: The mean life of the tyres manufactured by a company follows normal distribution with standard deviation 3200 kms. A sample of 250 tyres is taken and it is found that the average life of the tyres is 50000 kms with a standard deviation of 3500 kms. Establish the 99% confidence interval within which the mean life of tyres of the company is expected to lie.

Solution: Here, we are given that

$$n = 250,\ \sigma = 3200,\ \bar{X} = 50000,\ S = 3500$$

Since population standard deviation i.e. population variance σ^2 is known, therefore, we use $(1-\alpha)$ 100% confidence limits for population mean when population variance is known which are given by

$$\bar{X} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z and for 99% confidence interval, we have $1-\alpha = 0.99 \Rightarrow \alpha = 0.01$. For α = 0.01 we have, $z_{\alpha/2} = z_{0.005} = 2.58.$

Therefore, the 99% confidence limits are

$$\bar{X} \mp 2.58\frac{\sigma}{\sqrt{n}}$$

By putting the values of n, $\bar{X}$ and σ, the 99% confidence limits are

$$50000 \mp 2.58 \times \frac{3200}{\sqrt{250}}$$

$$50000 \mp 522.20 = 49477.80 \text{ and } 50522.20$$

Hence, 99% confidence interval within which the mean life of tyres of the company is expected to lie is

$$[49477.80,\ 50522.20]$$

7.4.2 Confidence Interval for Population Mean when Population Variance is Unknown

In the cases, described in previous sub-section we assume that variance σ^2 of the normal population is known but in general it is not known and in such a situation the only alternative left is to estimate the unknown σ^2. The value of sample variance (S^2) is used to estimate the σ^2 where,

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

In this situation, we know that the variate

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

follows t-distribution (described in Unit 3 of this course) with $(n-1)$ df. Therefore, probability density function of statistic t is given by

$$f(t) = \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)\sqrt{n-1}}\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}} \;;\quad -\infty < t < \infty$$

Since distribution of statistic t is independent of the parameter to be estimated, therefore, t can be taken as pivotal quantity. So we introduce two constants $t_{(n-1),\alpha/2}$ and $t_{(n-1),(1-\alpha/2)} = -t_{(n-1),\alpha/2}$ (since t-distribution is symmetrical about t = 0 line see Fig. 7.5) such that

$$P\left[-t_{(n-1),\alpha/2} \le t \le t_{(n-1),\alpha/2}\right] = 1 - \alpha \qquad \ldots (3)$$

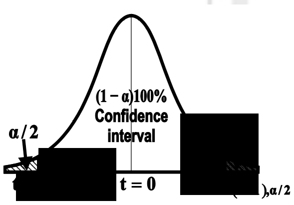

Fig. 7.5

where, $t_{(n-1),\alpha/2}$ is the value of the variate t with n–1 df having an area of $\alpha/2$ under the right tail of the probability curve of variate t as shown in Fig. 7.5.

By putting the value of variate t in equation (3), we get

$$P\left[-t_{(n-1),\alpha/2} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{(n-1),\alpha/2}\right] = 1 - \alpha$$

Now, for converting the above interval for parameter μ, we multiply each term in above inequality by $S/\sqrt{n}$ then we get

$$P\left[-t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}} \le \bar{X} - \mu \le t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

After subtracting $\bar{X}$ from each term in above inequality, we get

$$P\left[-\bar{X} - t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}} \le -\mu \le -\bar{X} + t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

Now, by multiplying each term by (-1) in above inequality, we get

$$P\left[\bar{X} + t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}} \ge \mu \ge \bar{X} - t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha \quad \left[\begin{array}{l}\because \text{by multiplying}(-1)\\ \text{the inequality is reversed}\end{array}\right]$$

This can be rewritten as

$$P\left[\bar{X} - t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{(n-1),\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

Hence, when variance σ^2 is unknown then $(1 - \alpha)$ 100% confidence interval for population mean of normal population is given by

$$\left[\bar{X} - t_{(n-1),\,\alpha/2}\frac{S}{\sqrt{n}},\ \ \bar{X} + t_{(n-1),\,\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad \dots (4)$$

and corresponding limits are given by

$$\bar{X} \mp t_{(n-1),\,\alpha/2}\frac{S}{\sqrt{n}} \qquad \dots (5)$$

Note 2: For different confidence intervals and different degrees of freedom, the values of $t_{(n-1),\,\alpha/2}$ are different. Therefore, for given values of α and n we read the tabulated value of t-statistic from the table of t-distribution (t-table) given in Appendix (at the end of Block 1 of this course) by using the method described in Unit 4 of this course.

For example, if we want to find the 95% confidence interval for μ when n = 8 then we have

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$$

From t-table, for $\alpha = 0.05$ and $\nu = n - 1 = 7$, we have the value of $t_{(n-1),\,\alpha/2} = t_{(7),0.025} = 2.365.$

As we have seen in t-table of the Appendix that when sample size is greater than 30 (n > 30) then all values of variate t are not given in this table so for convenient as we have discussed in Unit 2 of this course that when n is sufficiently large (> 30) then we know that almost all the distributions are very closely approximated by normal distribution. Thus in this case t-distribution is also approximated normal distribution. So the variate

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$

also follows the normal distribution with mean 0 and variance unity. Therefore, when population variance is unknown and sample size is large then the $(1 - \alpha)$ 100% confidence interval for population mean may be obtained by using the same procedure as we have followed in case when σ^2 is known by taking S^2 in place of σ^2 which is given as

$$\left[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad \dots (6)$$

and corresponding limits are given by

$$\bar{X} \mp z_{\alpha/2}\frac{S}{\sqrt{n}} \qquad \dots (7)$$

Following example will explain the application of the above discussion:

Example 2: It is known that the average weight of students of a Study Centre of IGNOU follows normal distribution. To estimate the average weight, a sample of 10 students is taken from this Study Centre and measured their weights (in kg) which are given below:

48, 50, 62, 75, 80, 60, 70, 56, 52, 77

Compute the 95% confidence interval for the average weight of students of Study Centre of IGNOU.

Solution: Since population variance is unknown, therefore, $(1-\alpha)$ 100% confidence limits for the average weight of students of Study Centre are given by

$$\bar{X} \mp t_{(n-1),\,\alpha/2}\frac{S}{\sqrt{n}}$$

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

Calculation for $\bar{X}$ and S:

S. No.	Weight (X)	$(X-\bar{X})$	$(X-\bar{X})^2$
1	48	–15	225
2	50	–13	169
3	62	–1	1
4	75	12	144
5	80	17	289
6	60	–3	9
7	70	7	49
8	56	–7	49
9	52	–11	121
10	77	14	196
Sum	$\sum X = 630$		$\sum\left(X-\bar{X}\right)^2 = 1252$

From the above calculation, we have

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{10}\times 630 = 63$$

$$S^2 = \frac{1}{n-1}\sum\left(X-\bar{X}\right)^2 = \frac{1}{9}\times 1252 = 139.11$$

$$\Rightarrow S = \sqrt{139.11} = 11.79$$

For 95% confidence interval, we have $1-\alpha = 0.95 \Rightarrow \alpha = 0.05.$ Also from t-table, we have, $t_{(n-1),\alpha/2} = t_{(9),0.025} = 2.306.$

Thus, the 95% confidence limits are

$$\bar{X} \mp t_{(n-1),\,0.025}\frac{S}{\sqrt{n}} = 63 \mp 2.306 \times \frac{11.79}{\sqrt{10}}$$

$$= 63 \mp 8.60 = 54.40 \text{ and } 71.60$$

Hence, required 95% confidence interval for the average weight of students of Study Centre of IGNOU is given by

$$\left[54.4,\ 71.60\right]$$

Example 3: The mean life of 100 electric blubs produced by a company is 2550 hours with a standard deviation 54 hours. Find 95% confidence limits for population mean life of electric blubs produced by the company.

Solution: Here, we are given that

$$n = 100,\ \bar{X} = 2550,\ S = 54$$

Since population variance is unknown and sample size is large (> 30), therefore, we can use (1−α) 100 % confidence limits for population mean which are given by

$$\bar{X} \mp z_{\alpha/2}\frac{S}{\sqrt{n}}$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z. For 95% confidence limits, we have $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and for $\alpha = 0.05$, we have $z_{\alpha/2} = z_{0.025} = 1.96$.

Thus, the 95% confidence limits for mean life of electric bulbs are

$$\bar{X} \mp 1.96\frac{S}{\sqrt{n}} = 2550 \mp 1.96\frac{54}{\sqrt{100}}$$

$$= 2550 \mp 1.96 \times 5.4$$

$$= 2550 \mp 10.58 = 2539.42 \text{ and } 2560.58$$

Now, it is time for you to try the following exercises to make sure that you have learnt about the confidence interval for population mean in different cases.

E4) Certain refined oil is packed in tins holding 15 kg each. The filling machine maintains this but have a standard deviation 0.30 kg. A sample of 200 tins is taken from the production line. If sample mean is 15.25 kg then find the 95% confidence interval for the average weight of oil tins.

E5) Sample mean of weights (in kg) of 150 students of IGNOU is found to be 65 kg with standard deviation 12 kg. Find the 95% confidence limits in which the average weight of all students of IGNOU expected to lie.

E6) It is known that the average height of cadets of a centre follows normal distribution. A sample of 6 cadets of the centre was taken and measured their heights (in inch) which are given below:

70	72	80	82	78	80

From this data, estimate the 95% confidence limits for the average height of cadets of the particular centre.

7.5 CONFIDENCE INERVAL FOR POPULATION PROPORTION

In Section 7.4, we have discussed the confidence interval for population mean. But in many real word situations, in business and other areas, the data are collected in form of counts or the collected data classified into two categories or groups according to an attribute or characteristic under study. Generally, such types of data are considered in terms of proportion of elements / individuals / units / items possess or not possess a given characteristic or attribute. For example, the proportion of female in the population, proportion of diabetes patients in a hospital, proportion of Science books in a library, proportion of defective articles in a lot, etc.

In such situations, we deal population proportion instead of population mean and one may want to obtain the confidence interval for population proportion.

For example, a sociologist may want to know the confidence interval for proportion of female in the population of a state. A doctor may want to know the confidence interval for proportion of diabetes patients in a hospital, a product manager may want to know the confidence interval for proportion of defective articles in a lot, etc.

Generally, population proportion is estimated by sample proportion.

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a population with population proportion P. Also let X denotes the number of observations or elements possess a certain attribute (successes) out of n observations of the sample then sample proportion p can be defined as

$$p = \frac{X}{n} \le 1$$

As we have seen in Section 2.4 of the Unit 2 of this course that mean and variance of the sampling distribution of sample proportion are

$$E(p) = P \text{ and } Var(p) = \frac{PQ}{n}$$

where, Q = 1– P.

But sample proportion is generally considered for large sample so if sample size is sufficiently large, such that np > 5 and nq > 5 then by central limit theorem, the sampling distribution of sample proportion p is approximately normally distributed with mean P and variance PQ/n. Therefore, the variate

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$$

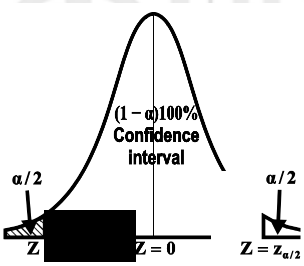

Fig. 7.6

is approximately normally distributed with mean 0 and variance unity. Since distribution of Z is independent of parameter P so it can be taken as pivotal quantity, therefore, we introduce two constants $z_{\alpha/2}$ and $z_{(1-\alpha/2)} = -z_{\alpha/2}$ such that

$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z as shown in Fig. 7.6.

Putting the value of Z, we get

$$P\left[-z_{\alpha/2} \le \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \le z_{\alpha/2}\right] = 1 - \alpha \qquad \dots (8)$$

For large sample, the variance P (1 - P) /n can be estimated by p(1 - p)/n therefore, putting p(1 - p) / n in place of P (1 - P) / n in equation (8), we get

$$P\left[-z_{\alpha/2} \le \frac{p - P}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right] = 1 - \alpha$$

Now, for converting the above interval for parameter P, we multiplying each term by $\sqrt{\frac{p(1-p)}{n}}$ and then subtracting p from each term in the above inequality, we get

$$P\left[-p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le -P \le -p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha$$

Now, by multiplying each term by (-1) in above inequality, we get

$$P\left[p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \ge P \ge p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha \quad \left[\begin{array}{l}\because \text{by multiplying} \\ (-1)\text{the inequality} \\ \text{is reversed}\end{array}\right]$$

This can be written as

$$P\left[p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le P \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha$$

Hence, $(1-\alpha)$ 100% confidence interval for population proportion is given by

$$\left[p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}},\ p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] \qquad \ldots (9)$$

Therefore, corresponding confidence limits are

$$p \mp z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \qquad \ldots (10)$$

Following example will explain the application of the above discussion:

Example 4: A sample of 200 voters is chosen at random from all voters in a given city. 60% of them were in favour of a particular candidate. If large number of voters cast their votes then find 99% and 95% confidence intervals for the proportion of voters in favour of a particular candidate.

Solution: Here, we are given

$n = 200,\ p = 0.60$

First we check the condition of normality as

$\because np = 200 \times 0.60 = 120 > 5$ and $nq = 200 \times (1-0.60) = 200 \times 0.40 = 80 > 5$ so $(1-\alpha)100\%$ confidence limits for the proportion are given by

$$p \mp z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

For 99% confidence interval, we have $1-\alpha = 0.99 \Rightarrow \alpha = 0.01$. For $\alpha = 0.01$, we have $z_{0.005} = 2.58$ and for $\alpha = 0.05$, $z_{0.025} = 1.96$.

Therefore, 99% confidence limits of voters in favour of a particular candidate are

$$\begin{aligned} p \mp z_{0.005}\sqrt{\frac{p(1-p)}{n}} &= 0.60 \mp 2.58 \times \sqrt{\frac{0.60 \times 0.40}{200}} \\ &= 0.60 \mp 2.58 \times 0.03 \\ &= 0.60 \mp 0.08 = 0.52 \text{ and } 0.68 \end{aligned}$$

Hence, required 99% confidence interval for the proportion of voters in favour of a particular candidate is given by

[0.52, 0.68]

Similarly, 95% confidence limits are given by

$$p \mp z_{0.025}\sqrt{\frac{p(1-p)}{n}} = 0.60 \mp 1.96 \times 0.03$$

$$= 0.60 \mp 0.06 = 0.54 \text{ and } 0.66$$

Hence, 95% confidence interval for the proportion of voters in favour of a particular candidate is given by

[0.54, 0.66]

It is your time to try the following exercise.

E7) A random sample of 400 apples was taken from a large consignment and 80 were found to be bad. Obtain the 99% confidence limits for the proportion of bad apples in the consignment.

7.6 CONFIDENCE INTERVAL FOR POPULATION VARIANCE

In Sections 7.4 and 7.5, we discussed the confidence interval for population mean and proportion respectively. But there are many practical situations where one may be interested to obtain the interval estimate of the population variance. For example, a manufacturer of steel ball bearings may want to obtain the interval estimate of the variation of diameter of steel ball bearing, an economist may wish to know the interval estimate for the variability in income of the person living in a city, etc.

Similar to confidence interval for the population mean we can determine confidence interval for population variance into two following cases:

1. When population mean is known and
2. When population mean is unknown.

These two cases are described one by one, in Sub-sections 7.6.1 and 7.6.2 respectively.

7.6.1 Confidence Interval for Population Variance when Population Mean is Known

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population having mean μ and variance σ^2 where μ is known, that is, μ has a specified value. In this case, we know that the variate

$$\chi^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2_{(n)}$$

follows the chi-square distribution with n degrees of freedom whose probability density function is given by

$$f(\chi^2) = \frac{1}{2^{n/2}\left\lceil n/2 \right.} e^{-\chi^2/2}(\chi^2)^{\frac{n}{2}-1}; \;\; 0 < \chi^2 < \infty$$

Since distribution of χ^2 is independent of parameter σ^2 therefore, χ^2 can be taken as pivotal quantity, therefore, we introduce two constants $\chi^2_{(n),\,\alpha/2}$ and $\chi^2_{(n),\,(1-\alpha/2)}$ such that

$$P\left[\chi^2_{(n),\,(1-\alpha/2)} \le \chi^2 \le \chi^2_{(n),\,\alpha/2}\right] = 1-\alpha \qquad \ldots (11)$$

where, $\chi^2_{(n),\,\alpha/2}$ and $\chi^2_{(n),\,(1-\alpha/2)}$ are the value of the χ^2 variate at n df having area of $\alpha/2$ under the right tail and $\alpha/2$ under the left tail respectively of the probability curve of χ^2 as shown in Fig. 7.7.

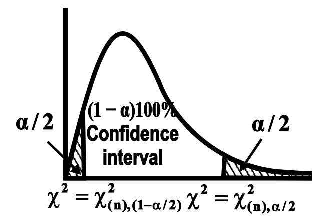

Fig. 7.7

Putting the value of χ^2 in equation (11), we get

$$P\left[\chi^2_{(n),\,(1-\alpha/2)} \le \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \le \chi^2_{(n),\,\alpha/2}\right] = 1-\alpha$$

Now, for converting this interval for σ^2, we divide each term by $\sum_{i=1}^{n}(X_i-\mu)$ in above inequality then we get

$$P\left[\frac{\chi^2_{(n),(1-\alpha/2)}}{\sum_{i=1}^{n}(X_i-\mu)^2} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{(n),\,\alpha/2}}{\sum_{i=1}^{n}(X_i-\mu)^2}\right] = 1-\alpha$$

Reciprocal each term of the above inequality

$$P\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,(1-\alpha/2)}} \ge \sigma^2 \ge \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,\alpha/2}}\right] = 1-\alpha \quad \left[\because \begin{array}{c}\text{by reciprocaling, the}\\ \text{inequality is reversed}\end{array}\right]$$

This can be written as

$$P\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,\alpha/2}} \le \sigma^2 \le \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,(1-\alpha/2)}}\right] = 1-\alpha$$

Hence, $(1-\alpha)100\%$ confidence interval for population variance when population mean is known in normal population is given by

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,\alpha/2}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,(1-\alpha/2)}}\right] \qquad \ldots (12)$$

and the corresponding $(1-\alpha)$ 100% confidence limits are given by

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,\alpha/2}} \text{ and } \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,(1-\alpha/2)}} \qquad \ldots (13)$$

Note 3: For different confidence interval and degrees of freedom the values of $\chi^2_{(n),\,\alpha/2}$ and $\chi^2_{(n),\,(1-\alpha/2)}$ are different. Therefore, for given values of α and n we read the tabulated value of these from the table of χ^2-distribution **(χ^2-table)** (given in

Appendix at the end of Block 1 of this curse) by the method described in Unit 4 of this course.

For example, if we want to find the 95% confidence interval for σ^2 then

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05$$

From the χ^2-table, for α = 0.05 and if n =10, we have

$$\chi^2_{(n),\,\alpha/2} = \chi^2_{(10),\,0.025} = 20.48 \text{ and } \chi^2_{(n),\,(1-\alpha/2)} = \chi^2_{(10),\,0.975} = 3.25$$

Therefore, 95% confidence interval for variance is given by

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{20.48}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{3.25}\right]$$

Following example will explain the application of the above discussion.

Example 5: Diameter of steel ball bearing produced by a company is known to be normally distributed. To know the variation in the diameter of steel ball bearings, the product manager takes a random sample of 10 ball bearings from the lot having average diameter 5.0 cm and measures diameter (in cm) of each selected ball bearing. The results are given below:

S. No.	1	2	3	4	5	6	7	8	9	10
Diameter	5.0	5.1	5.0	5.2	4.9	5.0	5.0	5.1	5.1	5.2

Find the 95% confidence interval for variance in the diameter of steel ball bearings of the lot from which the sample is drawn.

Solution: Here, we are given that

$$n = 10, \mu = 5.0$$

Since population mean is given, therefore, we use $(1-\alpha)100\%$ confidence interval for population variance when population mean is known which is given by

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,\alpha/2}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(n),\,(1-\alpha/2)}}\right]$$

Calculation for $\sum_{i=1}^{n}(X_i-\mu)^2$:

S. No.	**Diameter (X)**	$(X-\mu)$	$(X-\mu)^2$
1	5.0	0	0
2	5.1	0.1	0.01
3	5.0	0	0
4	5.2	0.2	0.04
5	4.9	–0.1	0.01
6	5.0	0	0
7	5.0	0	0
8	5.1	0.1	0.01
9	5.1	0.1	0.01
10	5.2	0.2	0.04
Total			$\sum(X-\mu)^2 = 0.12$

From the above calculation, we have

$$\sum(X_i - \mu)^2 = 0.12$$

For 95% confidence interval, we have $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ then from **χ^2-table**, we have

$$\chi^2_{(n),\,\alpha/2} = \chi^2_{(10),\,0.025} = 20.48 \text{ and } \chi^2_{(n),(1-\alpha/2)} = \chi^2_{(10),0.975} = 3.25$$

Thus, 95% confidence interval for variance in the diameter of steel ball bearings of the lot is given by

$$\left[\frac{0.12}{20.48}, \frac{0.12}{3.25}\right]$$

or $\quad [0.0059, 0.0369]$

7.6.2 Confidence Interval for Population Variance when Population Mean is Unknown

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population with unknown mean μ and variance σ^2. In this case, the value of sample mean $\bar{X}$ is used to estimate μ. As we have seen in Section 4.2 of Unit 4 of this course that the variate

$$\chi^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad \text{where, } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

follows the chi-square distribution with (n - 1) degrees of freedom whose probability density function is given by

$$f(\chi^2) = \frac{1}{2^{\frac{(n-1)}{2}}\left\lceil\frac{n-1}{2}\right.} e^{-\chi^2/2}\left(\chi^2\right)^{\frac{n-1}{2}-1}; \quad 0 < \chi^2 < \infty$$

Since distribution of χ^2 is independent of parameter to be estimated, therefore, χ^2 can be taken as pivotal quantity. So we introduce two constants $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ such that

$$P\left[\chi^2_{(n-1),(1-\alpha/2)} \le \chi^2 \le \chi^2_{(n-1),\alpha/2}\right] = 1-\alpha \qquad \ldots (14)$$

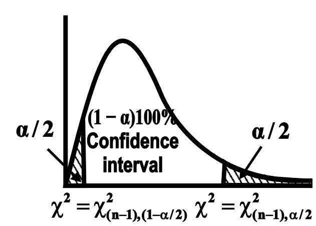

Fig. 7.8

where, $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ are the value of the χ^2-variate at (n –1) df having area of $\alpha/2$ under the right tail and $\alpha/2$ under the left tail of the probability curve of χ^2 as shown in Fig. 7.8.

Putting the value of χ^2 in equation (14), we get

$$P\left[\chi^2_{(n-1),(1-\alpha/2)} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{(n-1),\,\alpha/2}\right] = 1-\alpha$$

Now, for converting this interval for σ^2 we dividing each term in above inequality by $(n-1)S^2$ then we get

$$P\left[\frac{\chi^2_{(n-1),\,(1-\alpha/2)}}{(n-1)S^2} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{(n-1),\,\alpha/2}}{(n-1)S^2}\right] = 1-\alpha$$

By taking reciprocal of each term of above inequality, we get

$$P\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,(1-\alpha/2)}} \ge \sigma^2 \ge \frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}}\right] = 1-\alpha \quad \left[\begin{array}{l}\because \text{by reciprocaling, the} \\ \text{inequality is reversed}\end{array}\right]$$

This can be written as

$$P\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{(n-1),\,(1-\alpha/2)}}\right] = 1-\alpha$$

Hence, the $(1-\alpha)$ 100% confidence interval for population variance when population mean is unknown is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}},\ \frac{(n-1)S^2}{\chi^2_{(n-1),(1-\alpha/2)}}\right] \qquad \ldots (15)$$

and corresponding confidence limits are

$$\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}} \text{ and } \frac{(n-1)S^2}{\chi^2_{(n-1),(1-\alpha/2)}} \qquad \ldots (16)$$

where, $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ are the values of χ^2-variate at $(n-1)$ degrees of freedom and the values of these can be read from χ^2-table. For example, if we want to find the 95% confidence interval for σ^2 then

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05$$

From χ^2-table for α = 0.05 and n = 10 degrees of freedom, we have

$$\chi^2_{(n-1),\,\alpha/2} = \chi^2_{(9),\,0.025} = 19.02 \text{ and } \chi^2_{(9),0.975} = 2.70.$$

Therefore, 95% confidence interval for variance is given by

$$\left[\frac{9S^2}{19.02},\ \frac{9S^2}{2.70}\right]$$

Let us do an example based on above discussion.

Example 6: A random sample of 10 workers is taken from a factory. The wages (in hundreds) per months of these workers are given below:

48, 50, 62, 75, 80, 60, 70, 56, 52, 77

Obtain 95% confidence interval for the variance of wages of all the workers of the factory.

Solution: Here, the population mean is unknown therefore, we use $(1-\alpha)$ 100% confidence interval for population variance when population mean is unknown which is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}},\frac{(n-1)S^2}{\chi^2_{(n-1),\,(1-\alpha/2)}}\right]$$

where, $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ are the values of χ^2 variate at (n−1) degrees of freedom, whereas

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

Calculation for $\bar{X}$ and S^2 :

S. No.	Weight (X)	$\left(X-\bar{X}\right)$	$\left(X-\bar{X}\right)^2$
1	48	–15	225
2	50	–13	169
3	62	–1	1
4	75	12	144
5	80	17	289
6	60	–3	9
7	70	7	49
8	56	–7	49
9	52	–11	121
10	77	14	196
Sum	$\sum X = 630$		$\sum\left(X-\bar{X}\right)^2 = 1252$

From the above calculation, we have

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{10}\times 630 = 63$$

$$S^2 = \frac{1}{n-1}\sum\left(X-\bar{X}\right)^2 = \frac{1}{9}\times 1252 = 139.11$$

For 95% confidence interval
$1-\alpha = 0.95 \Rightarrow \alpha = 0.05$
and $\alpha/2 = 0.025$ &
$1-\alpha/2 = 0.975$.

From χ^2-table, we have $\chi^2_{(n-1),\,\alpha/2} = \chi^2_{(9),\,0.025} = 19.02$ and

$\chi^2_{(n-1),(1-\alpha/2)} = \chi^2_{(9),0.975} = 2.70$

Thus, 95% confidence interval for the variance of wages of all the workers of the factory is given by

$$\left[\frac{10\times 139.11}{19.02}, \frac{10\times 139.11}{2.70}\right] \text{ or } \left[73.14,\ 515.22\right]$$

Now, you can try following exercises to see how much you have followed.

E8) A study of variation in weights of soldiers was made and it is known that the mean weight of soldiers follows the normal distribution. A sample of 12 soldiers is taken from the soldier's population and sample variance is found 60 pound^2. Estimate the 95% confidence interval for the variance of soldier's weight of the population from which the sample was drawn.

E9) If $X_1 = -5$, $X_2 = 4$, $X_3 = 2$, $X_4 = 6$, $X_5 = -1$, $X_6 = 4$, $X_7 = 0$, $X_8 = 10$ and $X_9 = 7$ are the sample observations taken from normal population $N(\mu, \sigma^2)$, obtain confidence interval for σ^2.

7.7 CONFIDENCE INTERVAL FOR NON-NORMAL POPULATIONS

So far in this unit, we have kept our discussion on the confidence interval of the normal population except population proportion. But one may be interested to find out the confidence interval when the population under study is not normal. The aim of this section is to give an idea how we can obtain confidence interval for non-normal populations. For example, one may be interested to estimate, say, 95% confidence interval of parameter θ when population under study follows exponential distribution(θ).

We know that when the sample size is large then almost all the sampling distributions of the statistics $\bar{X}$, S^2, etc. follow normal distribution. So when the sample is large we can also obtain the confidence interval as follows:

Let $X_1, X_2, ..., X_n$ be a random sample of size n (sufficiently large $\text{i.e.}\, n > 30$) taken from f(x,θ) then according to the central limit theorem the sampling distribution of sample mean $\bar{X}$ is normal, that is,

$$\bar{X} \sim N\left[E\left(\bar{X}\right),\ Var\left(\bar{X}\right)\right]$$

Then the variate

$$Z = \frac{\bar{X} - E\left(\bar{X}\right)}{\sqrt{Var\left(\bar{X}\right)}} \sim N\left(0,1\right)$$

is approximately normally distributed with mean 0 and variance unity. Since distribution of Z is independent of parameter so it can be taken as pivotal quantity therefore we introduce two constants $z_{\alpha/2}$ and $z_{(1-\alpha/2)} = -z_{\alpha/2}$ such that

$$P\left[-z_{\alpha/2} \le Z \le z_{\alpha/2}\right] = 1 - \alpha \qquad \ldots (17)$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z.

By putting the value of Z in equation (17), we get

$$P\left[-z_{\alpha/2} \le \frac{\bar{X} - E\left(\bar{X}\right)}{\sqrt{Var\left(\bar{X}\right)}} \le z_{\alpha/2}\right] = 1 - \alpha$$

After this, we have to convert this interval for parameter θ as discussed in Section 7.3.

Following example will explain the procedure more clearly.

Example 7: Obtain 95% confidence interval to estimate θ when a large sample is taken from exponential population whose probability density function is given by

$$f\left(x,\theta\right) = \theta e^{-\theta x}\ ; \qquad x > 0,\ \theta > 0$$

Solution: Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from exponential population whose probability distribution is given by

$$f\left(x,\theta\right) = \theta e^{-\theta x}\ ; \qquad x > 0,\ \theta > 0$$

We know that for exponential distribution

$$E(X)=\frac{1}{\theta} \text{ and } Var(X)=\frac{1}{\theta^2}$$

Since $X_1, X_2, ..., X_n$ are independent and come from same exponential distribution, therefore,

$$E(X_i)=E(X)=1/\theta \text{ and } Var(X_i)=Var(X)=1/\theta^2 \text{ for all } i=1, 2, ..., n$$

Now consider,

$$E(\bar{X})=E\left[\frac{1}{n}(X_1+X_2+...+X_n)\right] \text{[By defination of sample mean]}$$

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

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

$$=\frac{1}{n}\left(n\frac{1}{\theta}\right)=\frac{1}{\theta}$$

and

$$Var(\bar{X})=Var\left[\frac{1}{n}(X_1+X_2+...+X_n)\right]$$

$$=\frac{1}{n^2}\left[Var(X_1)+Var(X_2)+...+Var(X_n)\right]\left[\begin{array}{l}\text{If X and Y are}\\ \text{two independent}\\ \text{random variable}\\ \text{then Var(aX + bY)}\\ =a^2Var(X)+b^2Var(Y)\end{array}\right]$$

$$=\frac{1}{n^2}\left(\underbrace{\frac{1}{\theta^2}+\frac{1}{\theta^2}+...+\frac{1}{\theta^2}}_{n-\text{times}}\right)$$

$$=\frac{1}{n^2}\left(n\frac{1}{\theta^2}\right)$$

$$Var(\bar{X})=\frac{1}{n\theta^2}$$

Thus, the variate

$$Z=\frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}}=\frac{\bar{X}-1/\theta}{\sqrt{1/n\theta^2}}\sim N(0,1)$$

is approximately normally distributed with mean 0 and variance unity. Since distribution of Z is independent of parameter so it can be taken as pivotal quantity, therefore we introduce two constants $z_{\alpha/2}$ and $z_{(1-\alpha/2)} = -z_{\alpha/2}$ such that

$$P\left[-z_{\alpha/2}\le Z\le z_{\alpha/2}\right]=1-\alpha$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z.

For 95% confidence interval, $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$, we have $z_{\alpha/2} = z_{0.025} = 1.96$. So confidence interval for θ is

$$P\left[-1.96\le Z\le 1.96\right]=1-\alpha$$

Putting the value of Z, we have

$$P\left[-1.96 \le \frac{\bar{X}-\frac{1}{\theta}}{\sqrt{\frac{1}{n\theta^2}}} \le 1.96\right] = 0.95$$

$$\Rightarrow P\left[-1.96 \le \frac{\frac{1}{\theta}\left(\theta\bar{X}-1\right)}{\frac{1}{\theta}\sqrt{\frac{1}{n}}} \le 1.96\right] = 0.95$$

$$\Rightarrow P\left[-1.96 \le \sqrt{n}\left(\theta\bar{X}-1\right) \le 1.96\right] = 0.95$$

$$\Rightarrow P\left[\frac{-1.96}{\sqrt{n}} \le \left(\theta\bar{X}-1\right) \le \frac{1.96}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[1-\frac{1.96}{\sqrt{n}} \le \theta\bar{X} \le 1+\frac{1.96}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[\frac{1}{\bar{X}}\left(1-\frac{1.96}{\sqrt{n}}\right) \le \theta \le \frac{1}{\bar{X}}\left(1+\frac{1.96}{\sqrt{n}}\right)\right] = 0.95$$

Hence, 95% confidence interval for parameter θ is

$$\left[\frac{1-1.96/\sqrt{n}}{\bar{X}}, \ \frac{1+1.96/\sqrt{n}}{\bar{X}}\right]$$

7.8 SHORTEST CONFIDENCE INTERVAL

It may be noted that for a confidence coefficient, there are many confidence intervals for a parameter are possible. For example, from normal table (given in the end of the Block 1 of this course) we can have many sets of a's and b's to give 95% confidence interval for μ some of them are given below:

$$P\left[-1.65 \le Z \le 3.0\right] = 0.95, \ P\left[-1.68 \le Z \le 2.70\right] = 0.95$$

$$P\left[-1.70 \le Z \le 2.54\right] = 0.95, \quad P\left[-1.96 \le Z \le 1.96\right] = 0.95, \text{etc.}$$

Therefore, we need some criterion with the help of which we may choose the best (best in the sense of minimum length) confidence interval among these set of confidence intervals.

An obvious criterion (method) of selecting the shortest one out of these is that, we chose a's and b's in such a way that the length of interval is minimum. In above case, the lengths of these intervals are

$L = T_2 - T_1$

So, $L_1 = 3.0-(-1.65) = 4.65$, $L_2 = 2.70-(-1.68) = 4.38$

$L_3 = 2.54-(-1.70) = 4.24$, $L_4 = 1.96-(-1.96) = 3.92$

Hence, the last one has the minimum length. Therefore, it is best confidence interval for μ in all the above intervals on the basis of the minimum length criterion.

7.9 DETERMINATION OF SAMPLE SIZE

So, far you have become familiar with the main goal of this block. The discussion of whole block centered on the theme of estimate some population parameter of interest. To estimate some population parameter we have to draw a random sample. A natural question which may arise in your mind "how large should my sample be?" And this is very important question which is commonly asked. From statistical point of view, the best answer of this question is "take as large sample as you can afford. That is, if possible 'sample' the entire population that means study all units of the population under study because by taking all units of the population we will have all the information about the parameter and we will know the exact value of population parameter which is better than any estimate of that parameter. Generally, this is impractical to take entire population to be sampled due to economic constraints, time constraints and other limitations. So the answer, "take as large sample as you can afford is best if we ignore all costs because as you have studied in Section 1.4 of Unit 1 of this course that the larger the sample size smaller the standard error of the statistic that means less the uncertainty.

When the recourses in terms of money and time are limited then the question is "how to found the minimum sample size which will satisfy some precision requirements" In such cases, we should require first the answers of the following three questions about his / her requirement about the survey:

1. How close do you want your sample estimate to be to the unknown parameter? That means what should be the allowable difference between the sample estimate and true value of population parameter. This difference is known as sampling error or margin of error and represented by E.
2. The next question is, what do you want the confidence level to be so that the difference between the estimate and the parameter is less than or equal to E? That is, 90%, 95%, 99%, etc.
3. The last question is what is the population variance or population proportion as may be the case?

When we have the answers of these three questions then we will get an answer of the minimum required sample size.

In this section, we will describe the procedure of determining of sample size for estimating population mean and population proportion.

Determination of Minimum Sample Size for Estimating Population Mean

For determining minimum sample size for estimating population mean we use confidence interval. Since as you seen in Section 7.4 of this unit that confidence interval for population mean depends upon the nature of population and population variance (σ^2) is known or unknown so following cases may arise:

Case I: Population is normal and population variance (σ^2) is known

If σ^2 is known and population is normal then we know that $(1-\alpha)100\%$ confidence interval for population mean is given by

$$P\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Also $$P\left[-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha \qquad \ldots (18)$$

Since the normal distribution is symmetric so we can concentrate on the right-hand equality so

$$\bar{X}-\mu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \ldots (19)$$

This inequality implies that the largest value that the difference $\bar{X}-\mu$ can assume is $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Also the difference between the estimator (sample mean $\bar{X}$) and the population parameter (population mean μ) is called the sampling error so

$$E = \bar{X}-\mu = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \ldots (20)$$

Solving this equation for n, we have

$$n = \frac{z_{\alpha/2}^2\sigma^2}{E^2} \qquad \ldots (21)$$

When population is finite of size N and sampling is to be done without replacement then finite population correction $\sqrt{\frac{N-n}{N-1}}$ is required so equation (20) becomes

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

which gives

$$n = \frac{N\,z_{\alpha/2}^2\sigma^2}{E^2\left(N-1\right)+z_{\alpha/2}^{\;2}\sigma^2} \qquad \ldots (22)$$

If the finite population correction is ignored then equation (22) is reduced to equation (21).

Case II: Population is non-normal and population variance (σ^2) is known

If the population is not assumed to be normal and the population variance σ^2 is known then by central limit theorem we know that sampling distribution of mean approximate normally distributed as sample size increases. So the above method can be used determining minimum sample size. Once the required sample size is obtained, we can check to see it that sample size greater than 30 and if it does we may be confident that our method of solution was appropriate.

Case III: Population is normal or non-normal and population variance (σ^2) is unknown

In this case, we use value of sample variance (S^2) to estimate the population variance but S^2 is also calculated from a sample and we have not take a sample yet. So in this case, determination of sample size is not directly obtained. The most frequently methods for estimating σ^2 are as follows:

1. A pilot or preliminary sample may be drawn from the population under study and the variance computed from this sample may be used as an estimate of σ^2.
2. The variance of previous or similar studies may used to estimate σ^2.
3. If the population is assumed to be normal then we may use the fact that the

range is approximately equal to six times of standard deviation i.e. $\sigma = R/6$. This approximate require only knowledge of largest and smallest value of the variable under study because range may be defined as

R = largest value – smallest value

Determination of Minimum Sample Size for Estimating Population Proportion

The method of determination of minimum sample size for estimating population proportion is similar as that described in estimating population mean. So the formula for minimum sample size is given by

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{E^2} \qquad \ldots (23)$$

where, P is the population proportion and E = p – P is the sampling error or margin of error.

When population is finite of size N and sampling is to be done without replacement then finite population correction $\sqrt{\frac{N-n}{N-1}}$ is required so

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{E^2\left(N-1\right) + z_{\alpha/2}^2 P(1-P)} \qquad \ldots (24)$$

Here, sample size is depended upon the population proportion and it is generally unknown so the most frequently methods for estimating P are as follows:

1. A pilot or preliminary sample may be drawn from the population under study and the sample proportion computed from this sample may be used as an estimate of P.
2. The proportion of previous or similar studies may used to estimate P.
3. Absence of any information, we may use P = 0.5.

Now, it is time to do some examples based on determination of sample size.

Example 8: A hospital administrator wishes to estimate the mean weight of babies born in her hospital. How large a sample of birth records should be taken if she wants a 99 percent confidence that the estimate within the range of 0.4 pound? Assume that a reasonable estimate of σ is 0.5 pound.

Solution: Here, we are given that

E = margin of error = 0.4, confidence level = 0.99 and σ = 0.5

Also for 99% confidence, $1-\alpha = 0.99 \Rightarrow \alpha = 0.01$ and $\alpha/2 = 0.005$, we have

$z_{\alpha/2} = z_{0.005} = 2.58$.

Hospital administrator wishes to obtain the minimum sample size for estimating the mean weight of babies born in her hospital so we have

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{(2.58)^2 (0.5)^2}{(0.4)^2} = 10.40 \sim 11$$

Hence, hospital administrator should obtain a random sample of at least 11babies.

Example 9: The manufacturers of a car want to estimate the proportion of people who are interested in a certain model. The company wants to know the population proportion, P, to within 0.05 with 95% confidence. Current company records indicate that the proportion P may be around 0.20. What is the minimum required sample size for this survey?

Solution: Here, we are given that

E = margin of error = 0.05, confidence level = 0.95 and P = 0.20

Also for 95% confidence, $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$, we have $z_{\alpha/2} = z_{0.025} = 1.96$.

The manufacturers of a car interested to obtain the minimum sample size for estimating the population proportion so the required formula is given below

$$n = \frac{z^2_{\alpha/2}P(1-P)}{E^2} = \frac{(1.96)^2(0.20)(0.80)}{(0.05)^2} = 245.86 \sim 246$$

Hence, the company should require a random sample of at least 246 people.

You will be more cleared about this, when you try the following exercises.

E10) A survey is planned to determine the average annual family medical expenses of employees of a large company. The management of the company wishes to be 95% confident that the sample average is correct to within ± 100 Rs of the true average family expenses. A pilot study indicates that the standard deviation can be estimated can be estimated as 400 Rs. How large a sample size is necessary?

E 11) The manager of a bank in a small city would like to determine the proportion of the depositors per week. The manager wants to be 95% confident of being correct to within ± 0.10 of the true proportion of depositors per week. A guess is that the parentage of such depositors is about 8%, what sample size is needed?

With this we have reached end this unit. Let us summarise what we have discussed in this unit.

7.10 SUMMARY

In this unit, we have covered the following points:

1. The interval estimation.
2. The method of obtaining the confidence intervals.
3. The method of obtaining confidence interval for population mean of a normal population when variance is known and unknown
4. The method of obtaining confidence interval for population proportion of a population.
5. The method of obtaining confidence interval for population variance of a normal population when population mean is known and unknown.
6. The method of obtaining confidence interval for population parameters of non-normal populations.
7. The concept of the shortest interval.
8. Determination of sample size.

7.11 SOLUTIONS / ANSWERS

E1) We know that the length of confidence interval $[T_1, T_2]$ is given by

$$L = T_2 - T_1$$

Therefore, in our case we have

(i) $L = 3.0 - (-1.65) = 4.65$

(ii) $L = 2.70 - (-1.68) = 4.38$

(iii) $L = 2.54 - (-1.70) = 4.24$

(iv) $L = 1.96 - (-1.96) = 3.92$

E2) We know that if we have confidence interval for parameter θ is

$$P\left[T_1 \le \theta \le T_2\right] = 1 - \alpha$$

then

Lower confidence limit (LCL) = T_1

Upper confidence limit (UCL) = T_2

Confidence coefficient (CC) = $1 - \alpha$

Therefore in our cases, we have

(i) LCL = 0, UCL = 1.5 & CC = 0.90

(ii) LCL = −1, UCL = 2 & CC = 0.95

(iii) LCL = −2, UCL = 2 & CC = 0.98

(iv) LCL = −2.5, UCL = 2.5 & CC = 0.99

E3) Refer Section 7.3.

E4) Here, we are given that

$$n = 200,\ \sigma = 0.30, \bar{X} = 15.25$$

Since population standard deviation or variance is known so we use $(1-\alpha)$ 100% confidence limits for population mean when population variance is known which are given by

$$\bar{X} \mp z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where, $z_{\alpha/2}$ is the value of the variate Z having an area of $\alpha/2$ under the right tail of the probability curve of Z. For 95% confidence interval, we have $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ and for $\alpha = 0.05$, we have, $z_{\alpha/2} = z_{0.025} = 1.96.$

Thus, the 95% confidence limits for the average weight of oil tins are

$$\bar{X} \mp 1.96 \frac{\sigma}{\sqrt{n}} = 15.25 \mp 1.96 \times \frac{0.30}{\sqrt{200}}$$

$$= 15.25 \mp 0.04 = 15.21 \text{ and } 15.29$$

E5) Here, we are given that

$$n = 150,\quad \bar{X} = 65,\ S = 12$$

Since population variance is unknown and sample size is large sample, so we use (1 - α) 100% confidence limits for population mean when population variance is unknown which are given by

$$\bar{X} \mp z_{\alpha/2} \frac{S}{\sqrt{n}}$$

For 95% confidence interval, we have $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and for α = 0.05, we have, $z_{\alpha/2} = z_{0.025} = 1.96$.

Thus, the 95% confidence limits in which the average weight of all students of a Study centre of IGNOU expected to lie are given by

$$\bar{X} \mp 1.96 \frac{S}{\sqrt{n}} = 65 \mp 1.96 \times \frac{12}{\sqrt{150}}$$

$$= 65 \mp 1.92 = 63.08 \text{ and } 66.92$$

Hence, required confidence limits are

63.08 and 66.92.

E6) Since population variance is unknown and sample size is small (n < 30), therefore, (1 - α) 100% confidence limits for average height of credits of particular centre are given by

$$\bar{X} \mp t_{(n-1),\, \alpha/2} \frac{S}{\sqrt{n}}$$

where, $\bar{X} = \frac{1}{n}\sum X$, $S = \sqrt{\frac{1}{n-1}\sum\left(X-\bar{X}\right)^2}$

Calculation for $\bar{X}$ and S:

S. No	X	$(X-\bar{X})$	$(X-\bar{X})^2$
1	70	-7	49
2	72	-5	25
3	80	3	9
4	82	5	25
5	78	1	1
6	80	3	9
	$\sum X = 462$		$\sum\left(X-\bar{X}\right)^2 = 118$

Form the above calculation, we have

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{6} \times 462 = 77$$

$$S^2 = \frac{1}{n-1}\sum\left(X-\bar{X}\right)^2 = \frac{1}{6-1} \times 118 = 23.6$$

$$\Rightarrow S = 4.86$$

For 95% confidence interval $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$.

From t-table, we have $t_{(n-1),\alpha/2} = t_{(5),0.025} = 2.571$.

Thus, 95% confidence limits are given by

$$\bar{X} \mp t_{(n-1),\, 0.025} \frac{S}{\sqrt{n}} = 77 \mp 2.571 \times \frac{4.86}{\sqrt{6}}$$

$$= 77 \mp 2.571 \times 1.98$$

$$= 77 \mp 5.09 = 71.91 \text{ and } 82.09$$

E7) We have

$$n = 400,\ X = 80$$

$$p = \frac{X}{n} = \frac{80}{400} = \frac{1}{5} = 0.20$$

$\because np = 400 \times 0.20 = 80 > 5$ and $nq = 400 \times (1 - 0.20) = 400 \times 0.80 = 320 > 5$ so $(1 - \alpha)100\%$ confidence limits for the proportion are given by

$$p \mp z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

For $\alpha = 0.01$, we have, $z_{\alpha/2} = z_{\alpha/2} = 2.58$.

For 99% confidence interval $1 - \alpha = 0.99 \Rightarrow \alpha = 0.01$ and $\alpha/2 = 0.005$.

Therefore, 99% confidence limits are

$$0.20 \mp 2.58\sqrt{\frac{0.20 \times 0.80}{400}}$$

$$= 0.2 \mp 2.58 \times 0.02$$

$$= 0.2 \mp 0.05 = 0.15 \text{ and } 0.25$$

E8) Here, we are given that

$$n = 12,\ S^2 = 60$$

Since population mean is unknown, therefore, we use $(1-\alpha)$ 100% confidence interval for population variance when population mean unknown which is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{(n-1),\,(1-\alpha/2)}}\right]$$

For 95% confidence interval $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$.

where, $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ are the values of χ^2 variate at $(n-1)$ degrees of freedom. From χ^2-table, we have

$$\chi^2_{(n-1),\,\alpha/2} = \chi^2_{(11),\,0.025} = 21.92 \text{ and } \chi^2_{(n-1),(1-\alpha/2)} = \chi^2_{(11),0.975} = 3.82$$

Thus, 95% confidence interval for the variance of soldier's weight is given by

$$\left[\frac{11 \times 60}{21.92}, \frac{11 \times 60}{3.82}\right] \Rightarrow [30.11,\ 172.77]$$

E9) Since population mean is unknown, therefore, we use $(1-\alpha)$ 100% confidence interval for population variance which is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{(n-1),\,(1-\alpha/2)}}\right]$$

where, $\chi^2_{(n-1),\,\alpha/2}$ and $\chi^2_{(n-1),\,(1-\alpha/2)}$ are the values of χ^2 variate at $(n-1)$ degrees of freedom, whereas

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

Calculation for $\bar{X}$ and S^2:

X	$(X-\bar{X})$	$(X-\bar{X})^2$
-5	-8	64
4	1	1
2	-1	1
6	3	9
-1	-4	16
4	1	1
0	-3	9
10	7	49
7	4	16
$\sum X = 27$		$\sum (X-\bar{X})^2 = 166$

Therefore,

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{9}(27) = 3$$

Also, $(n-1)S^2 = \sum\left(X-\bar{X}\right)^2 = 166$ and from χ^2-table, we have

For 99% confidence interval
$1-\alpha = 0.99 \Rightarrow \alpha = 0.01$
and $\alpha/2 = 0.005$

$\chi^2_{(n-1),\,\alpha/2} = \chi^2_{(8),\,0.005} = 21.96$ and $\chi^2_{(n-1),(1-\alpha/2)} = \chi^2_{(8),0.995} = 1.34$

Thus, 99% confidence interval for σ^2 is

$$\left[\frac{166}{21.96}, \frac{166}{1.34}\right] \text{ or } [7.56, 123.88]$$

E10) Here, we are given that

E = margin of error = 100, confidence level = 0.95 and σ = 400

Also for 95% confidence, $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$, we have $z_{\alpha/2} = z_{0.025} = 1.96$.

The management of the company wishes to obtain the minimum sample size for estimating the average annual family medical expenses of employees of the company so the required formula is given below

$$n = \frac{z^2_{\alpha/2}\sigma^2}{E^2} = \frac{(1.96)^2(400)^2}{(100)^2} = 61.46 \sim 62$$

Hence, the management of the company should require a random sample of at least 62 employees.

E 11) Here, we are given that

E = margin of error = 0.1, confidence level = 0.95 and P = 0.08

Also for 95% confidence, $1-\alpha = 0.95 \Rightarrow \alpha = 0.05$ and $\alpha/2 = 0.025$, we have $z_{\alpha/2} = z_{0.025} = 1.96$.

The manager wants to obtain the minimum sample size for determining the proportion so the required formula is given below

$$n = \frac{z^2_{\alpha/2}P(1-P)}{E^2} = \frac{(1.96)^2(0.08)(0.94)}{(0.05)^2} = 115.56 \sim 116$$

Hence, the manager should require a random sample of at least 116 depositors.