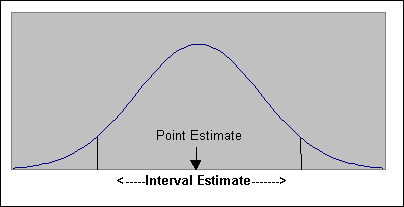
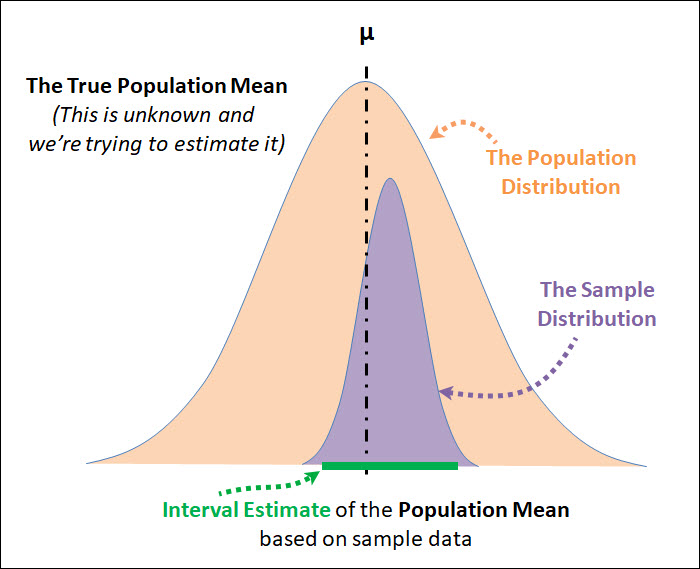
**Notes– Ch 8 Interval Estimation**



**Interval Estimate:** An interval estimate describes a range of values within which a population parameter is likely to lie. The problem with using a [point estimate](http://itfeature.com/category/estimate-and-estimation/point-and-inteval-estimation) is that although it is the single best guess you can make about the value of a population parameter, it is also usually wrong. Interval estimate overcomes this problem using interval estimation technique which is based on point estimate and margin of error.



**Interval Estimate of Mean:** The interval estimate of population mean is of the form

Sample mean ± margin of error

**Margin of Error:** The ± value added to and subtracted from a point estimate in order to develop an interval estimate of a population parameter.

There are 2 cases for interval estimate of mean:

1. σ Known: The population parameter σ is available to compute the margin of error.
2. σ Unknown: σ is not available, and sample standard deviation s is used to compute the margin of error.

**1. For** σ **Known:**

In most applications σ is not known, and s is used to compute the margin of error. In some applications, however, large amounts of relevant historical data are available and can be used to estimate the population standard deviation prior to sampling. Also, in quality control applications where a process is assumed to be operating correctly, or “in control,” it is appropriate to treat the population standard deviation as known. We refer to such cases as the σ known case.

**Confidence Level:** The probability that the interval estimate contains the true population parameter is given by the confidence level. This probability indicated how confident we are that the interval estimate will include the population parameter. A higher probability means more confidence. In estimation, the most commonly used confidence levels are 90%, 95% and 99%, but we are free to apply any confidence level. So if we have a 95% confidence level, we can be confident that 95% of the time, our interval estimate will contain the population mean. And 5% of the times our interval estimate will not contain the population mean.

**Confidence Coefficient (1- α):** The confidence coefficient give the probability that the interval estimate will contain the population parameter. α is the risk that you will not accurately contain the true population parameter.

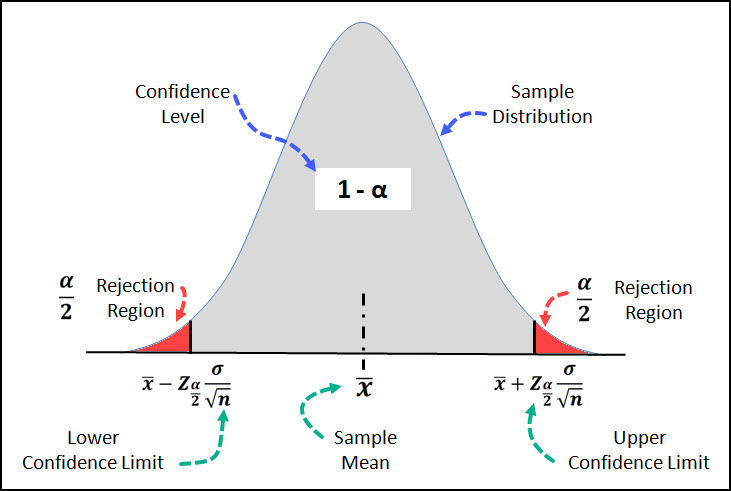
Confidence Coefficient = 1 – α

So if your confidence level is 95% then your Confidence Coefficient or (1- α) is 0.95

Which means that α = 0.05 and α/2 = 0.025

zα/2 = 1.960 provides area of 0.025 in the upper tail of the standard normal distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Confidence**  **Level** | **Confidence Coefficient**  **(1 - α)** | **α** | **α/2** | **zα/2** |  |
| 90% | 0.90 | 0.10 | 0.05 | 1.645 | 90% confidence level means zα/2 = 1.645 provides area of α/2 = 0.05 in the upper tail of the standard normal distribution |
| 95% | 0.95 | 0.05 | 0.025 | 1.960 | 95% confidence level means zα/2 = 1.960 provides area of α/2 = 0.025 in the upper tail of the standard normal distribution |
| 99% | 0.99 | 0.01 | 0.005 | 2.576 | 99% confidence level means zα/2 = 2.576 provides area of α/2 = 0.005 in the upper tail of the standard normal distribution |



**Confidence Interval:** It is the range of the estimate we are making. We express the confidence interval in standard errors rather than in numerical values.

A high confidence level seems to signify a high degree of accuracy in the estimate. In practice, however, high confidence levels will produce large confidence intervals, and such large intervals are not precise; they give fuzzy estimates.

Interval estimate is given by Sample mean ± margin of error.

Interval Estimate is given where (1 - α) is the confidence coefficient and zα/2 is the z value providing an area of α/2 in the upper tail of the standard normal probability distribution.

**Example:** Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius) 102.5, 101.7, 103.1, 100.9, 100.5, and 102.2 on 6 different samples of the liquid. He calculates the sample mean to be 101.82. If he knows that the standard deviation for this procedure is 1.2 degrees, what is the interval estimation for the population mean at a 95% confidence level?

**Ans.** Here,

At 95% confidence interval, confidence coefficient(1-α) = 0.95

So, α/2 = 0.025 and

0.96

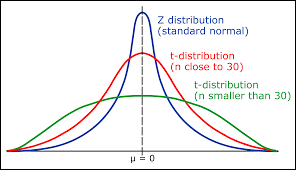
Interval estimate is given by given

2. **For** σ **Unknown:**

When *s* is used to estimate *σ*, the margin of error and the interval estimate for the population mean are based on a probability distribution known as the *t* distribution.

**Students’ t-distribution or t-distribution:** The *t-*distribution is a family of similar probability distributions, with a specific *t-*distribution depending on a parameter known as the degrees of freedom.

**Degree of freedom:** The degrees of freedom refers to the number of independent observations in a set of data. When the t-distribution is used in the computation of an interval estimate of a population mean, the appropriate *t-d*istribution has (n – 1) of freedom, where *n* is the size of the simple random sample.

The *t-*distribution with one degree of freedom is unique, as is the *t*-distribution with two degrees of freedom, with three degrees of freedom, and so on. As the number of degrees of freedom increases, the difference between the *t-*distribution and the standard normal distribution becomes smaller and smaller.

A *t-*distribution with more degrees of freedom exhibits less variability and more closely resembles the standard normal distribution. Note also that the mean of the *t-*distribution is zero.

For sample size > 30, we use normal distribution to approximate t-distribution.

We place a subscript on *t* to indicate the area in the upper tail of the *t* distribution.

Interval Estimate is given where (1 - α) is the confidence coefficient and tα/2 is

the t value providing an area of α/2 in the upper tail of the t-distribution with n – 1 degree of freedom.

**Procedure for Finding** tα/2:

1. Go across the row marked “Confidence level” in the t table until you find the column with the desired confidence level(area in upper tail) at the top. The tα/2  value is in this column somewhere.
2. Go down the column to find degrees of freedom = n -1 on the left. The number in that row and column is the desired value of tα/2.

**Example:** 12 bank tellers were randomly sampled and it was determined they made an average of 3.6 errors per day with a sample standard deviation of 0.42 error. Construct a 90% confidence interval for the population mean of errors per day.

**Ans.** Here,

At 90% confidence interval, confidence coefficient(1-α) = 0.90

So, α/2 = 0.05, degree of freedom = n – 1 = 11 and 1.796

0.217

Interval estimate is given by given

**Determining sample size:** We can choose a sample size large enough to provide a desired margin of error (E).

If the computed sample size is not a whole number, then round the value up to the next whole number.

A planning value for the population standard deviation σ must be specified before the sample size can be determined. Three methods of obtaining a planning value for σ are discussed here

However, even if σ is unknown, we can use a preliminary or planning value for σ. In practice, one of the following procedures can be chosen.

1. Use the estimate of the population standard deviation computed from data of previous studies as the planning value for σ.

2. Use a pilot study to select a preliminary sample. The sample standard deviation from the preliminary sample can be used as the planning value for σ.

3. Use judgment or a “best guess” for the value of σ. For example, we might begin by estimating the largest and smallest data values in the population. The difference between the largest and smallest values provides an estimate of the range for the data. Finally, the range divided by 4 is often suggested as a rough approximation of the standard deviation and thus an acceptable planning value for σ.

**Example:** You want to estimate the mean number of friends for all users of the website. How many users must be included in the sample if you want to be 95% confident that the sample mean is within seven friends of the population mean? Assume the sample standard deviation is about 53.0.

**Ans.** E = 7, = 53

At 95% confidence interval, confidence coefficient(1-α) = 0.95

So, α/2 = 0.025 and

= 220.23

You should include at least 221 (rounding up to obtain a whole number) users in your sample

**Interval Estimate of proportion:** The interval estimate of population proportion is of the form

± margin of error

Interval Estimate is given by where (1 - α) is the confidence coefficient and zα/2 is the z value providing an area of α/2 in the upper tail of the standard normal probability distribution.

The sampling distribution of proportion can be approximated by normal distribution when

np 5 and n(1 – p) 5

**Example:** You’ve surveyed 1,000 individuals from your neighbourhood to determine how many of them would be willing to pay a higher Homes Owners Associated fee in order to build a neighbourhood pool.  612 said yes. Find the 95% confidence interval for the true population proportion for your neighbourhood.

**Ans.:** Here,

At 95% confidence interval, confidence coefficient(1-α) = 0.95

So, α/2 = 0.025 and

= 0.030

Interval estimate is given by given

**Determining the Sample Size:** Let us consider the question of how large the sample size should be to obtain an estimateof a population proportion at a specified level of precision. The margin of error(E) associated with an interval estimate of a population proportion is

If the computed sample size is not a whole number, then round the value up to the next whole number.

In practice, the planning value p\* can be chosen by one of the following procedures.

1. Use the sample proportion from a previous sample of the same or similar units.

2. Use a pilot study to select a preliminary sample. The sample proportion from this

sample can be used as the planning value, p\*.

3. Use judgment or a “best guess” for the value of p\*.

4. If none of the preceding alternatives apply, use a planning value of p\* = 0.50.

Some possible values of p\*(1- p\*)

|  |  |  |
| --- | --- | --- |
| p\* | p\*(1- p\*) |  |
| 0.10 | (0.10)(0.90) = 0.09 |  |
| 0.30 | (0.30)(0.70) = 0.21 |  |
| 0.40 | (0.40)(0.60) = 0.24 |  |
| 0.50 | (0.50)(0.50) = 0.25 | Largest value of p\*(1- p\*) |
| 0.60 | (0.60)(0.40) = 0.24 |  |
| 0.70 | (0.70)(0.30) = 0.21 |  |
| 0.90 | (0.90)(0.10) = 0.09 |  |

This value of p\* is frequently used when no other information is available. The sample size is proportional to the quantity p\*(1 - p\*). A larger value for the quantity p\*(1 - p\*) will result in a larger sample size. The largest value of p\*(1 - p\*) occurs when p\*= 0.50. In case of any uncertainty about an appropriate planning value, we know that p\* = 0.50 will provide the largest sample size recommendation. In effect, we play it safe by recommending the largest necessary sample size. If the sample proportion turns out to be different from the 0.50 planning value, the margin of error will be smaller than anticipated. Thus, in using p\* = 0.50, we guarantee that the sample size will be sufficient to obtain the desired margin of error.

**Example:** A newspaper article about the results of a poll states: "In theory, the results of such a poll, in 99 cases out of 100 should differ by no more than 5percentage points in either direction from what would have been obtained by interviewing all voters**.”** Find the sample size suggested by this statement.

**Ans:** Margin of error(E) = 0.05; confidence level = 99%

At 99% confidence interval, confidence coefficient(1-α) = 0.99

So, α/2 = 0.005 and

Taking planning proportion as p\* = 0.50

The sample size should be 664 (rounding off to the next whole number)