Mr.Gangadhar Immadi,

Immadi.gangadhar@gmail.com

9986789040

CONFIDANCE INTERVAL

INTRODUCTION TO CONFIDENCE INTERVAL

- Confidence interval is the range in which the value of a population parameter is likely to lie with certain probability

- Confidence interval provides additional information about the population parameter that will be useful in decision making

- The objective of confidence interval is to provide both location and precision of population parameters

Interval Estimate and Confidence Level

- An interval estimate of a population parameter such as mean and standard deviation is an interval or range of values within which the true parameter value is likely to lie with certain probability.

- Confidence level, usually written as $(1 - \alpha)100\%$, on the interval estimate of a population parameter is the probability that the interval estimate will contain the population parameter. When $\alpha = 0.05$, 95% is the confidence level and 0.95 is the probability that the interval estimate will have the population parameter

Significance and Confidence Level

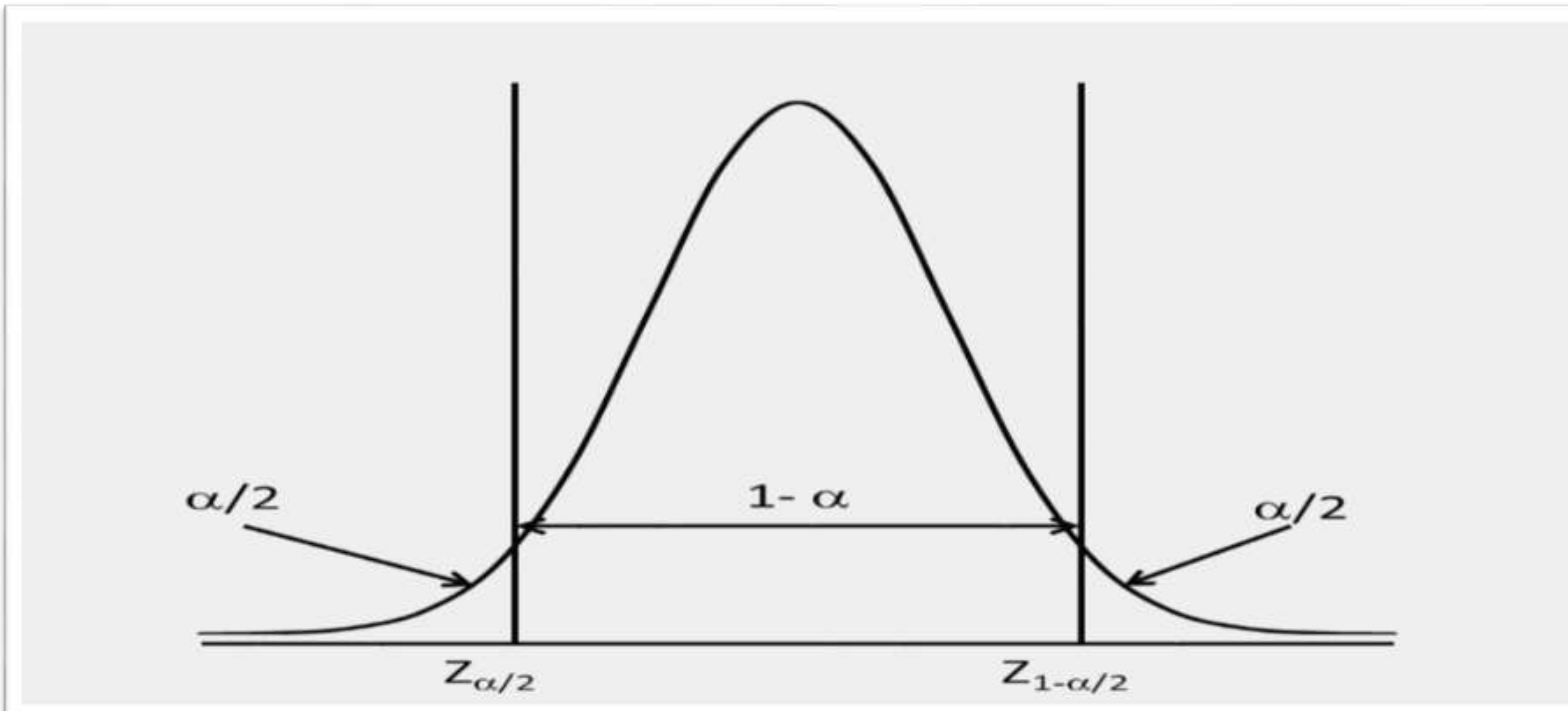
- The value of α is called significance and signifies that the chance of not observing the true population mean in the interval estimate is 1 out of 20. Alternatively, 95% confidence implies that in 19 out of 20 cases, the true population mean will be within the interval estimate.

- Confidence interval is the interval estimate of the population parameter estimated from a sample using a specified confidence level

Confidence Interval for Population Mean

- Let X_1, X_2, …, X_n be the sample means of samples S_1, S_2, …, S_n that are drawn from an independent and identically distributed population with mean μ and standard deviation σ. From central limit theorem we know that the sample means X_i follow a normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$
- The variable $Z = \dfrac{X_i - \mu}{\sigma / \sqrt{n}}$ follows a standard normal variable.

Assume that we are interested in finding $(1 - \alpha)$ 100% confidence interval for the population mean. We can distribute α (probability of not observing true population mean in the interval) equally ($\alpha/2$) on either side of the distribution as shown in Figure



CI for the population mean when population standard deviation is known

- In general, $(1 - \alpha)$ 100% the confidence interval for the population mean when population standard deviation is known can be written as

$$\bar{X} \pm Z_{\alpha/2} \times \sigma / \sqrt{n}$$

- Above equation is valid for large sample sizes, irrespective of the distribution of the population. The above equation is equivalent to

$$P(\bar{X} - Z_{\alpha/2} \times \sigma / \sqrt{n} \leq \mu \leq \bar{X} + Z_{\alpha/2} \times \sigma / \sqrt{n}) = 1 - \alpha$$

CI for Different Significance Values

- That is, the probability that the population mean takes a value between $\bar{X} - Z_{\alpha/2} \times \sigma/\sqrt{n}$ and $\bar{X} + Z_{\alpha/2} \times \sigma/\sqrt{n}$ is $1 - \alpha$.
- The absolute values of $Z_{\alpha/2}$ for various values of α are shown below:

α	$\lvert Z_{\alpha/2} \rvert$	Confidence interval for population mean when population standard deviation is known
0.1	1.64	$\bar{X} \pm 1.64 \times \sigma/\sqrt{n}$
0.05	1.96	$\bar{X} \pm 1.96 \times \sigma/\sqrt{n}$
0.02	2.33	$\bar{X} \pm 2.33 \times \sigma/\sqrt{n}$
0.01	2.58	$\bar{X} \pm 2.58 \times \sigma/\sqrt{n}$

Example

A sample of 100 patients was chosen to estimate the length of stay (LoS) at a hospital. The sample mean was 4.5 days and the population standard deviation was known to be 1.2 days.

(a) Calculate the 95% confidence interval for the population mean.

(b) What is the probability that the population mean is greater than 4.73 days?

Solution

(a) 95% confidence interval for population mean: We know that $\bar{x}$ 4.5 and σ = 1.2 and thus

The 95% confidence interval is given by $\sigma / \sqrt{n} = 1.2 / \sqrt{100} = 0.12$

$$\boxed{(\bar{X} - Z_{\alpha/2} \times \sigma / \sqrt{n}, \bar{X} + Z_{\alpha/2} \times \sigma / \sqrt{n}) = (4.5 - 1.96 \times 0.12,\ 4.5 + 1.96 \times 0.12) = (4.2648,\ 4.7352)}$$

For current problem CONFIDENCE(0.05, 1.2, 100) = 0.235196. The corresponding confidence interval is

(4.5 – 0.235196, 4.5 + 0.235196) = (4.2648, 4.7352)

Note that 4.73 is the upper limit of the 95% confidence interval from part (a), thus the probability that the population mean is greater than 4.73 is approximately 0.025.

Confidence Interval for Population Mean when Standard Deviation is Unknown

- if the population follows a normal distribution and the standard deviation is calculated from the sample, then the statistic given in Eq will follow a *t*-distribution with ($n-1$) degrees of freedom

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

- Here S is the standard deviation estimated from the sample (standard error). The *t*-distribution is very similar to standard normal distribution; it has a bell shape and its mean, median, and mode are equal to zero as in the case of standard normal distribution. The major difference between the *t*-distribution and the standard normal distribution is that *t*-distribution has broad tail compared to standard normal distribution. However, as the degrees of freedom increases the *t*-distribution converges to standard normal distribution.

Confidence Interval for Population Mean when Standard Deviation is Unknown

- The $(1 - \alpha)100\%$ confidence interval for mean from a population that follows normal distribution when the population mean is unknown is given by

$$\boxed{\overline{X} \mp t_{\alpha/2, n-1} \times \frac{S}{\sqrt{n}}}$$

- In above Eq, the value $t_{\alpha/2, n-1}$ is the value of t under t-distribution for which the cumulative probability $F(t) = 0.025$ when the degrees of freedom is $(n - 1)$.
- Here the degrees of freedom is $(n - 1)$ since standard deviation is estimated from the sample

- The absolute values of $t_{\alpha/2,n-1}$ for different values of α are shown in Table along with corresponding $Z_{\alpha/2}$ values.

α	$\lvert t_{\alpha/2,10}\rvert$	$\lvert t_{\alpha/2,50}\rvert$	$\lvert t_{\alpha/2,500}\rvert$	$\lvert Z_{\alpha/2}\rvert$
0.1	1.812	1.675	1.647	1.64
0.05	2.228	2.008	1.964	1.96
0.02	2.763	2.403	2.333	2.33
0.01	3.169	2.677	2.585	2.58

It is evident from table that the values of $t_{\alpha/2,n-1}$ and $Z_{\alpha/2}$ converge for higher degrees of freedom. In fact, as the sample size nears 100, the *t*-distribution gets very close to a normal distribution.

Example

- An online grocery store is interested in estimating the basket size (number of items ordered by the customer) of its customers so that it can optimize its size of crates used for delivering the grocery items. From a sample of 70 customers, the average basket size was estimated as 24 and the standard deviation estimated from the sample was 3.8. Calculate the 95% confidence interval for the basket size of the customer order.

Solution

We know that $\bar{X} = 24$, n = 70, S = 3.8 and $t_{0.025,\,69}$ = 1.995

$$\boxed{\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} = 24 \pm 1.995\frac{3.8}{\sqrt{70}} = 24 \pm 0.9061}$$

Thus the 95% confidence interval for the size of the basket is (23.09,24.91).

Confidence Interval for Population Variance

- Let $S_1^2, S_2^2, \ldots, S_k^2$ be the sample variance estimated from samples of size n drawn from a normal distribution with variance σ^2. Then the random variable defined by

$$\boxed{\frac{(n-1) \times S_i^2}{\sigma^2}}$$

follows a χ^2-distribution with $(n - 1)$ degrees of freedom

Note that Eq. is valid only when the samples are drawn from a normal population; it is not valid otherwise

Confidence Interval for Population Variance

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

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

- The confidence interval for standard deviation, σ, is given by

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

Example

- Time taken to manufacture an aircraft door is a random variable due to several manual processes and assembly of more than 1000 parts to make the aircraft door. The sources of variability in door assembly include factors such as non-availability of parts, manpower, and machine tools. It is known that the time to assemble a door follows a normal distribution. The variance of the time taken to manufacture the door was estimated to be 324 hours based on a sample of 50 doors. Calculate a 95% confidence interval for the variance in manufacturing aircraft door.

Solution

- We know that n = 50, S^2 = 324, $\chi^2_{0.025,49} = 70.22$ and $\chi^2_{0.975,49} = 31.55$

- The 95% confidence interval for variance is given by

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \quad \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right] = \left[\frac{49 \times 324}{70.22}, \frac{49 \times 324}{31.55}\right] = [226.09, \;\; 503.20]$$

- The 95% confidence interval for standard deviation is [15.04, 22.43].