**Interval Estimations**

* As population parameters are estimated based on the sample observations drawn from the population, with each estimate an error is associated; however good are the properties of the estimators.
* For example, if we use sample mean to estimate the mean of a population, it is unlikely that sample mean () will be exactly equal to .
* Hence it is often required to find out the interval taking into account the error of estimation within which the population parameter would lie with a very high probability, i.e. confidence.
* Confidence interval for a parameter is the interval constructed based upon a given sample observations such that the parameter is contained in the interval with a very high probability.
* The two end points are called confidence limits.
* The probability with which the interval contains the true value of the parameter is known as confidence coefficient.

For example, let and be two statistics of the random sample , collected from a specific population. Let and are such that and . Putting , then, we have,

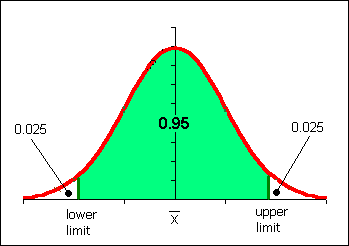
Then, is called the confidence coefficient,

is called the lower confidence limit for ,

is called the upper confidence limit for ,

is called the confidence interval for .

If the confidence interval is constructed by excluding equal tail areas, then what we get is the ***Central Interval***. If the distribution is symmetric, it can be shown that the ***central interval is the shortest interval*** with the same confidence coefficient.



**Central Interval (95 % Confidence)**

**Implication of 100(1-)% confidence interval of a parameter**

If a large number of times random samples of the same size are taken and each time confidence intervals is determined, then ***100(1-)% of these intervals will contain the parameter*** .

**Note:**

* Pivotal statistic/function is fundamental to the construction of confidence interval, and
* The precision of the confidence interval depends on the sample size.

**Pivotal statistic:** Pivotal statistic is a function of observations and unobservable parameters such that the function's probability distribution does not depend on the unknown parameters. In other words, a statistic is said to be pivotal if its sampling distribution does not depend on unknown parameter(s).

It is important to note that to facilitate construction of the confidence interval of a parameter, the pivotal statistic should involve the observations in the form of the estimator (*T*) or some statistic involving the estimator and the parameter (**)**, and the sampling distribution of the pivotal statistic must be known. For examples:

|  |  |
| --- | --- |
|  |  |

***The desired properties of the pivotal statistic are as follows:***

1. It depends on only through a sufficient statistic,
2. It involves the parameter on which confidence interval is wanted and no other, and
3. It has a fixed distribution independent of the parameter.

**Procedure for construction of confidence interval for a parameter**

* Let there exists a statistic and a pivotal function of *T* and , which is defined for all and the distribution of is independent of (i.e. same for all ).
* Then we can find from the sampling distribution of , two constants and , depending on and but not on , such that
  + - , and
* This implies that.
* Suppose further that it is possible to re-write the inequality in the form , where and are independent of .
* Then,

Hence, for any given set of observations , the values and of and respectively, are a pair of confidence limits to with confidence coefficient .

* **Let , where is known. Determine the confidence interval of based on a random sample of size .**

Here, the pivotal function is . Note that is a statistic which involves the parameter as well as the estimator and the sampling distribution of the statistic is independent of the parameter .

From standard normal table we can find out the critical value and such that

, and

This implies that

Thus, 100(1-)% confidence interval of is

* **, where is unknown. Determine the confidence interval of based on a random sample of size .**

Here the pivotal function is

From the statistical table for *t* distribution, we can find out the critical values and such that

This implies that

Thus, 100(1-)% confidence interval of is

* **and where populations of and are independent, and variances and are known. Determine the confidence interval for based on random samples of sizes and taken from the two populations.**

Here, the pivotal function is

From standard normal table we can find out the critical value and such that

Thus, 100(1-)% confidence interval of is

* **and where populations of and are independent variances, and and are unknown. Determine the confidence interval for based on random samples of sizes and ( and are small) taken from the two populations. It is reasonable to assume that .**

Let the sample means for the two populations be denoted by and respectively, and the sample variances by and respectively. Since both and are estimators of the common variance , we may obtain a combined (or pooled) estimator of that is better than or individually as

Then, the pivotal function is

From statistical table for t-distribution we can find out the critical value and such that

Thus, 100(1-)% confidence interval of is

* ***Confidence interval of for paired observations***

Consider the following cases:

* A chemical engineers wish to compare the fuel economy obtained by two different formulations of gasoline. Here, it would make more sense to select pairs of cars of the same make and model and driven under similar circumstances, and compare the fuel economy of the two cars in each pair.
* A dietician wishes to understand the effect of a diet for six months on weight of some selected boys. Here, it is necessary to measure the weight each boy before they are subjected to the change of diet and after a lapse of six months.
* Similarly, in many experimental situations, there are only n different experimental units and the data are collected in pairs; that is two observations are made on each unit.

In all such cases, it would be incorrect to analyze the data using the previous formulas for differences of mean of two normal distributions, since the samples were not drawn independently. What is correct is to compute the difference in the values in each pair (subtracting in the same order each time) and treat the differences as the data.

The new sample of differences may be considered as a random sample of size  selected from a normal population with mean and standard deviation . This approach essentially transforms the paired two-sample problem into a one-sample problem.

Suppose the data consists of pairs (, (, (,..., (. The random variables and have means and respectively. Let represents the difference between the random variables in the pair, that is, . Then, can be estimated as = and can be estimated as ,

The pivotal statistic is

Thus, 100(1-)% confidence interval of is

* **, is unknown. Determine the confidence interval of based on a random sample of size .**

Here, the pivotal statistic , where

From the statistical table of distribution, we can find out the critical values and such that

and

This implies that

Thus, confidence interval of is

* **, is known. Determine the confidence interval of based on a random sample of size *n*.**

Here, the pivotal statistic , where

From the statistical table of distribution, we can find out the critical values and such that

and

This implies that

Thus, confidence interval of is

* **and where populations of and are independent and variances and are unknown. Determine the confidence interval for the ratio**  **based on random samples of sizes and taken from the two populations.**

Here the pivotal statistic is

From the statistical table of distribution, we can find out the critical values and such that

and

This implies that

Or,

Or,

Or,

Thus, confidence interval of is

or

**Confidence Interval for population proportion ( is large)**

* Sample proportion = , where , if there is a success, , otherwise
* We know that and
* We further know that , if is large and
* Then, confidence interval (CI) of can be obtained considering as the pivotal statistic, i.e. for large ,

From standard normal table we can find out the critical value and such that

, and

This implies that

Thus, 100(1-)% confidence interval of is

Unfortunately, the upper and lower limits of confidence interval obtained above contain the unknown parameter . However, a satisfactory solution is to replace by in the standard error, which leads to the approximate 100(1-)% confidence interval of as

**Confidence Interval for population proportion ( is not large)**

If the value of *n* is not large, we need to develop a test for every . Such a test can be constructed as follows:

* For any we can always find approximate values of and such that and or are as close as possible to , where is the number of success and is the probability of success.
* Then,
* The above inequality can, equivalently, be written as

* So, confidence interval can be defined as

**Construction of CI when pivotal function is not available**

In the general case, if a function is not available, then we may proceed as follows:

* Upon getting the value of a statistic from a random sample, we may try to find two values of the parameter such that the observed value of becomes the and points of the sampling distribution of , respectively.
* These two values of may be considered as the confidence limits of .

**Summary of Confidence Interval of different Parameters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **For the parameter** | **Pivotal Statistic** | **Two-sided Confidence Interval** |
| **1** | in   * known |  |  |
| **2** | in   * unknown and * sample size small ( |  |  |
| **3** | in   * unknown and * sample size small ( |  |  |
| **4** | in independent &   * and are known |  |  |
| **5** | in independent &   * and are unknown * , are large |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **6** | in independent and   * unknown * , are small | where, |  |
| **7** | in independent and   * unknown * , are large |  |  |
| **8** | in independent &   * and are unknown * , are not large | where, |  |
| **9** | for paired observations. |  |  |
| **10** | in   * known |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| 11 | in   * unknown |  |  |
| 12 | in   * unknown and large |  |  |
| 13 | inindependent and |  | Or, |
| 14 | in   * and |  |  |
| 1 | and in independent and   * and are large |  |  |

**Notes**

* If the sample size is large, it can be shown that for samples from the distribution that is approximately normal, the approximate sampling distribution of is given by

* Let and , and both and are unknown. Further, the sample sizes and are not large. Then,

(approximately), ,

**Choice of Sample Size**

* The length of a confidence interval is a measure of the precision of estimation. For example, the precision of the confidence interval of (when is known) is .
* This means that while using as an estimate of , the error is less or equal to with a confidence of .
* In situations where we have the freedom to determine the sample size, we can choose so that we are confident that the error in estimating is less than or equal to a specified bound on the error *E.*
* So, the appropriate sample size is found by choosing such that .
* Therefore, solving this we get the sample size as

If value of is not an integer, it must be rounded up. This will ensure that level of confidence does not fall below .

**Numerical problems: Examples**

**Problem 1:** In a population, a random variable follows a normal distribution with an unknown mean and a standard deviation of 2.

1. In a sample of 400 selected at random, a sample mean of 50 was obtained. Determine the confidence interval with a confidence level of 97% for the average population. Given that .
2. With the same confidence level, what minimum sample size should it have so that the interval width has a maximum length of 1?

**Ans:**

1. Given that , , , i.e.

We know that .

Therefore, 97% confidence interval of average population () is

****

1. The interval width has a maximum length of 1. Therefore, maximum error, . So, we must have

, or , or

So, sample size for the 97% confidence level will be at least 76.

**Problem 2:** The quantity of hemoglobin in the blood stream of a man follows a normal distribution with a standard deviation of 2 g/dl. Calculate the confidence level for a sample of 12 men which indicates that the population mean blood hemoglobin is between 13 and 15g/dl.

**Ans**



Given that confidence interval is (13, 15) g/dl., *n* = 12, and g/dl. So,

Or,

Or,

Or,

From, standard normal table we get, .

So,

Therefore, the required confidence level is 1 - (0.0418×2) = 0.9164 = 91.64%.

**Problem 3:** A meteorologist who sampled 13 thunderstorms found that the average speed at which they travelled across a certain state is 15 miles per hour. Standard deviation of the sample was 1.7 miles per hour. Find the 99% confidence interval of the mean.

**Ans:** Given that , ,, i.e.

We know that

Therefore, from the statistical table for *t* distribution, we can find out two critical values and such that

Therefore, 99% confidence interval of average population () is

**Exercises**

**Exercise 1:** An optical firm purchases glasses for fitting into lenses. The refractive index (RI) of 20 pieces of glass, randomly selected from a large shipment, has a variance of . Assuming the distribution of RI to be normal construct a 95% confidence interval of .

**Exercise 2:** It was observed that 40 one-gallon cans of a certain kind of paint had covered, on the average, 513.3 sq. ft. with a standard deviation of 31.5 sq. ft. Assuming area covered by a one-gallon can is approximately normal, find the 99% confidence interval of .

**Exercise 3:** A random sample of 100 farms in a certain year gave an average yield of barley of 2000 lbs per acre with SD of 192 lbs, whereas a random sample of 100 farms in the same year showed average yield of wheat of 2100 lbs with SD of 224 lbs. Construct a 95% confidence interval for mean difference of yields of barley and wheat.

**Exercise 4:** A random sample of size 25 gives a sample mean of 20. The sample is collected from a population. Construct a 95% confidence interval for .

**Exercise 5:** 10 bullets from an enemy gun show an average diameter of 5 micron with a SD of 0.02 micron. Obtain a 99% confidence interval for the diameter of the gun barrel taking the diameter of the gun barrel to be 0.01 micron more than that of the bullets.

**Exercise 6:** A casting is supplied by two vendors A and B. From the supplies of the vendors, random samples of size 10 and 12 respectively, were examined for weight. Results obtained were as under.

Vendor A Vendor B

Average 1.1 kg 1.2 kg

SD 0.013 kg 0.017 kg

Construct a 95% confidence interval expected difference in weight under the assumption of normal population equal variance.

**Exercise 7:** A random sample of 100 bolts collected from lot of 1000 bolts shows that 10 out of them are defective. Obtain a 99% interval estimate for the expected number of non-defective bolts.

**Exercise 8:** A random sample of 40 microscopes from a shipment shows that 5 of them do not meet the specification. Construct a 99% confidence interval for the expected number of defectives in a shipment of 1000 such microscopes.

**Exercise 9:** A random sample of 12 piston rings showed an average width of 90 mm with a SD of 0.2 mm. Assuming a normal distribution for width, construct a 99% confidence interval for, where is the true SD.

**Exercise 10**: For a class project, a political science student at a large university wants to estimate the percent of students who are registered voters. He surveys 500 students and finds that 300 are registered voters. Compute a 90% confidence interval for the true percent of students who are registered voters, and interpret the confidence interval.

**Exercise 11**: The weights (in pound) at birth for 15 babies born in a hospital are obtained as follows: 6.2, 6.7, 7.1, 6.9, 7.5, 5.7, 4.8, 6.8, 7.6, 7.8, 8.1, 5.0, 5.8, 7.9 and 8.5. Give two limits between which the mean weight at birth for all such babies is likely to lie. Consider the confidence coefficient is 0.99.

**Exercise 12**: The weight (in pound) of 10 boys before they are subjected to a change of diet and after a lapse of six months are recorded as follows:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Before | 109 | 112 | 98 | 114 | 102 | 97 | 88 | 101 | 89 | 91 |
| After | 115 | 120 | 99 | 117 | 105 | 98 | 91 | 99 | 93 | 89 |

Give two limits between which the mean differences of weights is likely to lie if the diet has no effect. Consider the confidence coefficient is 0.99.