Frequently Asked Questions

1. What do you mean by Interval Estimation?

Answer:

A confidence interval gives an estimated range of values, which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

Hence an interval estimator is a rule for calculating two numbers say A and B to create an interval that we are fairly certain contains the parameter of interest that is population mean μ.

In statistics, a confidence interval (CI) is a kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate.

2. Why the statistical procedures are important?

Answer:

Statistical procedures are important because they provide the two types of information

- Methods of making the inference
- A numerical measure of the goodness or reliability of the inference

3. Why interval estimation is required as compared to point estimation?

Answer:

In a decision making process the parameters of interest can be estimated either using Point Estimation or Interval estimation techniques. If we have to estimate the population mean either we go for method of moments or method of maximum likelihood in case of point estimation which gives a single number as an estimate of the population mean. But in general however good an estimator may be it cannot be expected to coincide with the actual value of the parameter. In fact, the estimator in general has a continuous distribution and the probability that that it is equal to a particular value is zero. Because of variation in sample statistics, estimating a population parameter with a confidence interval is often preferable to using a point estimate.

In statistics, a **confidence interval** (**CI**) is a kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. We measure the probability using the confidence coefficient, designated by (1-α). A confidence interval can be used to describe how reliable survey results are but we cannot estimate the reliability of the estimates in case of point estimation.

4. What are the elements of interval estimation?

Answer:

The main elements of interval estimation are

- Level of confidence
    - Confidence in which the interval will contain the unknown population parameter
- Precision (range)
    - Closeness to the unknown parameter
- Cost
    - Cost required to obtain a sample of size n

5. Write a note on Confidence limits and confidence coefficient (level).

Answer:

Confidence Limits

Confidence limits are the lower and upper boundaries / values of a confidence interval, that is, the values which define the range of a confidence interval.

The upper and lower bounds of a 95% confidence interval are the 95% confidence limits. These limits may be taken for other confidence levels, for example, 90%, 99%, 99.9%.

Confidence Level

The confidence level is the probability value $(1 - \alpha)$ associated with a confidence interval.

It is often expressed as a percentage. For example, say $\alpha = 0.05 = 5\%$, then the confidence level is equal to (1-0.05) = 0.95, i.e. a 95% confidence level.

If the experiment is repeated, how frequently the observed interval contains the parameter is determined by the **confidence level** or **confidence coefficient**. The amount of confidence (in terms of probability) with which a parameter is expected to lie in the confidence interval is known as confidence coefficient.

6. What do you mean by construction of 95% confidence Interval?

Answer:

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 98%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

In an interval estimation, if we say that we obtained 95% confidence interval it means that if a large number of samples is taken and if the Confidence interval for each of the sample is obtained then in 95% of the cases the parameters value can be expected to lie in the confidence interval obtained.

7. Derive an interval estimate for the mean of the population when the population variance is known.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the SRS random samples drawn from a population of size N with mean $\bar{Y}$ and variance σ^2 we know that

We know $\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

$Z = (\bar{y} - \bar{Y}) / S.E(\bar{y}) \sim N(0,1)$

We can always find two quantities $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ from Standard Normal variate tables such as

$P[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}] = 1 - \alpha$

$P[-Z_{\alpha/2} \leq (\bar{y} - \bar{Y}) / S.E(\bar{y}) \leq Z_{\alpha/2}] = 1 - \alpha$

$P[-Z_{\alpha/2} S.E(\bar{y}) \leq (\bar{y} - \bar{Y}) \leq Z_{\alpha/2} S.E(\bar{y})] = 1 - \alpha$

$P[\bar{y} - Z_{\alpha/2} S.E(\bar{y}) \leq \bar{Y} \leq \bar{y} + Z_{\alpha/2} S.E(\bar{y})] = 1 - \alpha$

Therefore 100 (1- α) % C.I for the population mean $\bar{Y}$ when the variance is known as σ^2 is given by

$[\bar{y} - Z_{\alpha/2} S.E(\bar{y}), \bar{y} + Z_{\alpha/2} S.E(\bar{y})]$

Under SRSWR $S.E.(\bar{y}) = \sqrt{V(\bar{y})} = \sqrt{\sigma^2/n}$

95% Confidence Interval for the population mean $\bar{Y}$ is

$[\bar{y} - 1.96 \sqrt{\sigma^2/n}, \bar{y} + 1.96 \sqrt{\sigma^2/n}]$

8. Derive an interval estimate for the unknown mean of the population when the variance is unknown .

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the random samples drawn from a population of size N with mean $\bar{Y}$
We know that

We know $\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

$t = (\bar{y} - \bar{Y}) / S.E(\bar{y}) \sim t_\alpha(n-1)$

We can always find two quantities – $t_\alpha(n-1)$ and $t_\alpha(n-1)$ from Students t distribution table such that

$P[- t_\alpha(n-1) \le t \le t_\alpha(n-1)] = 1- \alpha$

$P[\bar{y} - t_\alpha(n-1)\ S.E(\bar{y}) \le \bar{Y} \le \bar{y} + t_\alpha(n-1)\ S.E(\bar{y})] = 1- \alpha$

Therefore, 100 (1- α) % C.I for the population mean $\bar{Y}$ when the variance is unknown is given by

$[\bar{y} - t_\alpha(n-1)\, S.E(\bar{y}) , \bar{y} + t_\alpha(n-1)\, S.E(\bar{y})]$

Therefore, 100 (1- α) % C.I for the population mean $\bar{Y}$ when the variance is unknown under SRSWR is given by

$[\bar{y} - t_\alpha(n-1) \sqrt{s^2/n} , \bar{y} + t_\alpha(n-1) \sqrt{s^2/n}]$

Where

$t_\alpha(n-1)$ **is the *t* distribution value with *(n -1)* degrees of freedom.**

9. A Simple Random Sample of size 25 was drawn from a population which has the mean value of 50 and its s value is 8. Set up 95% Interval estimate for the population mean μ.

Answer:

Therefore, 100 (1- α) % C.I for the population mean $\bar{Y}$ when the variance is unknown under SRSWR is given by

$[\bar{y} - t_\alpha(n-1) \sqrt{s^2/n} , \bar{y} + t_\alpha(n-1) \sqrt{s^2/n}]$

From the table of students t distribution t(24) = 2.0639 at 5% level of significance. Hence the interval is

$[50 - 2.0639 \sqrt{64/25} , \quad 50 + 2.0639 \sqrt{64/25}]$

$[46.69, 53.30]$

Hence, an Interval estimate for the population mean μ is $[46.69, 53.30]$

10. Give examples to state the application of interval estimation of population mean.

Answer:

Practical problems very often lead to estimation of μ, the mean of the population. A confidence interval for a mean specifies a range of values within which the unknown population parameter, in this case the mean, may lie. These intervals may be calculated by, for example

1) The average achievement of college students

2) The average number of deaths per age category

3) The average demand for a new cosmetic product

When the sample size n is large, the sample mean $\bar{y}$ is the best point estimator for the population mean μ. Since its sampling distribution is approximately normal, it can be used to construct a confidence interval according to the general approach

11. What are the steps to be followed for the construction of the C.I for the unknown population mean with known variance of the population?

Answer:

Steps to obtain confidence interval for mean when the variance is known

- First, obtain the point estimate of μ, that is, the sample mean .
- If *n* is large, then the Central Limit Theorem can be used and y bar is normally distributed with mean μ and standard deviation sigma by root n
- Select a confidence level. The most common level is 95%.
- Obtain the margin of error associated with the confidence level. For a normal distribution, the interval from *Z* = -1.96 to *Z* = 1.96 contains 95% of the area under the curve or of the sample means.
- Then the confidence interval is[y bar – 1-96 . S.E(y bar), y bar + 1-96 . S.E(y bar)]

12. State the procedure to be followed for the construction of the C.I for the unknown population mean with an unknown variance of the population?

Answer:

Steps to obtain confidence interval for mean when the variance is unknown

- First, obtain the point estimate of μ, that is, the sample mean .
- Select a confidence level. The most common level is 95%.
- When the population standard deviation sigma is unknown and n is small say (less than 30) then the confidence interval is
  [y bar – t alpha (n - 1) . S.E(y bar), y bar + t alpha (n- 1) . S.E(y bar)]

In fact, when n is greater than 30 the t value for 95% confidence level is 1.96, same as Z value.

13. How do you determine the standard deviation σ of the population?

Answer:

- In order to construct an interval estimate, it is necessary to obtain some estimate of σ, the variability of the population from which the sample is drawn. This is required to obtain an estimate of the standard error of the sample mean
- Generally, the sample standard deviation *s* is used as an estimate of σ. For large sample size, assume the CLT holds and assume *s* provides a reasonable estimate of σ. For a small sample, where $n < 30$, the *t*-distribution should be used, again using *s* as an estimate of σ.
- In practice, σ is rarely known. In addition, as *n* increases, the *t*-distribution approaches the normal distribution. Thus, so long as $n > 30$, it is acceptable to use *s* as an estimate of σ for purposes of constructing an interval estimate.
- Certain factors may affect the confidence interval size including size of sample, level of confidence, and population variability. A larger sample size normally will lead to a better estimate of the population parameter.

14. A social worker is faced with a population of 1000 families and wants to estimate the average expenditure of the families of the population. A sample of 50 families is selected with average expenditure of $1076.39, standard deviation of $273.62. Set up the 95% confidence interval estimate of the average expenditure of the families in the population.

Answer:

Given, n=50, $\bar{y} = \$1076.39$ and s =$273.62

95% Confidence Interval for the population mean μ is

$$[\ \bar{y} - t_{\alpha}(n\text{-}1)\sqrt{s^2/n}\ ,\ \bar{y} + t_{\alpha}(n\text{-}1)\sqrt{s^2/n}\ ]$$

1 -α = 0.95 which implies α=0.05 and since n is greater than 30

$t_{\alpha}(n-1) = 1.96$**.**

By substituting the given values in the above we get

[1076.39–(1.96) 273.62/√50 , 1076.39+(1.96) 273.62/√50]

=[1026.39 , 1126.39]

The 95% confidence interval for the population average expenditure of the families is between $1026.39 and $1126.39

15. A random sample of 25 students is selected from a University to estimate an average time spent by a student in a library in a month. A sample data reveals that a sample

average = 50 hours and a standard deviation is 8 Set up 95% Confidence Interval estimate for the average time spent by the students of University in a library.

Answer:

Given, n=25, $\bar{y} = 50$ and s =8

95% Confidence Interval for the population mean μ is

$$\left[\, \bar{y} - t_{\alpha}(n-1)\sqrt{s^2/n}\ ,\ \bar{y} + t_{\alpha}(n-1)\sqrt{s^2/n}\,\right]$$

1 -α = 0.95 which implies α=0.05 and $t_{\alpha}(n-1)$ = $t_{0,05}(24)$=2.06 from the table of students t distribution

By substituting the given values we get

[50–(2.06) 8/√25 , 50+(2.06) 8/√25]

=[46.704 ,53.296]

The 95% confidence interval for the average time spent by the students of a university in a library is between 46.70 hours and 53.30 hours.