Frequently Asked Questions

1. What do you mean by Interval Estimation?

A**nswer:**

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.

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. It is an observed interval (i.e. it is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest. Hence interval estimation is a process of obtaining an interval in which the parameter value is expected to lie.

2. Explain the desired properties of confidence intervals.

Answer:

The desirable properties may be described as: validity, optimality and invariance. Of these "validity" is most important, followed closely by "optimality". "Invariance" may be considered as a property of the method of derivation of a confidence interval rather than of the rule for constructing the interval. In non-standard applications, the same desirable properties would be sought.

- *Validity.* This means that the nominal coverage probability (confidence level) of the confidence interval should hold, either exactly or to a good approximation.

- *Optimality.* This means that the rule for constructing the confidence interval should make as much use of the information in the data-set as possible. Recall that one could throw away half of a dataset and still be able to derive a valid confidence interval. One way of assessing optimality is by the length of the interval, so that a rule for constructing a confidence interval is judged better than another if it leads to intervals whose lengths are typically shorter.

- *Invariance.* In many applications the quantity being estimated might not be tightly defined as such. For example, a survey might result in an estimate of the median income in a population, but it might equally be considered as providing an estimate of the logarithm of the median income, given that this is a common scale for presenting graphical results. It would be desirable that the method used for constructing a confidence interval for the median income would give equivalent results when applied to constructing a confidence interval for the logarithm of the median income: specifically the values at the ends of the latter interval would be the logarithms of the values at the ends of former interval.

3. 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

4. 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**.

More specifically, the meaning of the term "confidence level" is that, if confidence intervals are constructed across many separate data analyses of repeated (and possibly different) experiments, the proportion of such intervals that contain the true value of the parameter will approximately match the confidence level; this is guaranteed by the reasoning underlying the construction of confidence intervals

5. Derive an interval estimate for the mean of the population when the population variance is known under SRSWR.

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 ($\bar{y}$) ≤ $\bar{Y} \leq \bar{y}$ + Z $_{\alpha/2}$ S.E ($\bar{y}$)] = 1- α

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 $\sqrt{\sigma^2/n}$]

6. Derive an interval estimate for the unknown mean of the population when the variance is unknown under SRSWR.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the SRSWOR 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 ($\bar{y}$) ~ $t_\alpha(n-1)$

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

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

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

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 ($\bar{y}$)]

Under SRSWR when the variance is unknown the S.E is given by

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.**

7. Obtain an interval estimate for the mean of the population when the variance is known under SRSWOR.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the SRSWOR 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 SRSWOR $S.E.(\bar{y}) = \sqrt{V(\bar{y})} = \sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}}$

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

$$\left[\bar{y} - 1.96\sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}} \;, \bar{y} + 1.96\sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}}\;\right]$$

8. Deduce an interval estimate for the mean of the population when the variance is unknown under SRSWOR.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the SRSWOR 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

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

Under SRSWOR when the variance is unknown $S.E.(\bar{y}) = \sqrt{V(\bar{y})} = \sqrt{\frac{N-n}{N} * \frac{s^2}{n}}$

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

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

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 using With replacement technique 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. Derive an interval estimate for the population total when the variance is known under SRSWR.

Answer:

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

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

$$Z = \frac{\hat{Y} - Y}{S.E(\hat{Y})} \sim N(0,1)$$

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

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

$$P\left[-Z_{\alpha/2} \leq \frac{\hat{Y} - Y}{S.E(\hat{Y})} \leq Z_{\alpha/2}\right] = 1-\alpha$$

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

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

Therefore 100 (1- α) % C.I for the population Total

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

Under SRSWR $\quad S.E.(\hat{Y}) = \sqrt{V(\hat{Y})} = \sqrt{N^2\sigma^2/n}$

An estimate of the population total is $\hat{Y} = N\bar{y}$

95% Confidence Interval for the population total is

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

11. Derive an interval estimate for the population total when the standard deviation is known under SRSWOR.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the SRS without replacement samples drawn from a population of size N with mean $\bar{Y}$ and variance σ^2 we know that

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

$$Z = \frac{\hat{Y} - Y}{S.E(\hat{Y})} \sim N(0,1)$$

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

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

$$P\left[-Z_{\alpha/2} \le \frac{\hat{Y} - Y}{S.E(\hat{Y})} \le Z_{\alpha/2}\right] = 1 - \alpha$$

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

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

Therefore 100 (1- α) % C.I for the population Total

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

Under SRSWOR $$S.E.(\hat{Y}) = \sqrt{N^2 \times \frac{(N-n)}{N-1} \times \frac{\sigma^2}{n}}$$

An estimate of the population total is $\hat{Y} = N\bar{y}$

95% Confidence Interval for the population total is

$$\left[N\bar{y} - 1.96\sqrt{N^2 \times \frac{(N-n)}{N-1} \times \frac{\sigma^2}{n}},\ N\bar{y} + 1.96\sqrt{N^2 \times \frac{(N-n)}{N-1} \times \frac{\sigma^2}{n}} \right]$$

12. What do you mean by determination of sample size?

Answer:

Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample. In practice, the sample size used in a study is determined based on the expense of data collection, and the need to have sufficient statistical power. In complicated studies there may be several different sample sizes involved in the study. Sample sizes may be chosen in several different ways:

In order to estimate the population parameters with a specified degree of precision one has to take a decision about the size of the sample. Suppose it is desired to estimate, with a confidence interval, the mean of a population (μ). One of the first question arise is how large the sample should be? This question must be given serious consideration, because it is a waste of time and resources to take a larger sample than is needed to achieve the desired results. Similarly, a sample that is too small may lead to results that are of no practical value.

13. Derive an expression for the sample size of the Simple random sample under SRSWOR for estimating the population mean when the population variance is known .

Answer:

If $\bar{Y}$ denote the mean of the population observations of size N and let $\bar{y}$ denote the mean of the sample observations of size n. We know that $\bar{y}$ is an unbiased estimate of population mean $\bar{Y}$. If the permissible error in estimating $\bar{Y}$ is "e" and the confidence coefficient is 1 – α, then the sample size n is determined by the equation

$P\,[|\ \bar{y} - \bar{Y}\ | \leq e] = 1 - \alpha$ ------- (1) which implies

$P\,[|\ \bar{y} - \bar{Y}\ | > e] = \alpha$ - ------(2)

Where α is the level of significance. If n is sufficiently large and we consider SRSWOR then the statistic

$$Z = \frac{\bar{y} - \bar{Y}}{S\sqrt{\frac{N-n}{Nn}}}$$ is a Standard Normal variate

Accordingly if we take α=0.05 then we have

$$P\left[\frac{|\bar{y} - \bar{Y}|}{S\sqrt{\frac{N-n}{Nn}}} > 1.96\right] = 0.05$$

Comparing this with equation (2) we get

$$e = 1.96\, S\sqrt{\frac{N-n}{Nn}}$$

$$n = \frac{N\, S^2 (1.96)^2}{Ne^2 + (1.96)^2\, S^2}$$

The formula gives the sample size in SRSWOR for estimating $\bar{Y}$.with the confidence coefficient 95% and the margin of error 'e', provided n is large.

14. Derive an expression for the sample size of the Simple random sample under SRSWOR for estimating the population mean when the population variance is unknown .

Answer:

When the population variance is unknown and the required sample size is small we make use of Student's t distribution with (n-1) degrees of freedom. If α is the level of significance and t_α is the significant value for

(n-1) d.f then the sample size 'n 'is given by the equation

$$P\left[\left|\bar{y} - \bar{Y}\right| \geq t_\alpha \sqrt{\frac{N-n}{N}\frac{s^2}{n}}\right] = \alpha$$

Whici implies e= $t_\alpha \sqrt{\frac{N-n}{N}\frac{s^2}{n}}$

Which implies $n = \frac{N\, t^2 s^2}{Ne^2 + t^2\, s^2}$ $\Longrightarrow$ $n = \frac{s^2 t^2}{e^2 + t^2\, s^2/N}$

15. What sample size is needed to estimate the population mean, of being correct within ±5 from a population of size 100 and the population variance is given by 1932.657 with a 95% CI?

Answer:

$$n = \frac{N\,(1.96)^2 s^2}{Ne^2 + (1.96)^2\, s^2} =$$

$$n = \frac{100(1.96)^2(1932.657)}{100(5^2) + (1.96)^2(1932.657)} = 74.80$$

≈ 75