Frequently Asked Questions

1. What do you mean by Interval Estimation?

Answer:

Interval estimation is a process of obtaining an interval in which a population parameter value is expected to lie with a certain level of confidence. Interval estimation results in a confidence interval, an estimated range of values that 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. Briefly explain the principle behind the interval estimation.

Answer:

Let us consider a population with probability density function or probability mass function f(x, θ) where the parameter θ may be real valued or vector valued which is unknown. In interval estimation our problem is to find a Confidence interval for the parameter θ. In an interval estimation problem of finding confidence interval for the parameter θ with certain amount of confidence (1-α), we need to find two quantities A and B based on the sample observations such that

$P[A < \theta < B] = 1-\alpha$ -----(1)

In general, there may be many sets of values for A and B, which satisfy the equation (1). However, usually A and B are found such that they are at equal distance from the true value of the parameter θ. Intervals so obtained are called central intervals and the others non-central. If A and B are numbers such that $P[\theta<A]=1-\alpha$ and

$P[\theta>B]=1-\beta$ then A and B are called one sided upper confidence limit for θ with confidence coefficient (1- α) and one sided lower confidence limit with confidence coefficient (1- β)

3. What do you mean by central confidence Intervals?

Answer:

Let us consider a population with probability density function or probability mass function f(x, θ), where θ is the parameter. Let x1, x2... xn be a sample of size n taken from the above population. Let A and B be two statistics.

If $P[A > \theta]=\alpha_1/2$ and $P[B < \theta]= \alpha_2/2$ such that $P[A < \theta < B] = 1-\alpha$ where $\alpha = \alpha_1/2+ \alpha_2/2$, confidence intervals of this type are called central confidence intervals

4. 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.

5. 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 obtaining a sample of size n

6. 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 that 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

7. What do you mean by one sided upper confidence limit one sided lower confidence limit in interval estimation?

Answer:

In Interval estimation we try to find a Confidence interval for the unknown parameter θ. In the process of obtaining CI for the parameter θ with certain amount of confidence ($1-\alpha$), we need to find two quantities A and B based on the sample observations such that

$P[A < \theta < B] = 1-\alpha$ -----(1)

If A and B are numbers such that $P[\theta < A] = 1-\alpha$ and

$P[\theta > B] = 1-\beta$ then A and B are called one sided upper confidence limit for θ with confidence coefficient ($1-\alpha$) and one sided lower confidence limit with confidence coefficient ($1-\beta$)

8. What do we understand by the statement that we have obtained 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.

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.

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.

9. Explain the steps to be followed in constructing interval estimates of a parameter?

Answer:

The general form for the interval estimate of a population parameter is

Point estimate of parameter ± Margin of error

- The **margin of error** is an amount that is added to and subtracted from the point estimate of a statistic, to produce an interval estimate of the parameter
- The size of the margin of error depends on
  - The type of sampling distribution for the sample statistic.
  - The percentage of the area under the sampling distribution that a researcher decides to include – usually 90%, 95%, or 99%. This is termed a **confidence level**
- Each interval estimate is an interval constructed around the point estimate, along with a confidence level.

10. Explain interval estimation technique with an example.

Answer:

Example for interval estimation:

- Statistics Canada reports that mean weekly food expenditures for Prairie households in 2001 were $127.78. However, the data were obtained from a sample so there is sampling error associated with this estimate. The "true" value of the mean is between $123.78 and $131.78, 68% of the time and between $119.78 and 135.78, 95% of the time.
- "The margin of error is estimated to be plus or minus 3.51 per cent, 19 times out of 20." From the Palliser electoral district poll reporting Conservative at 43.3%, NDP at 35.7%, Liberal at 17.3%, and Green at 3.5% of decided voters, conducted by Sigma Analytics.

11. What do you mean by confidence interval for mean?

Answer:

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, a producer who wishes to estimate his mean daily output; a medical researcher who wishes to estimate the mean response by patients to a new drug; etc. Hence when we are estimated in the average characteristics of the population we go for confidence interval for mean which gives us an interval in which a population average is expected to lie with a certain level of confidence.

12. Briefly explain the steps to obtain confidence interval for mean.

Answer:

The steps to be followed in the construction of CI for mean are:

- Obtain the point estimate of μ, that is, the sample mean .
- Determine the distribution of the sample mean. If n is large, then the Central Limit Theorem can be used and x 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[x bar – 1-96 . S.E(x bar), x bar + 1-96 . S.E(x bar)]

- When reporting a confidence interval, make sure you report both the interval and the confidence level. One without the other is meaningless.

13. How do you determine the standard deviation σ in case of interval estimation?

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.

14. Write a note on interpretation of the interval estimates.

Answer:

- The interval estimate is an interval of values of the sample mean . We hope that this interval contains the population mean μ.
- With repeated random sampling, if a 95% confidence level is selected, the probability is 0.95 that the intervals contain the population mean μ. A particular interval may or may not contain μ but the method employed here means that 95% of intervals are constructed so that they cross the population mean μ. (For example, 95% confidence intervals for the two poor samples – samples 65 and 171 – in the 192-sample simulation do not contain the population mean). See following slide for an illustration of this.
- When reporting a confidence interval, make sure you report both the interval and the confidence level. One without the other is meaningless.

15. What are the cautions to be taken while obtaining interval estimates?

Answer:

Cautions about interval estimates are

- There are many assumptions involved in interval estimation:
  - The sample is randomly selected from a population.
  - The sample size is sufficiently large to use the CLT.
  - The population standard deviation is known or *s* is a good estimate of σ.
  - The selection of a confidence level is an arbitrary process.
  - The population is not too skewed
- As a result, interval estimates are not precise, but are estimates or approximations.
- Larger *n*, repeated sampling, comparisons with other studies, and careful sampling and survey design and practice can improve the quality of the estimates.