What are the 6 assumptions that must be checked when calculating a confidence interval?

Authored by

stats writer

May 2, 2024

A confidence interval is a statistical tool used to estimate the range within which a population parameter, such as a mean or proportion, is likely to fall. In order for this estimation to be accurate, there are six key assumptions that must be checked.

Firstly, the sample must be randomly selected from the population of interest. This ensures that the sample is representative of the population and reduces the risk of bias in the results.

Secondly, the sample size should be large enough to accurately estimate the population parameter. A larger sample size leads to a narrower confidence interval and a more precise estimate.

Thirdly, the data should be normally distributed. This assumption is necessary for the use of certain statistical methods and for the calculation of the confidence interval.

Fourthly, the data should be independent. This means that each observation in the sample should not be influenced by any other observation.

Fifthly, the variability of the population should be known. This can be achieved through previous research or pilot studies.

Lastly, the sample should be a simple random sample, meaning that every possible combination of individuals has an equal chance of being selected. This ensures that the sample is representative of the population and eliminates any potential bias.

Overall, these six assumptions are crucial to consider when calculating a confidence interval as they ensure the accuracy and validity of the estimation.

The 6 Confidence Interval Assumptions to Check

When constructing confidence intervals, it's important that certain assumptions are met. If these assumptions are violated, then the confidence interval can become unreliable.

Here are the six assumptions you should check when constructing a confidence interval:

Assumption #1: Random Sampling

The data should be collected using a (a method in which each individual in a population is equally likely to be included in the sample) so that the sample data you're working with is of the overall population of interest.

Assumption #2: Independence

Each observation in the sample data should be independent of every other observation. This means that no two observations in a sample are related to each other or affect each other in any way.

If you use a random sampling method to collect the data, this assumption is typically met.

Assumption #3: Large Sample

In order to apply the , our sample size must be sufficiently large. In general, we consider "sufficiently large" to be 30 or larger. However, this number can vary based on the underlying shape of the population distribution.

In particular:

If the population distribution is symmetric, sometimes a sample size as small as 15 is sufficient.If the population distribution is skewed, generally a sample size of at least 30 is needed.If the population distribution is extremely skewed, then a sample size of 40 or higher may be necessary.

Assumption #4: The 10% Condition

The sample size should be less than or equal to 10% of the population size. This further ensures that the observations in the data are independent.

Assumption #5: The Success / Failure Condition

When working with confidence intervals that involve proportions, there should be at least 10 expected successes and 10 expected failures in a sample in order to use the normal distribution as an approximation.

Assumption #6: Homogeneity of Variances

When working with confidence intervals that involve two samples, it's assumed that the two populations that the samples came from have equal variances.

For example, if sample 1 has a variance of 24.5 and

sample 2 has a variance of 15.2 then the ratio of the larger sample variance to the smaller would be calculated as 24.5 / 15.2 = 1.61.

Since this ratio is less than 4, we could assume that the variances between the two groups are approximately equal.