



Confidence Interval in Statistics ^[1]

Siddharth Kalla ^[2]77.9K reads

In statistics, confidence interval refers to the amount of error that is allowed in the statistical data and analysis.

Typical Confidence Levels:

One Tailed

>.95 (>95%)

>.99

>.999

Two Tailed

>.975 and <.0275

Confidence Interval

Since statistics uses a sample space and predicts the trends for the whole population, it is quite natural to expect a certain degree of error and uncertainty. This is captured through the confidence interval.

You will frequently encounter this concept while looking at survey results, which take the data of a few people and extend it to the whole group.

Suppose the survey ^[3] shows that 34% of the people vote for Candidate A. The confidence that these results are accurate for the whole group can never be 100%; for this the survey would need to be taken for the entire group.

Therefore if you are looking at say a 95% confidence interval in the results, it could mean that the final result would be 30-38%. If you want a higher confidence interval, say 99%, then the uncertainty in the result would increase; say to 28-40%.

The confidence interval depends on a variety of parameters, like the number of people taking the survey and the way they represent the whole group.

For most practical surveys, the results are reported based on a 95% confidence interval. The inverse relationship between the confidence interval width and the certainty of prediction should be noted.

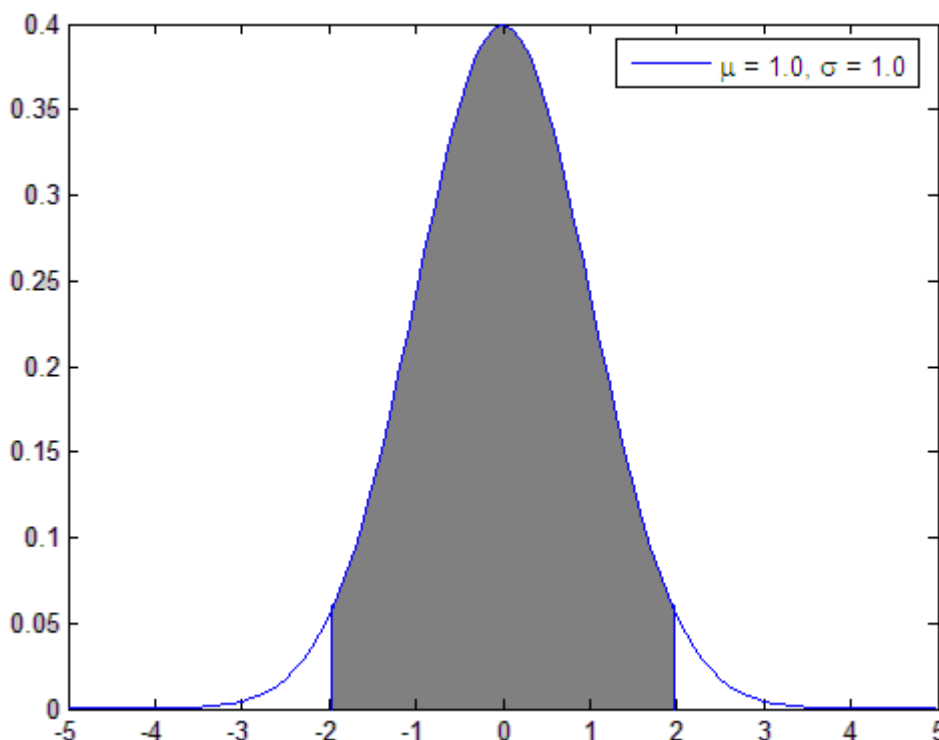
In normal statistical analysis, the confidence interval tells us the reliability of the sample mean as compared to the whole mean.

For example, in order to find out the average time spent by students of a university surfing the internet, one might take a sample student group of say 100, out of over 10,000 university students.

From this sample mean, you can get the average time spent by that particular group. In order to be able to generalize [4] this to the whole university group, you will need a confidence interval that reflects the applicability of this result for the given sample of students to the whole university.

The size of this interval naturally depends on the type of data and its distribution.

For sufficiently large values of sample size, it can be mathematically shown through the central limit theorem that the distribution is approximately normal distribution [5]. In such a case, the 95% confidence level occurs at an interval of 1.96 times the standard deviation [6]. This is shown in the figure below.



The figure can be interpreted as telling us that if one were to repeatedly take samples of the same size from the whole data that is represented through the normal distribution and the confidence interval is calculated in each case, then 95% of these intervals will contain the true mean of the whole population.

Source URL: <https://explorable.com/statistics-confidence-interval>

Links

- [1] <https://explorable.com/statistics-confidence-interval>
- [2] <https://explorable.com/users/siddharth>
- [3] <https://explorable.com/survey-research-design>
- [4] <https://explorable.com/what-is-generalization>
- [5] <https://explorable.com/normal-probability-distribution>
- [6] <https://explorable.com/calculate-standard-deviation>