# Confidence Interval

**G. Ezhumalai***

*Deanery Research Office,*
*Sri Balaji Vidyapeeth,*
*Pondicherry 607402.*

**ABSTRACT**

Most of the research papers are strictly using a p-value to describe a significant value but omitting a confidence interval. Now a day, the publishers are demanding the effect size in terms of confidence interval which is more meaningful than a p-value. Introduction of confidence interval in the research paper will limit the misinterpretation of p-values. This article will help the readers understand the importance of confidence interval while reporting a mean and standard deviation in their article.

**For Correspondence**
**G. Ezhumalai,*
*E-mail: ezhumalaig@mgmcri.ac.in*
Date of
Submisssion: 06-04-2018
Acceptance: 25-05-2018

Access this article online
Quick Response Code
https://www.jbcahs.org

## WHY CONFIDENCE INTERVAL IS NECESSARY?

After making a null hypothesis and an alternative hypothesis, we depend on a p-value to make a decision on whether to reject or accept the null hypothesis. For example, when we want to compare two means, we assume that the null hypothesis is true, i.e. there is no difference between the two means. We reject the null hypothesis when $p \leq 0.05$ and do not reject when $p > 0.05$. When we are rejecting the null hypothesis, alternatively we make a decision to accept the alternative hypothesis and vice versa. Hence the probability (p value) of getting the difference in our study is based on assuming the null hypothesis is true. Hence, when we compare two means, the p-value tells about whether the difference in means is statistically significant or not. It does not tell anything about the size of the effect between the two means[1]

Analysis of a sample data collected from a population deals with descriptive and inferential statistics. Describing the data into percentage, mean, median, standard deviation etc, is referred as descriptive statistics while applying some statistical test to make a conclusion about an unknown population is all about inferential statistics. Descriptive statistics are nothing but estimating the population parameter by a representatively selected sample population. This process is known as estimation.[2] Estimation is of two categories ie., point estimation and interval estimation. Point estimation is a single value calculated from a sample shall be named as an unbiased estimate of the entire population. Interval estimation is a process to estimate a lower and upper value that includes the population mean. This is called confidence interval.

## CONFIDENCE INTERVAL FOR MEAN

E.g., We have a set of 15 individuals (population) diastolic blood pressure as 80,82,86,78,74,90,87,74,70,97,78,70 ,73,70,80 with an average of 80

When we draw a sample of 5 from the population of 15, say 80,82,70,78,90, we get an average of 80 [(80+82+70+78+90)/5], Which is equal to the population mean 80. Hence, this estimate is an unbiased estimate of the population mean. Do we get the same 80 as mean every time when we draw a sample of 5 or any other size? Not always. Research is done only once with a sample, aiming to estimate a reliable estimate and to generalise the results about the population. Now, how to say that the point estimate what we have done from the sample represents the population value.

How confident about the estimate is realistic or accuracy? Can we repeat the research many times to get a reliable estimate? The solution is the interval estimate.

The interval estimate gives a lower limit and an upper limit for the estimate so as any sample value, tested many times, each will lie within these intervals with a confidence level. If we fix up a 95% confidence level, we are 95% sure that the population mean will lie within the interval. When we repeat the experiment 100 times, get 100 mean values and we calculate the mean of all 100 means that value is called population mean. But, for getting the population mean, we need not repeat our experiment 100 times but, we can estimate an interval from only one sample value in which the population mean will lie.

Now we will make a random sample of 5 out of 15. Let them be 78,74,70,80 and 87

**Table 1:** Working example of variance estimation

| Value (X) | Deviation (X-77.8) | Squared deviation |
|---|---|---|
| 78 | 0.2 | 0.04 |
| 74 | -3.8 | 14.44 |
| 70 | -7.8 | 60.84 |
| 80 | 2.2 | 4.84 |
| 87 | 9.2 | 84.64 |
| Sum 389 | | 164.8 |
| Average = 389/5 = 77.8 | | Variance = 164.8/4 = 41.2 |

Variance = 41.2

Standard deviation = √variance = √41.2=6.42

Standard error = S.D/√n =6.42/√5 = 2.88

Now, the mean is 77.8, Standard deviation =6.42, Standard error=2.88

The confidence interval is End points of interval= Estimate ± (confidence coefficient x standard error of the Estimation.[2])

Thus, Lower 95% confidence limit = 77.8 - 1.96 × 2.88 = 72.16

Upper 95% confidence limit = 77.8 + 1.96 × 2.88 = 83.44

It is understood that the population mean 80 will lie between 72.16 & 83.44. In general, we do not know the population mean but we are able to detect a minimum and maximum value, in between the population mean will lie. This is called 95% confidence interval. Likewise, we can calculate a 99% confidence interval. Here we have to replace 1.96 by 2.57 and the remaining values are same. Now we will get a wider confidence interval i.e., 70.4 to 85.2.

## CONFIDENCE INTERVAL FOR PROPORTIONS

Many of the researchers calculate the confidence interval for the mean value only but neglect the same in case of proportions. For any proportion, confidence interval is equally important as it is for mean. Calculation of CI for proportion is given in the following example

Let us assume that the hospital acquired infection (HAI) was 30% in a hospital from a sample of 60 patients.

Here, p=0.3 and 1-p=1-0.3=0.7

Now, standard error for the proportion will be = √p(1-p)/n = √0.3 *(0.7)/60 = 0.059

Standard error = 0.059

The confidence interval is as calculated above.

Thus, Lower 95% confidence limit = 0.3 - 1.96 × 0.059 = 0.184

Upper 95% confidence limit = 0.3 + 1.96 × 0.059 = 0.416

Hence, we are 95% confidence that the hospital acquired infection in the hospital will lie between 0.184 – 0.416 i.e., between 18.4% and 41.6%.

## INTERPRETATION

When we estimate a CI for a single mean, it tells about where the population mean will lie. When we are comparing two means, we should calculate confidence interval for the difference in means. If such

confidence interval does not include zero, the difference is significant otherwise not significant[1].

## CONCLUSION

p-value may determine whether the outcome is reliable and replication is possible.4 P-value tells whether the result is by chance or actually true, while confidence interval tells about the true effect occurred in the research. Thus, CI is more robust than the p-value. Many editorials demand for the real effect and the confidence intervals. Whenever we report mean & standard deviation in a research paper, it is advisable to include 95% CI.

## CONFLICTS OF INTEREST

None.

## *References*


1. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN et al Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016; 31:337-50
2. Mrs Spitz spring, Confidence interval for the means (small sample), Publisher Mitchel Smithwick, http://slideplayer.com/slide/2811914/ chapter 6.2, 2009
3. Mrs Spitz spring, Confidence interval for proportions, Publisher Millicent Barton, http://slideplayer.com/slide/5827799/ chapter 6.3, 2009
4. Yuko Y.Palesch, some common misperceptions about p-values, NIH public access Author Manuscript, 2014;45(12) e244-e246


## Work stress may be deadly for men with heart disease and diabetes

Men with heart disease, diabetes or a history of stroke are more likely to die prematurely when they have a stressful job even when they're relatively healthy, a large European study suggests and published in the Lancet Diabetes & Endocrinology.

For the current study, they examined data on 102,633 men and women living in Finland, France, Sweden and the UK who participated in one of seven studies examining the relationship between work stress and mortality. They found that work stress is particularly harmful for those with problems in the cardiovascular and metabolic systems, such as those with diabetes, heart disease or a history of stroke and this excess risk remained even if the person was free of conventional risk factors, such as smoking, high blood pressure or high cholesterol concentration.

Researchers examined two aspects of work stress: having high demands or responsibility but little control or authority, and having a large difference between effort and reward.