International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 6, ISSN No. 2455-2143, Pages 278-282
Published Online October 2019 in IJEAST (http://www.ijeast.com)

CONFIDENCE INTERVAL IS MORE INFORMATIVE THAN P-VALUE IN RESEARCH

Shyamal Kumar Das
Independent Researcher

Abstract - Statistical significance can be obtained from Hypothesis testing using both p-value and Confidence Interval but p-value fails to convey the complete picture of direction and strength of effect of an intervention, it provides a binary (yes / no) decision only not even an effective size. Where as Confidence Interval can provides the same including a range of possible plausible values for the target population as well as the probability with which the range covers the real value. Analysis and synthesis of research results based only on p-values may have less degree of intensity and its effect is not necessarily meaningful in the present day research. As for the same Confidence Interval is more informatics and provide higher degree of intensity than that of p-values.

Key words: Hypothesis testing, Confidence Interval, p-value, binary decision.

I. INTRODUCTION

Since long back many international journals do not accept articles on medical science, business, marketing etc. for publication if analyzing the significance was based only on P-value and always insisted upon calculating confidence interval which has become mandatory in modern research. I fill the same strategy are to be followed in educational research even as, in and for education homogeneous approach always goes for heterogeneous group, due to different demographic and SES of the parents, prevalent public and private sector of education. The statistics drawn from different random sample groups from same population fluctuates a lot, of course practically it may be a tedious job for a researcher. Again if the researcher goes for single statistics that may provide hardly any authenticated information.

The primary focus for most research studies is the parameter of the population [1] not statistics [2] calculated for the particular representative sample mean selected because the sample and statistics describing are important only insofar as they provide information about the unknown parameters. Once, the samples mean is achieved it is tried to 'estimate' the average population mean that hardly describe accurately the population of interest. It is pertinent to say statistics drawn from different sample groups from a population under study reflects different. So the inference makes on some specific unknown parameter based on statistics (point estimate, a single value) may be biased as we are generalising the true parametric value just from a sample, there may be high probability of some uncertainty associated in the estimation process though that is easy to calculate and comprehend but never provide any indication how accurate the estimation is? But to deal with uncertainty interval estimate is the best that provides the concept of confidence levels [3] and Confidence interval [4] range that best describe the population of interest. This is useful in Statistics when we would like to know how confident we are in answers we get from the data.

For example, suppose I like to study the scholastic achievement in mathematics of class-XII grade students. I shall take a random sample from the population and establish a mean achievement score of 71 marks. The mean of 71 is a point estimate of the population mean. A point estimate by itself is of limited usefulness because it does not reveal the uncertainty associated with the estimate; I do not have a good sense of how far this sample mean may be from the population mean. What I am missing at this point is the degree of uncertainty in this single sample.

Confidence intervals provide more information than point estimates. By establishing a 95% confidence interval using the sample's mean and standard deviation, and assuming a normal distribution as represented by the bell curve, I arrived at an upper and lower bound that contains the true mean 95% of the time. Assume the interval is 69 marks to 74 marks. If I take 100 random samples from the population of class-XII grade mathematics students as a whole, the mean should fall between 69 and 74 marks in 95 of those samples.

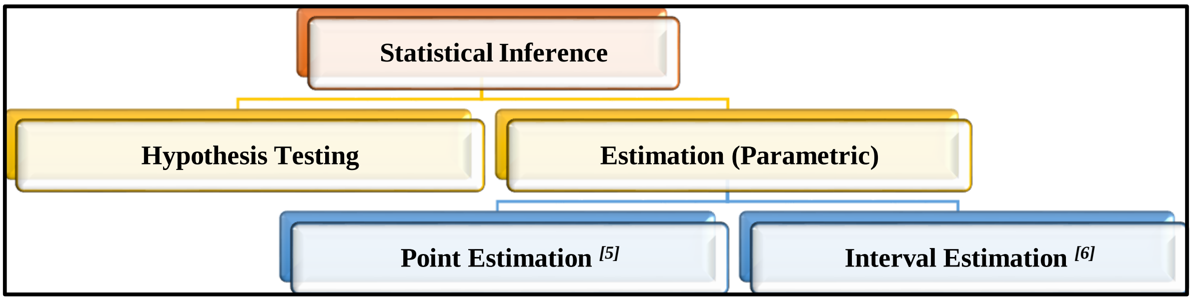
I may even go for greater confidence; they can expand the interval to 99% confidence. Doing so

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 6, ISSN No. 2455-2143, Pages 278-282
Published Online October 2019 in IJEAST (http://www.ijeast.com)

invariably creates a broader range, as it makes room for a greater number of sample means. If I establish the 99% confidence interval as 75 marks to 78 marks, that can expect 99 of 100 samples evaluated to contain a mean value between these numbers.

II. CONFIDENCE INTERVAL IS MORE INFORMATIVE THAN *P*-VALUE

The main aim of an educational research is to estimate the effect of an intervention or exposure in a certain population. Hypothesis testing (using *P*-values) and point / interval estimation (using confidence intervals) are two concepts of inferential statistics that help in making inference about population from samples. P-value and Confidence Interval are common statistics measures which provide complementary information about the statistical probability and conclusions regarding the clinical significance of study findings.



2.1 Hypothesis testing

People often think p-value means ‘probability’; they are related but not the same. The p-value is the standard against which we compare our results; significantly different or not significantly different. The computed p-value is compared with the p-value criterion to test statistical significance. The p-value criterion is traditionally set as $p \leq 0.05$. If the computed p-value is less than criterion, we can achieve statistical significance. In general smaller the p-value is better.

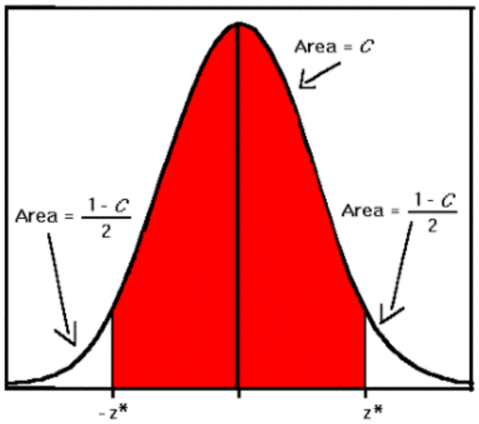
Reject the null hypothesis (reached statistical significance), if $p \leq 0.05$.

Fail to reject the null hypothesis (has not reached statistical significance), if $p \geq 0.05$.

It never interprets how large or small the effect is and how strong is the relationship? In exploratory studies, p-values enable the recognition of any statistically noteworthy findings.

2.2 Confidence Intervals

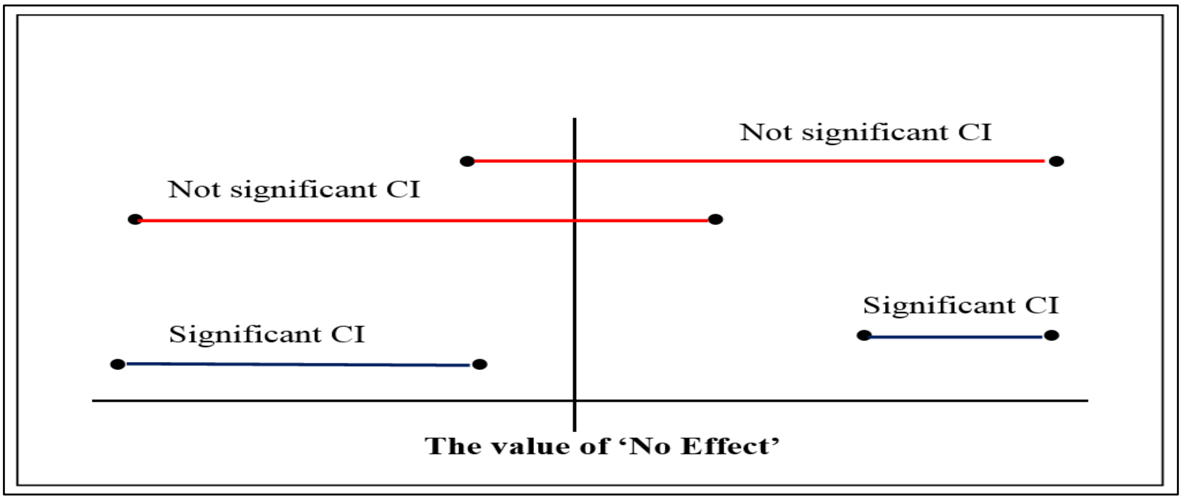
A confidence interval for the mean is a way of estimating the true population mean. Instead of a single number for the mean, in statistics, it is a type of range / interval estimate, computed from the statistics of the observed data that might contain the true value of an unknown population parameter. Most frequently, it is used to bind the mean or standard deviation. Measurements from a sample are only an estimate of the population. Confidence intervals serve as good estimates of the population parameter because the procedure tends to produce intervals that contain the parameter. Confidence intervals are comprised of the point estimate (the most likely value) and a margin of error around that point estimate. The margin of error indicates the amount of uncertainty that surrounds the sample estimate of the population parameter. It specifies how far above or below a sample based value the population value lies [within a given range]. It also gives information about the precision of an estimate (certainly). The centre of the confidence interval (the sample mean) is the most plausible value for the population mean. The ends of the confidence interval are less plausible values for the population mean. When using repeated measures or paired group design use a confidence interval around the mean difference. P-values can be estimated from the graphs of the confidence intervals of two independent sample means. If the confidence intervals overlap by half a margin of error, then, $p = 0.05$. When they are just touching, then $p = 0.01$.



The selection of a confidence level for an interval determines the probability that the confidence interval produced will contain the true parameter value. Common choice for the confidence level *C* is 0.95. These levels correspond to percentages of the area of the normal density curve. For example, a 95% confidence interval covers 95% of the normal curve, the probability of observing a value outside of this area is less than 0.05. Because the normal curve is symmetric, half of the area is in the left tail of the curve, and the other half of the area is in the right tail of the curve. As shown in the diagram to the right, for a confidence interval with level *C*, the area in each tail of the curve is equal to (1-*C*)/2. For a 95% confidence interval, the area in each tail is equal to 0.05/2 = 0.025. The value z^* representing the point on the standard normal density curve such that the probability of observing a value greater than z^* is equal to *p* is known as the upper *p* critical value of the standard normal distribution.

2.3 Confidence interval as a test for significance

For the assessment of statistical significance of any estimate, confidence interval can be used as hypothesis testing. If the confidence interval captures the value of 'no effect' this represents a statistically non significant result. If the confidence interval does not capture the value of 'no effect', this represents a difference that is statistically significant.



The value of 'no effect' is the absolute measure, i.e. absolute risk, absolute risk reduction and the number needed to treat. Means if a specific intervention leads to zero (0) risk reduction [Risk in control group – Risk in intervention group = 0], it has no effect compared with the control. Thus in situations dealing with absolute measures the value of no effect is zero.

So if the confidence interval measured for the absolute risk reduction of an exposure ranges [-3, 4] i.e. [-3, -2, -1, 0, +1, +2, +3], it contains zero (0) hence the risk reduction is not statistically significant.
Second the value of 'no effect' is the ratios i.e. relative risk or odd ratio. 1 in case of relative risk ratio or odds ratio [The incidence of outcome in the intervention (or exposed) group = The incidence of outcome in the control group]. Means intervention or exposure is neither beneficial nor harmful compared with the control group.
So if the confidence interval measured for the relative risk or odd ratio of an exposure ranges [0.8, 4] i.e. [+0.8, +0.9, +1, +2, +3, +4], it contains one (1) hence the relative risk ratio is not statistically significant.
Thus, not only 'statistical significance' ($P<0.05$) can be inferred from confidence intervals, but also these intervals show the largest and smallest effects that are likely to take place.

International Journal of Engineering Applied Sciences and Technology, 2019
Vol. 4, Issue 6, ISSN No. 2455-2143, Pages 278-282
Published Online October 2019 in IJEAST (http://www.ijeast.com)

	Value of no effect	95% CI	P-value
Measure of effect	0	Includes 0 (Not significant)	P > 0.05 (Not significant)
	0	Not include 0 (Significant)	P < 0.05 (Significant)
	1	Includes 1 (Not significant)	P > 0.05 (Not significant)
	1	Not include 1 (Significant)	P < 0.05 (Significant)

2.4 P-values and Confidence Intervals always agree about Statistical Significance

Both P-values and Confidence Intervals can be used to determine whether results are statistically significant. If hypothesis test produces both, these results will agree.

We have, Confidence Level = 1- α (alpha /significance level)

Hence if the alpha /significance level are 0.05, the corresponding confidence level is 95%.

● If the P-value is less than alpha/significance level, the hypothesis test is statistically significant.

● If the Confidence Interval does not contain the null hypothesis value, the results are statistically significant.

● If the P-value is less than alpha /significance level, the Confidence Interval will not contain the null hypothesis value.

In order to understand why the results always agree, it is better to know how both Significance Level and Confidence Level work does.

● The significance level defines the distance the simple mean must be from the null hypothesis to be considered statistically significant.

● The Confidence level defines the distance for how close the confidence limits are to sample mean.

2.5 P-values vs. Confidence Interval

● Statistical significance obtained from Hypothesis testing using p-value alone cannot convey the complete picture of effectiveness of an intervention.

● Hypothesis testing using p-value is a binary (yes / no) decision. P value indicates nothing about effect size; it only indicates that the difference is not by chance. It is the probability of committing Type I error.

● Though P- values are clear than the Confidence Interval yet Confidence Interval provide information about statistical significance [7], as well as, direction and strength of effect.

● Confidence Interval provides a range of possible plausible values for the target population, as well as the probability with which the range covers the real value.

● In contrast with confidence interval p-values gives the difference from a previously specified statistical α level.

● Statistical significance must be distinguished from medical relevance or biological importance.

III. CONCLUSION

In today's age of virtually instantaneous access to vast repository of knowledge and information, statistical analysis is being increasingly, universally used and it entwined almost every research field to a greater focus on p-values for simple detection of significant effect or different. Interpretation of results based only on p-values can be misleading and its effect is not necessarily meaningful in the real world. For instance, the effect might be too small to be of any practical value. But confidence interval provides more information than p-values. It provides magnitude of effect as well as its variability. Confidence interval should be calculated for each variable especially if p-values are insignificant. It's important to pay attention to the both the magnitude and the precision of the estimated effect. It allows assessing these characteristics along with the statistical significance. And for a narrow confidence interval where the entire range represents an effect that is meaningful in the real world.

IV. REFERENCE

1. Abdelhamid Attia, M.D.: "Why should researchers report the confidence interval in modern research"? Middle East Fertility Society Journal, Vol-10(1), (pp. 78-81), 2005

2. Berkson, J.: "Tests of significance considered as evidence". Journal of the American Statistical Association, Vol-37 (219), (pp. 325-335), 1942

3. Cohen, J.: "Some statistical issues in psychological research". In B. B. Wolman (Ed.), Handbook of clinical psychology, (pp. 95-121), New York: McGraw-Hill, 1965

4. Cumming, G.: "Pictures of confidence intervals and thinking about confidence levels". Teaching Statistics, Vol-29 (3), (pp. 89-93), 2007

5. Cumming, G., & Finch, S.: "Confidence intervals and how to read pictures of data".

American Psychologist, Vol-60 (2), (pp. 170-180), 2005

6. Fisher, R.A.: “Note on Dr. Berkson's criticism of tests of significance”. Journal of the American Statistical Association, Vol-38 (221), (pp. 103-104), 1943

7. Gardner M.J, Altman D.G.: “Confidence intervals rather than p values: estimation rather than hypothesis testing”. BMJ, Vol-292, (pp. 746-750), 1986

8. Harari G.: “Confidence interval or p-value--similarities and differences between two important methods of statistical inference of quantitative studies”. Harefuah, Vol-153,(pp. 48-51), 2014

9. Harari G, Herbst S.: “Confidence interval calculation for small numbers of observations or no observations at all”. Harefuah, Vol-153: (pp. 289-291, 304, 303), 2014

10. Knapp, T.R.: “Some reflections on significance testing”. Journal of Modern Applied Statistical Methods, Vol-1 (2), (pp. 240-242), 2002

11. Levin, J. R.: “Statistical significance testing from three perspectives”. Journal of Experimental Education, Vol-61, (pp. 378-382), 1993

12. Shaver, J. P.: “What statistical significance testing is, and what it is not”. Journal of Experimental Education, Vol-61, (pp. 293-316), 1993

End Notes

[1] A parameter is an unknown numerical summery of the population.

[2] A statistics is a known numerical summery of the sample which can be used to make inference about parameter.

[3] A confidence level is the probability that the interval estimate will include the true population parameter (such as the mean).

[4] A confidence interval is a range of values used to estimate a population parameter, based on one or more samples and is associated with a specific confidence level.

● A 95% confidence interval does not mean that 95% of the sample data lie within the interval or there is 0.95 probability that true parameter value lies within this interval.

● 95% Confidence Interval simply means that if I take out samples a large number of times (tending to infinity) from a population with 0.05 significance level then 95% of those intervals will encompass the true parameter value.

[5] In statistics, Point estimation, the process of finding an approximate value of some parameter of a population from random samples of the population.

[6] In statistics, interval estimation is the use of sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number.

[7] Major advantage of confidence interval is that we can also access significance from a confidence interval. If the confidence interval captures the value of ‘no effect’ (i.e. 1 in case of risk ratio or odds ratio and 0 in case of rate difference) this represents a statistically non significant result. If the confidence interval does not enclose the value of ‘no effect’, this represents a difference that is statistically significant. Thus, not only ‘statistical significance’ ($P<0.05$) can be inferred from confidence intervals, but also these intervals show the largest and smallest effects that are likely to take place.