

Statistical Analysis (Confidence Intervals and P-Values)

Posted on November 1, 2018

In statistical analysis, hypothesis testing is an essential factor. The primary purpose of a statistical study is to help us perform hypothesis testing. A hypothesis can be described merely as an educated guess about a specific process.

Its validity is always tested either through observations or experiments, and then its own conclusions are made about the hypothesis to determine whether it's true or false. One basic function of a hypothesis is to check the validity of an independent variable against a dependent variable through the use of hypothesis statements. An example of a hypothesis statement would be, 'If the teacher gives students a math test in the morning rather than in the afternoon, the student test scores will improve.' In this example, the independent variable is offering students exams in the morning rather than in the afternoon, while the dependent variable is the student scores. When using statistical methods, required data is collected, and after that, the statistical operation is done on the received data to come up with sound conclusions. The conclusions determine whether the stated hypothesis should be accepted or rejected. The paper shall discuss some of the key methods in statistical analysis, which are confidence intervals and p-values (Altman et al. 153).

The P-value approach helps researchers to determine the likeliness of an event occurring; that is, the p-value is useful in determining if the hypothesis statement holds or not. This is made possible through the use of the concept known as a null hypothesis. The concept works by first assuming that the null hypothesis is true, and then statistical experiments and operations are undertaken to decide whether it's true or false.

A small P-value indicates unlikeliness, while a large P-value supports the null hypothesis. Usually, the size of a P-value is measured using a statistic denoted as α (Alpha). If $P < \alpha$, this means that there is a situation of unlikeliness, and if $P > \alpha$, there is a high probability that the null hypothesis is true (Vickers 55). The following are the main steps in determining the hypothesis using the p-value approach:

- First, the null and alternative hypotheses are identified from the statement of the research problem.
- The null hypothesis is then presumed right, and a test statistic value is worked out from the sample data through the use of precise formulations and methods.
- Then, using the distribution of the test statistic, the p-value is calculated, and then conclusions are derived based on the value. For example, if the null hypothesis is correct, then we consider the probability of observing an extreme statistic about the alternative hypothesis.
- A level of significance is set, and it is usually denoted by the symbol α , representing the

probability of Type I errors.

- The p-value is, therefore, compared to the significance level, and if $P < \alpha$, then the null hypothesis is rejected. But if $P > \alpha$, then the null hypothesis is accepted.

The table below shows a summary of the relationship between hypothesis testing and p-value.

Decision		
	Accept H_0 :	Reject H_0 :
H_0 is true:	correct decision	type I error
	$1 - \alpha$ $P > \alpha$	α ($P < \alpha$)
H_0 = null hypothesis		
P = probability		

In statistical analysis, it is hard to understand or even use the concept of P-values without proper knowledge of the aspect of the confidence interval (CI). A confidence interval (CI) is a range of values that are well-defined, and there is a quantified probability that the values of a parameter lie within it. This simply implies that the confidence level represents the proportion of statistical data within which the true value of the corresponding constraints is contained. A confidence interval is, therefore, a range of values that act as a reliable estimate of an unknown parameter (sample or population).

Nevertheless, intervals that are worked out from a given sample do not mean that the true value is included within them. This is because the samples are simply random collections of data from the whole population under study, thus making the confidence interval from such data to be arbitrary parameters. In hypothesis testing, the confidence level is always set by the scholar, and it is a compliment to the significance level. For example, a confidence level of 95% is a reflection of a 5% level of significance. The confidence level is always presented in the research question and during problem definition. The most common confidence level used by researchers is 95%, although other levels may be used, such as 99% or 90%.

In conclusion, for us to get a better understanding, it is sensible to compare the concepts of point estimates and interval estimates. A point estimate represents a single value parameter, such as mean, median, and percentage estimates. On the other hand, an interval approximates the series within which the parameters are expected to be contained. Nevertheless, confidence intervals and point estimates are usually used together hand in hand for reliability purposes. The main factor in

estimating the size of an interval is the population or sample size, which is used in the estimation process (Schuller and Joseph 99). Hence, a confidence interval is mainly used to show the reliability of the results obtained from surveys or observations. For example, in a voting survey to determine the voting intentions of people, the results might be that around 40% of the total voters intended to vote for a particular politician. A 90% confidence interval would mean that the proportion of people from the whole population willing to vote for a particular politician would be 37%-43%.

Works Cited

Altman, Douglas, David Machin, Trevor Bryant, and Martin Gardner. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. New York, NY: John Wiley & Sons, 2013. Internet resource.

Vickers, Andrew. *What Is a P-Value Anyway?: 34 Stories to Help You Actually Understand Statistics*. Boston: Addison-Wesley, 2010. Print

Schmuller, Joseph. *Statistical Analysis with R for Dummies*. Hoboken, NJ: John Wiley & Sons, 2017. Print.