The p-value tells us about the likelihood or probability that the difference we see in sample means is due to chance. Thus, it really is an expression of probability, with a value ranging from zero to one.

Let us first consider an experiment where two samples are measured and their means are found to be different. Now this may happen due to two reasons. The populations may truly have different means. But there is also a small chance that the large difference observed would have occurred even if the population means were identical.

The p-value is a measure of how much evidence we have against the null hypothesis. The most important thing to remember about the p-value is that it is used to test hypotheses. It is a measure of how much evidence we have against the null hypothesis, which is the hypothesis of no change or no difference. The smaller the p-value, the more evidence we have against the null hypothesis.

Very often, a p-value less than 0.05 leads us to conclude that there is evidence against the null hypothesis and we say that we reject the same at 5%. A p-value less than 0.01 will under normal circumstances mean that there is substantial evidence against the null hypothesis.

One-tailed and Two-tailed P-values

P-values may either be one-tailed or two-tailed. A one-tail p-value is used when we can predict which group will have the larger mean even before collecting any data. But if the other group ends up with the larger mean, we should attribute that difference to chance, even if the difference is large. For this reason it is usually best to use a two-tail p-value as such a situation leads us to conclude that the difference is not statistically significant. This can be avoided by using two-tail p-values from the very beginning. Also a two-tail p-value is more consistent with the p-values reported by tests which compare three or more groups.

Misconception About the P-value

The main disadvantage of a p-value is that it is commonly misinterpreted. Many people misunderstand what question the p-value ultimately answers. For instance, if the p-value is 0.03, then what it means is that there is a 3% chance of observing a difference as large as observed in the particular experiment between the sample means even if the population means are identical.

It does not in any way imply that there is a 97% chance that the differences observed is due to real differences between populations and a 3% chance that the difference is due to chance.

Simply put, it means that if population means are identical then randomness in sampling would lead to smaller differences between sample means than we observed in 97% of experiments and larger differences in 3% of experiments.

To put it more simply, the p-value refers to the percentage of experiments in which the sample differences would be larger or smaller than we observed.

Dos and Don'ts

There are certain dos and don’ts that should be kept in mind when using p-values. The p-value should not be interpreted as the probability that the null hypothesis is true. A hypothesis is not a random event that can have a probability. Therefore we do not predict the probability of the hypothesis happening.

Rather, we try to infer whether it is true or not. We should be cautious while dealing with a small p-value. Also, a large p-value should not be taken as evidence in support of the null hypothesis as an inadequate sample size may have resulted in such a large value.