

# P Value, Statistical Significance and Clinical Significance

Padam Singh, PhD Gurgaon, India

---

### ABSTRACT

---

In medical research it is desirable to indicate the  $p$ -value while presenting the results of statistical tests of significance. The  $p$  value is calculated using the null distribution of test statistic. The computed  $p$ -value is compared with the level of significance while drawing the inference. All statistically significant differences need not be clinically significant. In view of this, it is suggested to appropriately frame the hypothesis to detect the Medically Important Clinical Difference (MICD) as significant. (J Clin Prev Cardiol. 2013;2(4):202-4)

**Keywords:** Clinical significance, Level of significance, Null distribution,  $p$ -value, Type I error

---

### Statistical Test of Significance

---

The theory of statistical test of significance essentially involves setting of a null Hypothesis and competing alternative hypothesis. The null hypothesis ( $H_0$ ) is a hypothesis of “no difference, no effect, no correlation, no association etc.” The null hypothesis is generally the negation of the research hypothesis.

The alternative hypothesis ( $H_1$ ) is the competing alternative to the null hypothesis and is usually the research hypothesis set out to investigate.

As an example, if we are investigating a new drug for its superiority over a standard drug, the null and alternative hypothesis would be as under:

Null hypothesis ( $H_0$ ): There is no difference in the treatment outcome between the new drug and the standard drug.

The alternative hypothesis ( $H_1$ ): The treatment outcome

with the new drug would be better as compared to the standard drug.

Mathematically,

$$H_0: p_n = p_s,$$

$$H_1: p_n > p_s$$

Where

$$p_n = \text{Efficacy of new drug}$$

$$p_s = \text{Efficacy of standard drug}$$

While drawing an inference in the testing of hypothesis following are used:

- i) **The level of significance ( $\alpha$ ):** The probability of Type I error (rejecting the null hypothesis ( $H_0$ ) when it is true)
- ii) The **estimated  $p$ -value** for given data set of the study.

If calculated  $p$ -value is less than the chosen level of significance then the null hypothesis is rejected.

### The $p$ -value and null distribution

---

The  $p$  value is calculated based on the Null distribution, which is a theoretical distribution of the test statistic when the null hypothesis is true. The commonly used null distributions and their applications are given as under:

---

From: Medanta - The Medicity, Sector 38, Gurgaon, Haryana-122001, INDIA

Corresponding Author: Padam Singh, Ph.D.  
Medanta - The Medicity, Sector 38, Gurgaon, Haryana-122001, INDIA

Ph: +91-124-4141414 I Fax: +91-124-4834111

Email: dr.padamsingh2013@gmail.com

Situation/ Application	Null distribution
Testing the difference of averages /proportions between 2 groups for large samples	The standard normal distribution
Tests for difference of averages between 2 groups for small samples	The t distribution
Tests for difference of proportions between 2 groups for small samples	The Binomial Distribution
Testing the independence of attributes	The chi-square distribution
Testing the difference of averages for more than two groups	The F distribution

Given the value of the test statistic and the null distribution of the test statistic, it is important to see whether the test statistic is in the middle of the distribution or out in a tail of the distribution. If in the middle of it is consistent with the null hypothesis and out in the tail making the alternative hypothesis seems more plausible. The large values of the test statistic seem more plausible under the alternative hypothesis than under the null hypothesis. The larger values of test statistic corresponds to the smaller the  $p$ -value, which are indicative of the stronger evidence against the null hypothesis in favor of the alternative.

### Interpretation of $p$ -value

The  $p$ -value indicates how *probable* the results are due to chance.

$p=0.05$  means that there is a 5% probability that the results are due to random chance.

$p=0.001$  means that the chances are only 1 in a thousand.

The choice of significance level at which you reject null hypothesis is arbitrary. Conventionally, 5%, 1% and 0.1% levels are used. In some rare situations, 10% level of significance is also used.

Statistical inferences indicating the strength of the evidence corresponding to different values of  $p$  are explained as under:

Values of $p$	Inference
$p > 0.10$	No evidence against the null hypothesis.
$0.05 < p < 0.10$	Weak evidence against the null hypothesis
$0.01 < p < 0.05$	Moderate evidence against the null hypothesis
$0.05 < p < 0.001$	Good evidence against null hypothesis.
$0.001 < p < 0.01$	Strong evidence against the null hypothesis
$p < 0.001$	Very strong evidence against the null hypothesis

Conventionally,  $p < 0.05$  is referred as **statistically significant** and  $p < 0.001$  as **statistically highly significant**. When presenting  $p$  values it is a common practice to use the asterisk rating system.

Important in medical research,  $p$  values less than 0.05 are often reported as statistically significant as we want there to be only a 5% or less probability that the treatment results, risk factor, or diagnostic results could be due to chance alone. Results that do not meet this threshold are generally interpreted as negative.

### $p$ -value and sample size

The magnitude of  $p$ -value depends on the sample size. In view of this, following is important to note:

- Given a sufficiently large sample size, even an extremely small difference could be detected as significant. However, the difference may not be of clinical importance.
- Conversely, potentially medically important differences observed in small studies, for which the  $p$  values are more than 0.05, are ignored.
- All statistically significant findings at ( $p < 0.05$ ) are assumed to result from real treatment effects, whereas by definition an average of one in 20 comparisons in which null hypothesis is true will result in  $p < 0.05$ .

### Clinical Significance

One should keep in mind the difference between statistical significance and clinical significance. The results of a study can be statistically significant but still be too small to be of any practical value.

We might encounter a situation where new treatment shows reduction in pain as statistically significant with a  $p$  value of 0.0001. The extremely low  $p$  value in this situation indicates that we are really sure that the results are not accidental-- the improvement is really due to the treatment and not just chance. However, it is important to note the magnitude of improvement while interpreting it clinically.

Usually, the extremes are easy to recognize and agree upon. If with the new treatment being investigated, patients get 90% pain relief, we can all agree that this is an effective, worthwhile treatment. But if reduction

in pain is just 3%, such a paltry effect may not be of any importance.

It is thus important to decide whether the observed treatment effect is large enough to make a difference.

But what would constitute the smallest amount of improvement (say  $\delta$ ) that would be considered worthwhile. After all we want the new treatment to make a difference. This is tricky. The researchers doing the study should explicitly state what this minimal amount of clinically important improvement is.

One can find the absolute amount of improvement in critical outcome measure, to decide if the improvement looked very large and can cross reference it with other sources to decide whether it met the Medically Important Clinical Difference (MICD). The null and alternative

hypothesis could be appropriately framed to detect the Medically Important Clinical Difference as significant.

The Null and Alternative Hypothesis in this situation would be as under:

$$\mathbf{H}_0: p_n = p_s + \delta$$

$$\mathbf{H}_1: p_n > p_s + \delta$$

It is therefore advised, not to just look at the  $p$  value but also try to examine if the results are robust enough to be clinically significant with desired improvements (MICD).

### **References**

1. Krikwood BR, Sterne JAC. Medical statistics, 2<sup>nd</sup> edition, Oxford Blackwell Science Ltd; 2003