Print

# The p-Value and Statistical Significance

In the previous video, we introduced the idea that you can expect a sample mean to fall within a range of values when the null hypothesis is true, whereas some values are unusual but plausible, and others are implausible, or extremely unlikely, under the null hypothesis.

In a statistical test, we ask, "How likely is our result under the null hypothesis?" and we set a boundary beyond which the sample is so implausible that the null hypothesis is simply not credible. We use words such as "implausible" and "credible" because a sample cannot tell us the truth about a population with certainty. Hypothesis tests are, ultimately, about decision making in the face of uncertainty. There is always a chance, however small, that you'll get an implausible sample mean when the null hypothesis is true.

Conversely, there's a chance that a population that is consistent with the alternative hypothesis delivers a sample mean that appears to have come from a null population. You learn more about these prospects for erroneous conclusions later. For now, let's focus on the idea of setting boundaries for rejecting the null hypothesis.

There is no hard and fast rule on where to set the boundary for rejection. Consider this question: What range of results would be surprising if the null hypothesis were true? In the 1920s, Sir Ronald A. Fisher suggested that events that occur 1 time or fewer out of 20 might be considered implausible. Fisher's rather casual "1 in 20" suggestion found wide acceptance as the conventional boundary.

In statistical tests, we use a "probability value," or p-value, to guide our decision making. A p-value is a measure of the strength of the evidence against the null hypothesis. The smaller the p-value, the stronger the evidence against the null. If we have a high p-value, we have little evidence against the null.

Sample results with p-values smaller than a chosen value, called the significance level, are referred to as statistically significant. Significant results are surprising or unusual enough to warrant attention, because they are unlikely to have occurred just due to random sampling variation.

Per Fisher's suggestion, the significance level, alpha, is often set at 0.05. However, this is not a hard and fast rule. Suppose that you conduct a test and determine that your results are statistically significant. Accordingly, you reject the null hypothesis and accept the alternative.

However, having significant results does NOT mean that you are correct in rejecting the null hypothesis! When you reject the null hypothesis, there is no guarantee that you are making the correct decision. The goal of statistical testing is to make informed decisions, but there is always a risk of committing an error.

In the terminology of statistical testing, when an unusually rare sample result leads you to mistakenly reject a correct null hypothesis, you have committed a Type I error. Your test has led to a false positive.

Conversely, if the null hypothesis is false and your test does not lead you to reject the null, you have committed a Type II error. Your test has led to a false negative.

Statistical tests are designed to minimize the risk of committing a Type I error. You, the analyst, determine the significance level for the test. This significance level, alpha, is the risk of a Type I error that you are willing to accept.

You learn more about Type I and Type II errors in the sample size lesson at the end of this module. Another thing to consider, in any statistical test, is that a low p-value does not signify that the result is meaningful or important from a practical perspective.

Statistical significance means that the results are unlikely to have occurred by chance alone, given the null hypothesis. You, the decision-maker, have knowledge of the decision context.

You must decide whether the difference between the observed results and the hypothesized value is large enough to be practically meaningful or relevant. We revisit this idea of practical significance in the next lesson.





Close