Print

# Equivalence Testing

You've learned that when you don't reject a null hypothesis, you cannot accept that the null hypothesis is true. You simply fail to reject it. For example, let's revisit the metal parts pilot scenario. Your target is 40 hundredths of an inch. You test the null hypothesis that the mean thickness is 40 hundredths of an inch, against the alternative that the mean is not 40.

The p-value for this two-tailed test is 0.2426. So you fail to reject the null hypothesis. You simply don't have evidence that the null is not true. This language is very specific, but also a bit unsatisfying. In a hypothesis test, you're trying to find evidence against the null. But you might be more interested in finding evidence that supports the null. In industrial settings, this is often the case. In the metal parts pilot scenario, what you actually want to know is whether the process is reasonably close to the target.

If your goal is to provide evidence that the deviation from the null hypothesis is not too great, then you would use an equivalence test. In equivalence testing, you specify a practically acceptable interval around the hypothesized value, and then test to see whether the sample mean is outside this interval.

To do this, you conduct two one-sided t tests. As a result, this is known as the TOST (two one-sided t tests) approach. First, you run a one-sided t test that the mean is the low value of the interval. For this test, the alternative is that the mean is greater than this value. Then you run a one-sided t test that the mean is the high value of the interval, using the alternative that the mean is less than this value. If you reject both of these tests, at some level of alpha, then you can conclude that the mean is "practically" equivalent to the hypothesized value.

We'll conduct an equivalence test for the metal parts scenario. Suppose that a difference of one hundredth of an inch from the target would be of practical importance to you. That is, if the true mean is one hundredth of an inch below or above the target, you consider the process to be on target.

We omit the technical details and focus on the interpretation and the implications. Here are the results for this equivalence test. The p-values for both tests are below a significance level of 0.05. We are able to reject both null hypotheses. That is, we accept both alternative hypotheses. The mean is both greater than 39 and less than 41.

The equivalence graph supports our conclusions. For this test, the 90% confidence interval for the mean thickness falls entirely within the equivalence region. That is, the estimated mean falls within the acceptable interval. From these test results, we can conclude that the mean thickness, based on the pilot data, is practically equivalent to 40. This provides the additional evidence your team needs to be confident that the piloted changes will bring the process to target.

To learn more about equivalence tests, see the Read About It for this module.