Exploring Sample Size and Power

To explore the relationship between sample size and power, we revisit the one-sample t test for the metal parts data.

Remember that the null hypothesis is that the process is on target, with a mean of 40. The p-value for this test is 0.2426. Based on this analysis, you fail to reject the hypothesis that the mean is 40. There is not enough evidence to reject the null hypothesis.

However, these results are based on only 20 observations. Would you have drawn different conclusions if you had more data?

Remember that the test statistic, the t ratio, is a ratio of the "signal," the difference between the observed and hypothesized values, and the "noise" in the data, the standard error. The t ratio for this test is 1.2. That is, the observed sample mean is 1.2 standard errors above the hypothesized value. You can see this graphically with the JMP PValue animation, which is available when you run a one-sample t test.

What if the same sample estimates were based on 50 observations rather than 20? The t ratio increases to 1.9, and the p-value for this test drops to 0.06.

If we further increase the sample size, this time to 100, the t ratio jumps to nearly 2.7. The test is now highly significant, even though the only thing that changed was the sample size.

Now, let's return to the original scenario. The sample mean, reported as the estimated mean, is 40.345, and the hypothesized value is 40. The difference between the means is 0.345 units. What if the difference between these two values was larger? If we increase the difference by just 0.5 units, by changing the hypothesized value to 39.5, the test is now highly significant.

As these results illustrate, by changing the sample size or by increasing the difference between the observed and hypothesized means, we are more likely to reject the null hypothesis.

Let's consider this from another perspective. The alternative hypothesis states that the population mean does not equal 40. There are infinitely many ways in which the null hypothesis can be false, so there are an infinite number of potential alternative sampling distributions.

This image was generated using the Power animation. This animation is also available when you run a one-sample t test in JMP.

The red curve represents the distribution of sample means under the null hypothesis. The blue curve is the sampling distribution under the hypothesis that the estimated mean represents the true mean. We ran a two-tailed test, so the red shading in the tails represents alpha, the significance level. The blue shading represents beta, the false negative rate.

Remember that the power of a test is 1 minus beta. The power for this test is only 0.21. That is, the probability of correctly rejecting the null hypothesis that the mean is 40, given a sample mean of 40.435, is only

0.21. But, for the same estimates, if we have 100 observations rather than 20, the power climbs to 0.76.

Likewise, if we have 20 observations, but our hypothesized mean is 39.5 rather than 40, the power is also much higher. It's close to 0.80.

What about the significance level, alpha? For each of these situations, the significance level is 0.05. With a sample size of 20 and a hypothesized value of 39.5, if we change alpha from 0.05 to 0.1, the power now increases to 0.89. It is much more likely that we'll detect this difference if we increase the significance level of the test.

To summarize, power is the ability to correctly reject a false null hypothesis. This is largely based on your sample size. However, as you have seen in this lesson, the power of your statistical test is also related to the critical difference that you need to detect, the variability in the population, and the significance level you choose.

In upcoming videos, you see how to calculate the sample size and power for one-sample t tests.

Close