Calculating the Sample Size for Two-Sample t Tests

In the previous video, you learned how to calculate the sample size for a one-sample t test. The same thinking applies to calculating the sample size for a two-sample t test.

For example, consider the Diameter Test scenario, which was introduced in the previous lesson. The characteristic of interest is the diameter of parts, in millimeters. Parts can be run on two different machines. You're concerned that the mean diameter for parts produced on the two machines might be different.

The estimate of the population standard deviation is 0.045 millimeters. You'd like a high probability of detecting a difference of 0.03 mm between the two machines. The total sample size required to detect this difference, with a power of 0.9, is 97. This is the total sample size between the two groups.

Because this is an odd number, we round this up. You need a total of 98 observations, or 49 parts from each machine, to have a high probability of detecting this difference.

Note that this calculation assumes that the sample size is the same for both machines. If the sample sizes for the two groups are not the same, meaning that you measure more parts from one machine than from the other, the actual power will be lower than 0.9.

Notice the trade-off between the sample size and the difference for a power of 0.9.

To detect differences below, say, 0.025, you need an extremely large sample size. If your critical difference is greater than, say, 0.035, you have a high probability of detecting this difference with much less data.

In the upcoming JMP demonstration videos, you see how to determine the sample size to detect critical differences for a two-sample t test. You also learn how to determine the sample size to detect a specified difference between any two means for one-way analysis of variance, or ANOVA.

Remember that ANOVA is used when you want to test the null hypothesis that the means for all groups are equal against the alternative hypothesis that at least two means are different.

Statistical Thinking for Industrial Problem Solving

Close