Sampling Distribution under the Null

To explore the fundamental logic of statistical testing, we revisit the idea of sampling error. In the metal parts scenario, the null hypothesis is that the mean is 40, and we assume that the null hypothesis is true.

In hypothesis testing, you make decisions based on one sample. But, remember that different samples will have slightly different means. You have to make a decision about the population in light of this sampling variation, using the one sample you have. If we assume that the null hypothesis is true, what would the distribution of potential sample means look like?

Here is the distribution of 10,000 sample means for the metal parts scenario, simulated under the null hypothesis (that is, assuming that the null hypothesis is true). Each mean was simulated using a population mean of 40, a standard deviation of 1.5, and a sample size of 20.

In this scenario, remember that your improvement team is looking for warning signs that the mean thickness is not 40, based on a pilot sample of 20 parts. However, when the population mean truly is 40, you can see how much the sample means vary. We rarely draw a sample with a mean of exactly 40. The distribution of sample means is centered at 40, which is reassuring. But, even when the null is true, it's common to obtain a random sample in which the mean thickness is below or above 40.

Fortunately, we can characterize this sampling variation and, we can judge the likelihood of getting different sample means. Means like these would be common, and hence unsurprising if the mean thickness truly were 40. Means like this would be possible, but they'd be unusual. And a mean like this, or like this, would be incredible!

A sample mean of 42, for this scenario, is well beyond the bounds of ordinary variation. It is so extreme that it challenges our belief that the null hypothesis describes reality. When we say that a sample mean like this would be "incredible," we are getting to the heart of statistical thinking.

If we believe that the null hypothesis is true, a sample mean of 42, based on 20 observations, almost certainly would not occur. Hence, in the face of a sample mean of this size, what's really incredible, or not believable, is the null hypothesis itself. That is, if we were to observe a sample mean of 42 from this sample, we would doubt the validity of the null hypothesis.

You can see that the sample mean of 40.345, which we observed for the pilot data, is consistent with ordinary sampling variation. That is, the sample mean is consistent with the null hypothesis that the true mean is 40.

Let's return to the idea of characterizing the variability in sample means. This is possible with the help of the Central Limit Theorem, which you learned about in a previous module.

Under a range of conditions, as the sample size n increases, the distribution of the sample mean approaches a normal distribution with the mean of mu and a standard deviation of sigma over the square root of n. Why is this a big deal? Although we can't be certain in advance about an individual sample mean, we have a good idea how close it will be to the true population mean based on ordinary sampling variation.

Means within the range of ordinary sampling variation are consistent with the null hypothesis. But, if a sample mean is extreme relative to the distribution of sample means, we really start to question whether the null hypothesis is correct.

Statistical Thinking for Industrial Problem Solving

Close