Calculating Capability for Nonnormal Data

You've learned how to conduct a capability study for data that are normally distributed.

For normally distributed data, you know that 99.73% of the observations will fall within plus or minus 3 standard deviations of the mean. This 6-standard-deviation window defines process spread used to calculate the capability indices.

Approximately 0.135 percent of the observations will be more than 3 standard deviations below the mean, and approximately 0.135 percent will fall above the mean.

What if your distribution is not normal?

If you use the normal distribution to estimate the percent of measurements out of spec for a nonnormal distribution, you will over- or under-estimate the capability of the process.

Take this example. The measurement for this characteristic has a lower bound of zero. When you fit a normal distribution to these data, the normal curve extends below zero. So if you estimate the capability using the normal distribution, you would estimate that you'd have many nonconformances below the lower bound of the data.

In fact, if you assume a normal distribution and conduct a capability analysis with a lower spec of 0, you see that an estimated 15% of the measurements will fall below this spec, even though the lowest value possible is zero.

Instead of estimating capability using the normal distribution, you can estimate the capability using the distribution that best fits the data. Let's look at an example.

The Impurity scenario involves the production of a polymer.

A catalyst is required for the chemical reactions to occur to produce the polymer.

The catalyst contains a chemical that can create an impurity in the polymer. The lower the impurity the better, because lower impurity means higher yield.

The target is 3%, and the upper spec is 7%.

There is no lower spec, but the natural lower bound for Impurity is zero.

The data appear to be right skewed. The curvature in the normal quantile plot confirms that the underlying distribution is not normal.

In fact, the data appear to follow a lognormal distribution. The lognormal distribution is often constrained by zero and is right skewed, so it makes sense here. Note that you learn how to select the best distribution in the next video.

When you fit a lognormal distribution to the data, the curve seems to fit the data well.

Here is a lognormal probability plot. You interpret this plot like you interpret the normal quantile (or probability) plot. The data more or less follow a straight line, with no obvious pattern. So it appears that the lognormal distribution fits the data well.

Let's conduct a capability study using the lognormal distribution instead of the normal distribution.

Instead of estimating capability based on the percentiles of the normal distribution, the indices will be based on the percentiles of the lognormal distribution. These percentiles (or quantiles) define the process spread for the lognormal distribution.

The same capability indices can now be computed based on this process spread. We don't have a lower spec limit for this example, so C_p isn't computed. The C_{pk} is only 0.164. This reflects the fact that the process mean is shifted towards the upper spec limit. The estimated percent out of spec is 25.18. Clearly this is a process in need of improvement!

In the next videos, you learn how to identify the underlying distribution and see how to estimate capability for nonnormal distributions.

Close