

Demo: Transforming Data

In this video, you learn two methods for transforming data: using virtual columns in analyses and using New Formula Column from the data table.

For this video, we use the file Queue Time.jmp. This file includes information about queue times for 100 batches of parts in a machining operation.

We'll start by graphing the data. To do this, we select Graph Builder from the Graph Menu.

We drag Queue Times to the X zone and click the Histogram icon.

You can see that these data are highly skewed. Queue time data often follow a lognormal distribution. When you apply a log transformation to lognormal data, the distribution of the transformed data is approximately normal.

Let's apply a log transformation to these data. To do this, we right-click the variable Queue Times in the column selection panel, select Transform, and then select Log.

This creates a virtual column, Log(Queue Times), which doesn't exist in the data table.

When we create a distribution of this new variable by dragging it to the X zone, we see that the distribution is indeed approximately normal.

Let's add this column to the data table. To do this, we right-click Log(Queue Times) in the column selection panel, and select Add to Data Table. When we look at the data table, we see this new column.

The formula in the column applies a log transformation to the data.

The log is just one of many transformations available under the Transcendental function group.

If you know that you want to transform a variable or create a derived variable, you can create a new formula column directly from the data table.

To create a log transformation of Queue Times from the data table, you right-click the column head, select New Formula Column, Transform, and Log.

New Formula Column is an efficient way to transform data. You can also use it to create formulas that apply a variety of different functions. The types of functions available depend on the type of data and the number of variables you select. For categorical data, you can easily apply many character functions. If you have date or time data, you can apply a number of date and time functions.

Close