Print

Deriving New Variables

Sometimes, the information you need is in your data table, but you have to do a bit of work to get to it.

In the component scenario, you used the formula editor to create a new variable, scrap rate.

Scrap rate is a derived variable, because it was computed using data in the data table.

Earlier, you saw examples of derived variables in the Measles scenario. A variable was created to bin the data in the periods before and after the measles vaccine program was implemented.

This formula was created using a simple IF-THEN statement. Another formula was used to create the variable Vaccine. For this variable, the data are binned in the time periods Before, Transition, and After.

This variable made it easy to compare the measles rates for the three time periods.

Another way of creating a derived variable is to apply a transformation.

For example, consider the population (in thousands) for countries in 2017.

The data are so skewed that it is difficult to learn much about the distribution from a histogram alone. There are two countries with very large populations (India and China), and the rest of the countries are lumped together close to zero. You can't really see the distribution.

If you want to understand the population geographically, you don't learn much.

A data transformation can make it easier to see the patterns in the data.

For example, when you apply a natural log transformation to Population, you see that the distribution of the log of Population is much less skewed than the original data.

Applying data transformations is a deep topic that is, for the most part, beyond the scope of this course. For more information about applying data transformations, including reasons why it might be beneficial to transform your data, see the Read About It for this module.

Formulas can also be used to extract information out of a column in the data table.

For example, the variable Lot includes a lot of embedded information. It includes

the company name (JMP or SAS) the part number, and the production date.

You explore the formula used to extract the part number in a practice.

Statistical Thinking for Industrial Problem Solving

Close