

Restructuring Data for Analysis

For most statistical analyses, your data set must have a data grid with the observations stored in rows and the variables stored in columns.

For example, let's suppose that you want to compare the average number of particles for three different polymers.

You have particle data on five samples per polymer, and your data for the polymers are stored in separate columns in a data table. This is known as a wide, or split, data table format.

In order to analyze these data, you need to stack the data so that the polymer labels appear in one column and the particles data are in a separate column. This is called a tall, or stacked, format.

Now you can use the methods you've already learned to compare particles for the different polymers.

Although most statistical analyses require data to be in a stacked format, if your data consist of paired readings or repeated measurements, you might need to split your data into separate columns.

For example, let's suppose that you want to compare the yield for two processes. You have one yield measurement for each process every month.

If you want to compare the yield over time for the two processes, you can graph the data in this stacked form.

But if you want to calculate the difference in the yield between these two processes each month, the data need to be in a wide format.

That is, you need to split the process data into two separate columns.

Now, for example, you can compute the differences between the yield values for the two processes each month, and you can directly graph or analyze these differences.

In these graphs, notice that the months are sorted in the correct order: January, February, and so on.

This is because there is a Value Ordering column property assigned for Month in this data table.

The value ordering column property tells JMP the order in which the months should appear in graphs and analyses.

Note that, in this example, the labeling of the months is a little inconsistent. Some months are abbreviated and some aren't.

If JMP recognizes the values as months, the values are automatically plotted in the correct order. If you have the full month names, use month numbers, or use the same three-character abbreviations, JMP plots the values in the correct order without requiring value ordering.

In this video, you learned about stacking and splitting your data. There are other ways you might want to reshape your data for analysis.

For example, you might want to create a new data table from a subset of your data, sort your data, or transpose the rows and columns in your data table.

These and other tools for reshaping your data table for analysis are available from the Tables menu in JMP. To learn more, search for reshape in the JMP Help files (at www.jmp.com/help).

Statistical Thinking for Industrial Problem Solving

Copyright © 2020 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close