Demo: Creating Tabular Summaries for Categorical Data

In this video, we show how to create tabular summaries of categorical data using Tabulate and Fit Y by X from the Analyze menu. We use the Chemical Manufacturing data, which has a categorical variable, Performance, with two values, Accept and Reject. Rejected batches did not have a yield greater than 80%.

We'll use tabular summaries to compare the frequencies (or counts) for accepted and rejected batches for three base particle sizes: small, medium, and large.

First, we select Tabulate from the Analyze menu.

We drag Performance to the Drop zone for columns. The results zone shows the counts for fails and passes. Next, we drag Base Particle Size to the Drop zone for Rows. The counts are now broken down by Base Particle Size. You can see that the counts are different for accepted and rejected batches across the three sizes.

We drag N and Column % from the statistics panel to the results zone. You now see both the counts and the percent of the accepted and rejected batches that fell within each size. For example, of the rejected batches, 35.7% were large, 14.3% were medium, and 50% were small.

Instead of showing Column %, you might be interested in Row %. We drag Row % on top of the Column % label. Now the data for each size is broken into Accept and Reject. Of the batches with small particle size, 75% were accepted and 25% were rejected.

Instead of using Tabulate, you can create a compact summary called a contingency table.

To create a contingency table in JMP, we select Fit Y by X from the Analyze menu.

We drag Performance to Y, Response and Base Particle Size to X, Factor. The graph at the top, the mosaic plot, summarizes the frequencies for levels of the two variables. Mosaic plots are described in the next video.

Let's focus on the results in the contingency table. This table reports the counts (or N), Column Percent, and Row Percent. It also reports the cumulative totals for both variables, the percent for each cell out of the total, and the percent for each level of a variable out of the total.

In this example, it appears that there might be a difference in performance for the different particle sizes.

Statistical Thinking for Industrial Problem Solving

Close