

## Demo: Summarizing Continuous Data with Column Viewer and Tabulate

In this video, we show how to summarize the continuous variables in the Impurity data using two platforms, Columns Viewer and Tabulate. You see how to summarize categorical variables in future videos.

We begin with Columns Viewer, which can be used to summarize many variables at a time.

To start, we select Columns Viewer from the Cols menu. We select the four continuous variables, from Impurity through Reaction Time. We select the box for Show Quartiles and click Show Summary.

For each variable, the minimum and maximum values are displayed, along with the mean, the median, and other statistics. If values are missing from any variable, you also see a column named N Missing. In this example, none of the four variables have missing values.

While in Columns Viewer, if you're interested in seeing the distributions of these variables, you can select the variables and click Distribution. This produces distribution output for the selected variables.

We'll close the Distribution output, and close the Columns Viewer output, and switch gears.

Let's take a look at the Tabulate platform.

Tabulate provides a flexible interface for creating tables of summary statistics. Here, we focus on summarizing the four continuous variables in the Impurity data set.

First, we select Tabulate from the Analyze menu. You can drag and drop variables and statistics to the different drop zones. There is a drop zone for columns and a drop zone for rows. The resulting cells panel is also a drop zone.

Here, we want to summarize all of the continuous variables. From the columns list, we select Impurity through Reaction Time and drag them to the Drop zone for columns.

The default statistic, which displays in the resulting cells panel, is the sum. This is simply the sum of all the values in each column, which isn't overly interesting here.

We'd like to calculate the mean for each of the variables instead of the sum. So we drag Mean from the statistics panel and drop it on top of Sum for any of the variables. By dropping the mean on top of the sum, we are telling JMP to replace the sum with the mean.

Note that Tabulate is very forgiving. If you make a mistake, or if want to make a change, you can use the Undo button in the control panel. Or you can start over entirely.

Let's say that we want to change the position of the label for the mean. We click Undo, and this time we drag Mean to the Drop zone for rows.

Let's add some other statistics. We select Min, Max, and Median, and drop them on top of the results panel. These statistics are added, in addition to the mean.

Next, we add a quantile. Drag and drop Quantiles to the resulting cells panel. We add a 90% quantile (or percentile). This quantile is now added, in addition to the other statistics.

To remove a statistic, we right-click the statistic name and select Delete.

If we want to change the number of decimal places that are displayed for the different statistics, we can use the Change Format option. We'll use the default formats here.

When we're finished building the table, we click the Done button to close the control panel.

Several options are available under the Tabulate red triangle. For example, we can re-open the control panel, or we can save the results to a data table.

---

*Statistical Thinking for Industrial Problem Solving*

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

Close