

## Demo: Summarizing Continuous Data with the Distribution Platform

In this video, we show how to create box plots and summary statistics using the Distribution platform in JMP for the Impurity data.

To start, we select Distribution from the Analyze menu.

We select Impurity for Y, Columns, and click OK. We select Stack from the top red triangle next to Distributions to change the layout to horizontal.

By default, you see a histogram and box plot, quantiles, and summary statistics for Impurity.

Histograms are described in another demonstration, so we'll focus on the other output that is provided.

Let's start with the box plot.

Recall that the lower end of the box in the box plot is the first quartile, the line in the box plot is the median, and the upper end of the box is the third quartile. The distance between the first and third quartiles is called the interquartile range, or IQR, and the lines drawn from either end of the box are called whiskers.

Two other pieces of information displayed with the outlier box plot are the sample mean and a measure called the shortest half.

The center of the diamond is the sample mean. In this example, we see that the mean is slightly higher than the median. This is an indication that the distribution is somewhat skewed.

The tips of the diamond define a confidence interval for the mean. You learn about confidence intervals in the Decision Making with Data module.

The shortest half shows the densest region of the data. This shows where the "tightest" grouping of 50% of the observations fall. Notice that the shortest half corresponds to the tallest bars in the histogram.

Let's take a look at the default summary statistics that are reported.

The Quantiles report includes the minimum, maximum, median, quartiles, and quantiles (or percentiles) in different increments.

We can right-click on the values in this table and select Format Column to change the data format for the values that are displayed. For example, we might want to show only three decimal places for the quantiles, and also for the summary statistics. It might make sense to do this if we have only a few significant digits.

We can also change the default quantile increments displayed in the Quantiles report or request custom quantiles using red triangle options. For example, to set quantile increments, we select Display options from the red triangle next to Impurity, and then enter the quantile increment value. To display quantiles in increments of 10%, we enter 0.1.

We'll repeat these steps and select Revert to default quantiles to show the original quantile values.

The Summary Statistics table reports the mean, along with several other measures.

We can add other summary statistics to this report. To add statistics, click the red triangle next to Summary Statistics and select Customize Summary Statistics.

A variety of measures for shape, centering, and spread of the distribution are available.

Here, as an example, we select N Missing and click OK. We can see that none of the values for Impurity are missing.

Note that, if we prefer different percentile increments each time we run the analysis, or if we'd like different summary statistics to display by default, we can set preferences. To do this, we go to File then Preferences (or JMP then Preferences on a Mac). We select Platforms from the Preference Group and select Distribution or Distribution Summary Statistics from the Platforms list.

---

*Statistical Thinking for Industrial Problem Solving*

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

Close