

# JMP® Introductory Lab Activities

## Activity 2: Describing Categorical Data



**Data Set:** Denim.jmp

### Summary

In this lab you will open data from the JMP Sample Data directory and will graph and summarize data in JMP using **Analyze > Distribution**.

You'll create a report with graphs and numerical summaries from JMP, along with your commentary and interpretation (required output and discussion is in italics).

### Using Data in the Sample Data Directory

In JMP, go to **Help > Sample Data**. This is the JMP Sample Data directory. Select **See an Alphabetical List of all Sample Data Files**, and open the data file **Denim.jmp**.

**Denim** is a data set for a fictitious company that manufactures jeans. The denim fabric they use is treated to make it appear “worn.” Although appearance of the fabric is important, the company is also concerned about the strength of the treated fabric.

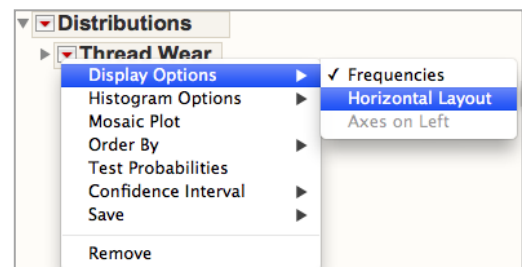
**Thread Wear Measured** rates the wear of the fabric on a numeric scale from 1 to 10. The variable **Thread Wear** converts the numerical values into categories, with 1 to 4 being Low, 5 to 9 as Moderate and 10 as Severe. The variable **Thread Wear** is an ordinal variable – Low, Moderate and Severe are categories with a natural order. If you display the distribution of a categorical variable (either ordinal or nominal), the displays offered will be different from the displays for a continuous variable.

### Graphing and Summarizing Categorical Data

We'll use **Analyze > Distribution** to display the distribution of **Thread Wear**. This option opens a dialog window that gives you the opportunity to specify which variables to analyze. Select **Thread Wear** for the **Y, Columns** variable. Then click **OK**.


There are many useful options under the red triangle next to **Thread Wear**. Use the red triangle to make the following changes to your display:

- To change the graph from a vertical to a horizontal layout, select **Display Options > Horizontal Layout**.
- To display a count axis, use **Histogram Options > Count Axis**.
- Add a mosaic plot (which may be unfamiliar to you) by selecting **Mosaic Plot**.



Create a new report including:

- The bar graph with a count axis and a mosaic plot for **Thread Wear**, and
- The Frequencies table, which has the counts and proportions in each category.

Use the **selection** tool  on your toolbar to select (highlight) the output to be copied. (Note: In Windows, if the toolbar does not appear, click the Alt key to display it.)

Hints: To extend a selection, hold the shift key. Or, click near the edge of the report to select all of the content in a window.

**Copy** the graph. Open the program where you'd like to paste the content, and select **Paste**. Add titles as needed.

*Describe the distribution of the fabric thread wear trial results, referring to specific aspects of the graph or numerical displays to support your answer.*

*Remember to write your description in the context of the situation. Also include a description of what the mosaic plot shows.*

## Graphing and Summarizing Continuous Data

The variable **Thread Wear Measured** contains the original results from the fabric thread wear tests that were measured on a scale from 1 to 10.

Analyze the distribution of **Thread Wear Measured** using **Analyze > Distribution**. Select **Thread Wear Measured** for the **Y, Columns** variable and click **OK**. The output window will display a histogram, a boxplot and some numerical summaries.

(Note: You'll formally explore continuous data in **Activity 3, Describing Numerical Data**.)

*Copy the resulting displays into your report. Describe the shape of the distribution of the fabric thread wear as originally measured. Use appropriate graphical displays and numerical summaries to support your answer.*

## Discussion

*Briefly discuss the differences between the ordinal version of the results (**Thread Wear**) and the continuous (numerical) version of the results (**Thread Wear Measured**).*

*Consider the categories of **Thread Wear**. Might it have been better to use different categorical divisions to create Thread Wear, such as 1 to 3, 4 to 7, and 8 to 10? If so, why?*