

## JMP INTRO® Lab Activities

### Lab Activity – Exploring Categorical Data

Data Set: Denim

In this lab we'll look at some of the techniques used to make judgments about data in which there is an association based on categorical variables.

Denim is a data set from an experiment done by a company that manufactures blue jeans. Denim fabric naturally contains starch, creating stiffness in the fabric. Most customers find this stiffness uncomfortable so denim manufacturers subject the fabric to a variety of washing treatments to remove some of the starch and make it feel “worn”. Although the feel of the fabric is important, the company is also concerned about the strength of the treated fabric, as measured by a count of the destroyed threads.

Open the data file **DENIM** from the Sample Data folder. This data contains information on 98 samples, listing for each sample the lot number of the fabric, the method used to treat the fabric, the size of the load washed (in lbs.), whether or not the fabric was sand blasted, the thread wear (an actual count of broken threads between 1 and 10), the thread wear categorized as low, moderate or severe, and the resulting starch content (%). The size of the wash load, the thread wear measured and the starch content are numerical variables (called continuous by JMP INTRO). The remaining variables are all categorical (nominal or ordinal).

This lab will be a study of which, if any, of the categorical factors is related to the severity of the thread wear of the denim.

Is the method of aging the denim independent of thread wear severity? What do you think? Write a brief response to this question in your report. Include your reasons for believing that an association might or might not be present.

Check in the Data Table Panel to the left of the Data Table to make sure that Method is shown as a nominal variable and Thread Wear is shown as an ordinal variable. Click on the variable name and change its type if necessary. Open **Analyze**® **Fit Y by X**. Select **Method** for the **X, Factor** and **Thread Wear** (not Thread Wear Measured) for the **Y, Factor**.

For numerical data, this command displays a scatterplot, but this time it shows something different: a mosaic plot and a two-way table that JMP INTRO calls a **Contingency Table**. (If your table shows more than the counts in each cell, make your table easier to read by using the red triangle next to the heading **Contingency Table** to hide all values except the counts.)

By looking at either the graphs or the table, does this analysis confirm or refute your original belief about an association between the method of aging and the severity of the

thread wear? Copy the graphs and contingency table into your report and write a brief response to this question. Include your reasons for believing that an association might or might not be present based on the graphs and/or table.

One way to assess this is to calculate percentages, either in individual rows or individual columns, for the entries in the table. Use the red triangle next to the heading **Contingency Table** to show **column percentages**. Copy this contingency table into your report. Now hide the column percentages and show **row percentages**. Copy this new contingency table into your report. Are row or column percentages reflected in the mosaic plot? Do either row or column percentages suggest dependence or independence of **method** and **thread wear**? Summarize your observations based on these percentages in your report.

How about other factors among the categorical data? Repeat the steps above using the variable **Sand Blasted?** as the **X,Factor**. Were your ideas about independence of **sand blasted** and **thread wear** confirmed or refuted? Include graphs and some numerical information to support your answer. Remember to report your answers in the context of the problem.

Print a copy of your lab report. Proofread your copy, make corrections as needed, close JMP INTRO and the word processing program, and log off.

## JMP INTRO® Lab Activities

### Teacher Notes

#### Lab Activity – Exploring Categorical Data

##### Objectives:

- Use JMP INTRO to display a mosaic plot and contingency table for categorical data. Row and column percentages will also be found.
- Create a word processing document, incorporating graphs and tables from JMP INTRO.

Time Required: 45 minutes

##### Materials:

- Exploring Categorical Data student activity directions
- Denim data set

##### Prerequisites:

- Students should have experience producing numerical summaries and graphical displays of categorical data by hand and/or with graphing calculators.
- Students should have basic knowledge of how to use JMP INTRO.

##### JMPINTRO® Notes:

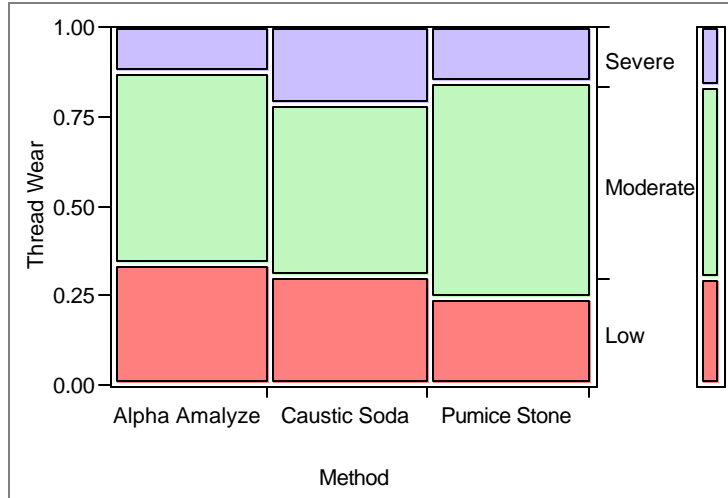
- The **Denim** data file is included in the Sample Data folder as part of JMP INTRO.
- The same command (**Analyze** ® **Fit Y by X**) is used for both numeric and categorical data. JMP INTRO decides which analysis to use based on the types of variables selected as factors.
- With bivariate categorical data, the mosaic plot that JMP INTRO displays includes a bar on the far right that represents all of the data.

## JMP INTRO® Lab Activities Answer Key

The following are suggested answers for Exploring Categorical Data:

### Contingency Analysis of Thread Wear By Method

#### Mosaic Plot



#### Contingency Table

Method By Thread Wear

Count	Low	Moderate	Severe	
<b>Alpha Analyze</b>	11	17	4	32
<b>Caustic Soda</b>	10	16	7	33
<b>Pumice Stone</b>	8	20	5	33
	29	53	16	98

Method By Thread Wear

Count	Low	Moderate	Severe	
<b>Alpha Analyze</b>	11	17	4	32
	37.93	32.08	25.00	
<b>Caustic Soda</b>	10	16	7	33
	34.48	30.19	43.75	
<b>Pumice Stone</b>	8	20	5	33
	27.59	37.74	31.25	
	29	53	16	98

Method By Thread Wear

Count	Low	Moderate	Severe	
<b>Alpha Analyze</b>	11	17	4	32
	34.38	53.13	12.50	
<b>Caustic Soda</b>	10	16	7	33
	30.30	48.48	21.21	
<b>Pumice Stone</b>	8	20	5	33
	24.24	60.61	15.15	
	29	53	16	98

Before looking at any analysis, students should guess whether thread wear and method are related or not. They should include a reasonable justification for their response.

The mosaic plots show the same information as the row percentages. The column percentages are fairly evenly spread for the low and moderate columns. In the severe column, caustic soda has a higher percentage than the other two methods. The row percentages show more differences. With all three methods, there are considerably more trial results in the moderate category than the other two – with two of the three containing more than half the trial results.

### Contingency Analysis of Thread Wear By Sand blasted?

**Mosaic Plot**



**Contingency Table**

Sand blasted? By Thread Wear

Count	Low	Moderate	Severe	
<b>no</b>	16	28	5	49
<b>yes</b>	13	25	11	49
	29	53	16	98

Sand blasted? By Thread Wear

Count Col %	Low	Moderate	Severe	
<b>no</b>	16 55.17	28 52.83	5 31.25	49
<b>yes</b>	13 44.83	25 47.17	11 68.75	49
	29	53	16	98

Sand blasted? By Thread Wear

Count Row %	Low	Moderate	Severe	
<b>no</b>	16 32.65	28 57.14	5 10.20	49
<b>yes</b>	13 26.53	25 51.02	11 22.45	49
	29	53	16	98

Before looking at any analysis, students should guess whether thread wear and method are related or not. They should include a reasonable justification for their response.

The mosaic plots show the same information as the row percentages. The column percentages for sandblasted appears to show some differences. They are similar for the low and moderate columns, with about 45% of the low and moderate thread wear results coming from fabrics that have been sandblasted.. In the severe column, the sandblasted fabric has a much higher percentage of severe thread wear than the non-sandblasted fabric. The row percentages show differences similar to those in the first analysis. With both sandblasted and non-sandblasted denim, there are considerably more trial results in the moderate category that the other two – with two of the three containing more than half the trial results.

