

JMP INTRO® Lab Activities

Lab Activity – X^2 Test of Independence

Data Set: Denim

As you may remember from the Exploring Categorical Data activity, 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. Since customers often don't like the stiff feel of the fabric, they wash the denim to make it feel "worn". Although the feel of the fabric is important, the company is also concerned about the strength of the treated fabric.

Open the data file **DENIM** from the Sample Data folder. The size of the load, the thread wear measured and the starch content are continuous numerical variables. The remaining variables are all categorical (nominal or ordinal).

In the Exploring Categorical Data activity, you looked at pairs of categorical variables and calculated row and column percentages. An additional feature of the percentage calculations is that it helps you determine what values would be expected in the table if there were no association between the categorical variables. These are values that would make the percentages in each row exactly the same for every column.

Recreate the your analyses from the Exploring Categorical Data activity by opening **Analyze** ® **Fit Y by X** and select **Method** for the **X, Factor** and **Thread Wear** for the **Y, Factor**. Also open **Analyze** ® **Fit Y by X** and select **Sand Blasted?** for the **X, Factor** and **Thread Wear** for the **Y, Factor**.

Contingency Table

Method By Thread Wear

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

Col. % of Total 29.59

Expected counts can be thought of in the following way: 29 of the 98 trials were "Low". The proportion of the total results that were "Low" is thus 29/98 or .2959. If there were no difference between the results produced by the different methods, you would expect about 29.59% of each of the methods to produce "Low" results. Therefore, the expected cell counts for the three methods are:

$$\text{Alpha Analyze - Low} = .2959 (32) = 9.4688$$

$$\text{Caustic Soda - Low} = .2959 (33) = 9.7647$$

$$\text{Pumice Stone - Low} = .2959 (33) = 9.7647$$

Figure the remaining expected values for this first table and show your calculations in your report.

In the red triangle pop-up menu, there is an option for the expected values. Add these values to each of your contingency tables (if necessary, hide any percentages to make your tables easier to read). Copy your graphs and contingency tables into your report. Did JMP INTRO give you the same results as figuring the expected counts by hand? To test for association, all of the expected counts must be 5 or more. Was this requirement (assumption) met in both of your analyses?

In comparisons of numerical data, the correlation coefficient measures how strongly two variables are associated. To evaluate this association for categorical data is somewhat similar, in that it includes a calculation of the sum of squares.

For each cell in the **Contingency Table**, you should have the observed value along with the expected value (if there were no association). If these values differ greatly, as a proportion of what was expected, it is evidence of a strong association.

For each cell: 1) subtract the observed value – the expected value
2) square this difference
3) divide by the expected value.

Then find the sum of all the cells' values. This is called the **chi-square statistic**, with formal notation X^2 (the Greek letter chi).

This is the formula:
$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Use the expected counts from the Sand Blasted-Thread Wear contingency table to calculate the chi-square test statistic by hand. Include your calculation in your report. Confirm your work by copying the Tests box in the JMP INTRO display into your report. Include only the Pearson Chi-Square Test results. Did these results agree with the values you calculated by hand?

To be statistically significant at the 0.05 level, you need to have a p-value or probability less than 0.05. What were your p-values? Were any of your results statistically significant? Remember to report your answers in the context of the problem.

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

JMP INTRO® Lab Activities

Teacher Notes

Lab Activity – X^2 Test of Independence

Objectives:

- Use JMP INTRO to figure expected values in a contingency table and run a chi-square test of independence.
- Create a word processing document, incorporating graphs and tables from JMP INTRO.

Time Required: 45 minutes

Materials:

- X^2 Test of Independence 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 JMP INTRO.
- Students should have completed Exploring Categorical Data (X^2 Goodness-of-fit Test can be done independently of X^2 Test of Independence.)

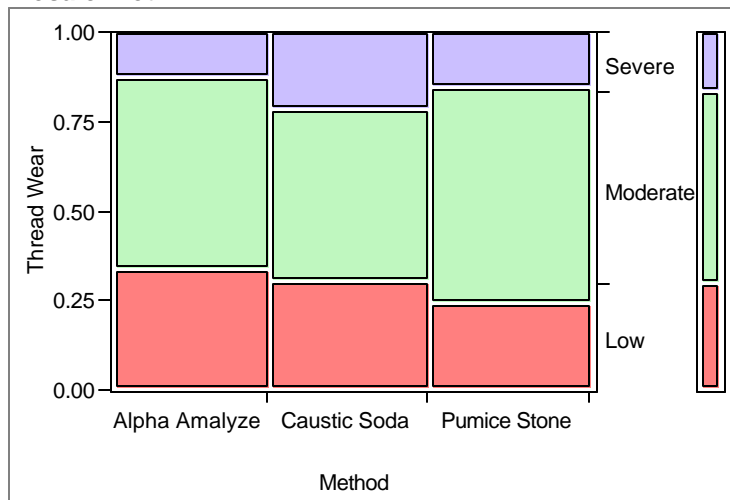
JMPINTRO Notes:

- The **Denim** data file is included in the Sample Data folder as part of JMP INTRO.
- The chi-square statistic that is calculated by the formula shown in the student directions and found in introductory textbooks is the Pearson test statistic.
- There are no categorical data sets provided in JMP INTRO that have significant chi-square test of independence relationships. You might want to add or substitute alternative data sets that are not independent. Additional data sets are available for download from the JMP website: www.jmpdiscovery.com.

JMP INTRO® Lab Activities Answer Key

Suggested answers for X^2 Test of Independence

Contingency Analysis of Thread Wear By Method Mosaic Plot



Contingency Table

Method By Thread Wear

Count Expected	Low	Moderate	Severe	
Alpha Analyze	11 9.46939	17 17.3061	4 5.22449	32
Caustic Soda	10 9.76531	16 17.8469	7 5.38776	33
Pumice Stone	8 9.76531	20 17.8469	5 5.38776	33
	29	53	16	98

Tests

Test	ChiSquare	Prob>ChiSq
Pearson	1.826	0.7678

Expected counts:

- alpha analyze – moderate = $0.5408 (32) = 17.3056$
- caustic soda – moderate = $0.5408 (33) = 17.8464$
- pumice stone – moderate = $0.5408 (33) = 17.8464$
- alpha analyze – severe = $0.1633 (32) = 5.2256$
- caustic soda – severe = $0.1633 (33) = 5.3889$
- pumice stone – severe = $0.1633 (33) = 5.3889$

The expected counts are essentially the same with differences due to rounding. All expected counts are greater than 5 so we can proceed with the chi-square test.

The p-value is 0.7678. This is not significant. The probability of getting results such as we got by chance alone is 0.7678. There is no evidence that the method to age the denim fabric and the thread wear are not independent.

Contingency Analysis of Thread Wear By Sand blasted?

Mosaic Plot



Contingency Table

Sand blasted? By Thread Wear

Count	Low	Moderate	Severe	
Expected				
no	16 14.5	28 26.5	5 8	49
yes	13 14.5	25 26.5	11 8	49
	29	53	16	98

Tests

Test	ChiSquare	Prob>ChiSq
Pearson	2.730	0.2554

All expected counts are at least 5. It is OK to continue with the chi-square test.

$$\begin{aligned}
 X^2 &= \frac{(15-14.5)^2}{14.5} + \frac{(28-26.5)^2}{26.5} + \frac{(5-8)^2}{8} + \frac{(13-14.5)^2}{14.5} + \frac{(25-26.5)^2}{26.5} + \frac{(11-8)^2}{8} \\
 &= .15517 + .08491 + 1.125 + .15517 + .08491 + 1.125 \\
 &= 2.73016
 \end{aligned}$$

JMP INTRO shows a chi-square test statistic of 2.730. This is essentially the same with differences due to rounding.

The p-value is 0.2554 which is not significant. There is no evidence that the method of sandblasting the fabric and thread wear are not independent.