JMP® Introductory Lab Activities

Activity 17: χ^2 Test of Independence

Data Set: Denim.jmp

Summary

We return to the **Denim.jmp** data, first presented and described in the **Exploring Categorical Data** activity.

In this lab, you'll calculate the χ^2 statistic for a two-way table using observed and expected counts, and will compare your results to those reported in JMP. You'll summarize your analysis and conclusions in a report (required output and discussion is in italics).

The Denim Data

Open the file **Denim.jmp** from the **Sample Data** directory. Recall that this data contains information on 98 samples of denim.

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 <u>expected</u> in the table if there were <u>no association</u> between the categorical variables. These are values that would make the percentages in each row exactly the same for every column.

Analyzing Categorical Data in JMP®

Recreate your analysis from the **Exploring Categorical Data** activity. Use **Analyze > Fit Y by X**. Select both **Method** and **Sand blasted?** for the **X, Factor** and **Thread Wear** for the **Y, Response**.

The contingency table for **Method** and **Thread Wear**, showing only the counts, is below (use the **red triangle** to remove the other values):

Contingency Table

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

Calculating Expected Values

As we saw in the activity χ^2 Goodness-of-Fit Test, expected counts are used to calculate the χ^2 test statistic. 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 0.2959.
- 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:

Alpha Amalyze - Low =	0.2959 (32) = 9.4688
Caustic Soda - Low =	0.2959 (33) = 9.7647
Pumice Stone - Low =	0.2959 (33) = 9.7647

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

Conducting a χ^2 Test for Independence in JMP®

Click on the **red triangle** next to **Contingency Table** in the two reports to display the expected values. Hint: Hold the control key (or if using a Mac, the command key) on your keyboard first to apply this selection to both reports – if you’re using a Mac, hold the command key.

Copy your graphs and contingency tables into your report. Did JMP give you the same results as calculating 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 (the value you would get 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, a chi-square value can be calculated as follows:

$$\frac{\left(\text{observed} - \text{expected}\right)^2}{\text{expected}}$$

Cell chi-square values are also available under the red triangle in the contingency table.

Contingency Table

Method	Count Expected Deviation Cell Chi^2	Thread Wear: Low	Moderate	Severe	
Method	Alpha Amalyze	11 9.46939 1.53061 0.2474	17 17.3061 -0.3061 0.0054	4 5.22449 -1.2245 0.2870	32
Method	Caustic Soda	10 9.76531 0.23469 0.0056	16 17.8469 -1.8469 0.1911	7 5.38776 1.61224 0.4825	33
Method	Pumice Stone	8 9.76531 -1.7653 0.3191	20 17.8469 2.15306 0.2597	5 5.38776 -0.3878 0.0279	33
		29	53	16	98

The sum of all cell chi-square values is the **chi-square statistic**, with formal notation χ^2 (χ is the Greek letter chi).

$$\chi^2 = \sum \frac{\left(\text{observed} - \text{expected}\right)^2}{\text{expected}}$$

This value is reported under **Tests** as the **Pearson ChiSquare** statistic, along with a p-value:

Tests

N	DF	-LogLike	RSquare (U)
98	4	0.90795611	0.0094

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	1.816	0.7696
Pearson	1.826	0.7678

Activity: Conducting a χ^2 Test for Independence

Use the expected counts from the ***Sand Blasted?*** *and* ***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 display into your report. Include only the Pearson Chi-Square Test results. Do 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.