

## Introduction

Frequency analyses are useful when you are analyzing discrete data, but also when you find that assumptions of ANOVA are not satisfied and that a non-parametric test will not solve the assumptions problem. JMP will shift to a frequency analysis whenever you ask it to work with nominal (or ordinal) variables in the Fit Y by X or Fit Model platform. JMP pays attention to whether the variable is classified as nominal or ordinal. Although Quinn and Keough describe many frequency analyses as having no independent variables (X's) and no dependent variable (Y), you must specify an X or Y variable to perform most frequency analyses in JMP (nominal X, nominal Y).

We will practice one-way (goodness of fit), two-way (contingency table), and multi-way analyses using JMP.

## Objectives

- Calculate expected values, chi-square and log-likelihood values for goodness of fit tests by hand and compare results to a chi-square distribution for significance testing.
- Learn to use JMP to perform goodness of fit tests.
- Calculate chi-square values for contingency table analysis and compare results to a chi-square distribution for significance testing.
- Learn to use JMP to perform contingency table analyses and obtain log-likelihood, chi-square, and exact probabilities (for small data sets).
- Perform a multi-way frequency analysis using the Fit Model Platform in JMP

## Exercise 1- Analyzing sex ratios for beetles in the genus *Phratora*

Evolutionary biologists have long wondered why most species consist of about 50% females and 50% males. R.A. Fisher, a famous evolutionist and statistician, proposed that sex ratios are even because any deviation from equality would result in an evolutionary advantage to the underrepresented sex (its better to be a male when males are rare, better to be a female when females are rare). Consequently, evolutionists have been very interested in cases where the sex ratio differs from 50:50, and one case is found in ladybird beetles. Female ladybirds sometimes pass a bacterium onto their daughters. It is in the evolutionary interest of the bacterium that a female ladybug produces more daughters (who pass on the bacterium) than sons. The bacterium has therefore evolved 'male-killing' behavior, and ladybug populations where the bacterium is common consist largely of females.

I was interested in determining whether the male-killing bacterium was found in willow leaf beetles, and I kept track of sex ratios for every species that I (or friends) collected while I lived in Europe. I was looking for cases where females predominated.

### *Opening the data file*

Open the file 'phratora\_sexratios.jmp.' Variables shown are species, locality, country, sex, and number of beetles sampled. I collected most of these beetles (except the Alaskan ones) while I lived in Europe. Usually, I collected beetles in small towns or villages. I collected the species *Phratora*

*polaris* W on willows in central Finland. A colleague collected the species *Phratora polaris* B on birch in northern Finland (Lapland), and it turns out that this beetle probably belongs to a different species than the willow feeding beetle.

#### *Conducting a goodness of fit test by hand*

1. To determine whether the observed sex ratio differs from the expected 50:50 ratio in one of these species, first select the data from *Phratora polaris* B. Calculate the expected frequencies and calculate chi-square values as shown on pages 380-381 of Quinn and Keough. Compare your chi-square value to the chi-square distribution (table provided to you in lecture), with one degree of freedom.
2. Repeat step 1, but this time calculate the log-likelihood value (G) using the formulae found in Sokal and Rohlf, given to you in a handout in lecture.
3. What do you conclude based on these tests? Do results meet assumptions about the size of the expected frequencies, mentioned in lecture?

#### *Conducting goodness of fit tests in JMP*

4. Now let's perform an array of goodness of fit tests in JMP. Select 'Analyze' and 'Distribution' and add 'sex' to the 'Y, Columns' box. Then add 'number' to the freq box and 'species' to the 'By' box. Adding 'number' tells JMP about the number of beetles collected and adding 'locality' to the 'By' box tells JMP to analyze each species separately.
5. Press OK, then examine output. You should see bar graphs for each species, showing the number of beetles of each sex collected.
6. For *Phratora polaris* B, click on the red triangle beneath 'sex' and select 'Test probabilities.' You will see a new small box open up in the output window. You should enter 0.5 into both boxes (indicating that you expect 50% of the beetles to belong to each sex).
7. Examine output closely. You should see the chi-square value (called Pearson), G value (log-likelihood), degrees of freedom, and *P* value. Compare these to the ones you calculated by hand.
8. Perform goodness of fit tests for the other species and make a table indicating the sex ratio for each sample, the log-likelihood value, and *P*.

#### *Contingency table analysis of sex-ratio data by hand*

Contingency tables answer very different questions than goodness of fit tests. While a goodness of fit test compares an observed frequency to some theoretical expectation, contingency tables make no *a priori* hypotheses about population values. For example, a contingency table analysis does not test the agreement between observed sex ratio and an expected value. Rather, it tests whether specified groups differ in sex ratio, or whether sex ratio *depends on another variable*. When a relationship does not exist, we say that the two variables are *independent* (see page 385, Quinn and Keough). We will test whether populations of *Phratora laticollis* differed in sex ratio.

9. Select the four rows for *Phratora laticollis* by holding the mouse key down and scrolling down four rows (on the left side). Then select 'Tables' and 'Subset' and press OK. You should see a new JMP table with only four rows.
10. Construct a table with sex in columns and locality in rows. Calculate expected frequencies, under the null hypothesis of *independence* or 'no-association' by using equation 14.4 (page 385) of Quinn and Keough.

11. Calculate the chi-square statistic, using formula 14.5 (page 385) of Quinn and Keough.
12. Compare the chi-square statistic to chi-square distribution [with  $(i-1)(j-1)$ , or 1 df] and evaluate the null hypothesis. Explain how the null hypothesis is described in the table of expected frequencies and describe how your observed data relate to this null hypothesis.

*Contingency table analysis in JMP (see pages 250-253 of your JMP manual for more information)*

13. Select the 'Analyze' and 'Fit Y by X' platform in JMP. Add 'locality' to the 'X, Factor' box and sex to the 'Y, Response' box. Then add number to the 'Freq' box and press OK.
14. Examine the output carefully and note that a  $P$  value is reported in three places. The Likelihood Ratio gives the value of  $G$  and its associated  $P$  value, and the Pearson value should match the chi-square you calculated. Finally, the JMP output provides a 2-tail exact probability, which is slightly higher than the probabilities calculated by the  $G$  or the Pearson chi-square test. JMP calculates exact probabilities for fairly small data sets. Whenever you obtain an exact probability, it is better to use it than the probability obtained by the  $G$  or the Pearson chi-square test.
15. Click on the red arrow beneath 'Contingency Table' and deselect the row, column, and total percents one by one, then click again to select the Expected option.

### **Exercise 2- Using contingency tables when assumptions of ANOVA are drastically invalid**

I recently began working with a Russian scientist, Elena Zvereva working in Finland who has looked at the relationship between willow chemistry, predation and parasitism on a leaf beetle (*Chrysomela lapponica*) in the northwest corner of Russia (Kola Peninsula, near Murmansk). She conducted a field experiment comparing predation and parasitism rates on a willow with low levels of salicylates (*Salix caprea*) versus another one, which the beetle prefers, and which contains high levels of salicylates (*Salix borealis*). We will use contingency tables to analyze an experiment where assumptions of ANOVA are so drastically violated that even the non-parametric Wilcoxon two-sample test should not be used (because variances are very unequal, see page 196 of Quinn and Keough).

#### *Opening the data file*

Open the file 'Pred\_Par\_Russia.jmp.' Variables shown are HostSpecies, PlantID, proportion parasitized, and proportion predated. There are two empty columns that you will fill in.

#### *Comparing means and verifying that ANOVA assumptions are invalid*

1. Run the 'Analyze' and 'Fit Y by X' platform, using HostSpecies as X and predation and parasitism rates as Y variables. Compare means and standard deviations. What is the trend as far as parasitism and predation on these two plant species?
2. Use the Normal Quantile plots to convince yourself that a serious heterogeneity of variance problem exists. Look at your data and propose an explanation for why the standard deviations are so different.

#### *Creating columns for contingency table analysis*

3. Point at the first row of the data in the 'par?' column and type a '1,' which indicates that parasitism occurred on that plant (parasitism rate is greater than zero and equals 0.07). Fill in the rest of the 'par?' column with ones if parasitism occurred, and zeroes if no parasitism occurred.
4. Repeat step 3 for predation.

5. Click on the blue 'c' in the left-hand side, and convert this variable from 'continuous' to 'nominal.' You have now created a nominal variable with two values, 1 = parasitism occurred on the plant, and 0 = no parasitism occurred.
6. Repeat step 5 for the 'pred?' variable.

#### *Creating contingency tables*

7. Point at the first row of the data in the 'par?' column and type a '1,' which indicates that parasitism occurred on that plant (parasitism rate is greater than zero and equals 0.07). Fill in the rest of the 'par?' column with ones if parasitism occurred, and zeroes if no parasitism occurred.
8. Design a contingency table, by hand, that compares the amount of parasitism on each host species.
9. Repeat step 8 for predation and interpret both tables.

#### *Contingency table analysis in JMP*

16. Select the 'Analyze' and 'Fit Y by X' platform in JMP. Add 'HostSpecies' to the 'X, Factor' box and 'par?' and 'pred?' to the 'Y, Response' box. Because you have one observation per row in these data, you don't need to add anything to the 'Freq' box. Press OK.
17. Examine the output carefully. Compare observed and expected frequencies for both tables and determine whether you can refute the null hypothesis. Write a brief summary of results
18. Compare the highlighted text in the paper that Elena and I published this year to your interpretation.

### **Exercise 3- Analyzing multiway tables in JMP**

If you find yourself in a willow patch full of beetles in the spring, you will notice that the beetles appear to be resting, looking for a mate, eating, or mating. In a few cases, you might observe a male trying to dislocate another male from his mating position (on the back of his mate). You may guess that even beetles that aren't mating might differ in behavior. Male reproductive success is affected by his mating frequency, while female reproductive success is determined by the number of offspring she produces.

We counted the number of female and male beetles engaged in one of these behaviors during the early morning, when it was still cool, and then again in mid-afternoon after it had warmed up. You will examine the behavior of the non-mating individuals at morning and afternoon. Note that you now have three variables (time of day, sex, and behavior) to compare, rather than just two. You cannot compare all three variables in one contingency table analysis. These data are more consistent with the three way examples shown on Table 14.3 of Quinn and Keough (page 389).

#### *Opening the data file*

Open the file 'time\_by\_sex\_by\_behavior.jmp.' Variables shown are time (am or pm), sex (female or male), and behavior (sitting, walking, or feeding).

#### *Constructing a multiway table*

1. Select the '**Tables** and '**Summary**' commands and group by all three variables. Construct a table showing the values, with a similar structure as Table 14.3a in Quinn and Keough. Include row and column totals on the table
2. Examine the patterns found on the table and write a brief description.

### *Contingency table analysis*

We will perform two contingency table analyses to compare the behavior of each sex for the morning and afternoon separately.

3. Select the 'Analyze' and 'Fit Y by X' platform in JMP. Add 'sex' to the 'X, Factor' box and 'behavior' to the 'Y, Response' box. Because you have two observation times, add 'time' to the 'By' box before pressing OK.
4. Examine the output and note any additional patterns that are evident from it. Look particularly closely at the differences between morning and afternoon. Unfortunately, there is no way to simultaneously compare behaviors among both sexes among both time points using two-way tables.

### *Multi-way analysis using Fit Model Platform*

JMP does not perform log-linear models in the format described in the rest of Chapter 14 in Quinn and Keough, but it does perform analyses that can answer the following question: Does the behavior of each sex change between morning and afternoon? JMP's method involves a categorical analogue to ANOVA, where you have nominal independent variables and a nominal response (Y) variable. This method is a variant on logistic regression, which we will cover next week.

5. Select the 'Analyze' and 'Fit Model' platform in JMP. Add 'sex' and 'time' and the 'sex\*time' interaction to the 'Construct Model Effects' box and 'behavior' to the 'Y' box. Note that the 'Personality' automatically toggles to 'Nominal Logistic' and press Run.
6. Scroll down to the bottom of the output and you will find a table that looks similar to an ANOVA table. These effects should be tested with log-likelihood values (G), but for some reason, JMP does not give these to you. So click on the red arrow next to 'Nominal Logistic Fit for Behavior' and select 'Likelihood Ratio Tests.'
7. Each test examines whether one of the main effects or the interaction among main effects relates to beetle behavior. Write out, in plain English, the three null hypotheses, and indicate whether they are refuted (the null hypothesis is considered refuted if  $P < 0.05$ ). Then write a brief description of the results in plain English. Discuss the result with your instructor if you are not sure how to interpret the analysis.