Biology 345: Biometry  
SONOMA STATE UNIVERSITY  
Lab Exercise 8 – One Way ANOVA and comparisons among means

Introduction

In this exercise, we will conduct one-way analyses of variance using two different platforms in JMP. We will also conduct planned and unplanned comparisons among means. Finally, we will learn how to conduct non-parametric correlation analysis.

Objectives

- Learn to conduct one-way ANOVA using the Fit Y by X and Fit Model platform.
- Conduct planned and unplanned comparisons among means.

Exercise 1- Conducting one-way ANOVA

Last week, we discussed the use of t tests (and their non-parametric equivalent) to compare two populations (groups or experimental treatments). Quinn and Keough (Chapter 8) and the JMP manual (Chapter 7) review the use of single factor analysis of variance (one-way ANOVA) to analyze experiments where more than two populations are compared. The results of the ANOVA tell us whether differences among group means exist (Fixed effects model). As discussed in class, a similar approach is used in ANOVA as in regression analyses, namely that the ANOVA model is fit to minimize the sum of squares in the error variable.

We will use a JMP data file that is closely related to the one we examined last week to practice conducting and interpreting one-way ANOVA.

Opening the data file

Open the file ‘OldFieldSiteFinnishData.JMP.’ This week’s data compares host plant characteristics and beetle performance on three species of Finnish willows, growing together in an abandoned field: *Salix myrsinifolia*, *Salix pentandra*, and *Salix phylicifolia*. Variables shown are the same as those used last week: LeafWater, LeafNitrogen, LeafToughness, TotalSalicylates, TotalPhenolics, RelLarGrowth, AreaEaten, and AveSurvival. Data for each host species are coded a different color.

Comparing species values and testing statistical assumptions for nitrogen content

1. Select the ‘Analyze’ and ‘Fit Y by X’ commands, and drag ‘HostSpec’ into the ‘X, Factor’ box and ‘LeafNitrogen’ into the ‘Y, Response’ box before clicking ‘OK.’
2. Select the red triangle, ‘Normal Quantile Plot’ and ‘Plot Actual by Quantile.’ You will now see quantile plots, shown separately for each species, to the right of the scatterplots. Your JMP manual describes these plots on page 148-150. Check to make sure that ANOVA assumptions (normality and homogeneity of variance) appear to be satisfied.
3. Select the red triangle, and ‘Means/Anova/t test.’ The output will show the Analysis of Variance results. On a separate sheet of paper, write down the numbers in the ‘Summary of Fit’ and ‘Analysis of Variance’ tables. Using the example on page 205 of your JMP

---

1 Or whether there is significant variation among groups (Random effects model).
manual and a hand calculator, recalculate mean square column in the table, the F value, the root mean square error, and r square. Show your work, as done in the JMP manual.

**Determining which variables can be analyzed parametrically**

4. Begin a table with nine rows, one for each variable, and repeat the inspection of quantile plots described in step 2. Next to each variable, write down whether the quantile plots appeared to show that the variable adheres to the normality and homogeneity of variance assumption. You should find three variables that clearly do not appear to adhere to the assumptions. For the variables that do adhere to assumptions, conduct ANOVAs and record the following as columns in your table: F, degrees of freedom in the numerator, degrees of freedom in the denominator, and P.

5. Carefully examine the three variables that failed to meet assumptions in the data table. One of them has an obvious problem for parametric analyses, because it affects the variance of one of the host species greatly. Which variable is it, and what is the problem?

**Transforming variables and analyzing them**

6. We will try log-transforming the other two variables to see whether the transformed data conform to ANOVA assumptions. Select ‘Column’ and ‘New Column.’ When the next box opens, give the column the name ‘logXXX’ where XXXX refers to the name of the original variable. Then select ‘New Property’ and ‘Formula.’ Using the formula editor (transcendental option), obtain the base 10 log of your variable, with the number 1 added to each value (e.g. logXXXX = log10(XXX + 1)).

7. Now select the ‘Analyze’ and ‘Fit Y by X’ platform again, and compare the quantile plots for each of the two transformed variables to the untransformed one. If you are satisfied that the transformation brings the variable into conformance with ANOVA assumptions, add two rows to your table, indicate the names of the log transformed variables, and include the same information for the transformed variables as you did for the untransformed ones in step 4.

**Using the Fit Model platform to save ANOVA residuals**

8. Unfortunately, you cannot save predicted values or residuals within the ‘Fit Y by X’ platform. Choose two variables that you analyzed above, select ‘Analyze’ and ‘Fit Model,’ place HostSpec into the ‘Construct Model Effects’ box, and your two response variables into the ‘Y’ box. Make sure that the ‘Personality’ box continues to say ‘Standard Least Squares’ and then press ‘Run Model.’

9. In the output window, select the red triangle next to ‘Response Variablename’ and select ‘Save Columns’ and ‘Residuals.’ Repeat this action for the second response variable (output shown below the first one).

**Determining whether residuals are normally distributed**

10. Select ‘Analyze’ and ‘Distribution’ and drag your two columns of residuals into the ‘Y, Columns’ box. Then perform Shapiro-Wilks tests on each set of residuals to establish

---

2 We add one to each value to avoid problems that arise when base 10 logs are taken from values between zero and one (the log would be a negative value).
whether they conform to expectations for a normal distribution. If you are unsure how to interpret the output from this analysis, please ask your instructor!

11. Write a brief summary of the results of your normality tests.

**Conducting non-parametric ANOVA**

12. Your table, above, should contain one variable that could not be analyzed parametrically. The non-parametric methods in JMP, which you used last week to compare two species, can also be used when comparing more than two groups (see pages 222-224 of your JMP manual). Select ‘Analyze’ and ‘Fit Y by X’ with your remaining response variable as Y. Then select ‘Nonparametric’ and ‘Wilcoxon Test’ to obtain the results of a non-parametric test of overall differences among the three host species. This test is equivalent to the ‘Kruskal-Wallis’ test that is described in many statistics textbooks.

13. Write a sentence describing your findings for this last variable, based on the non-parametric analysis. Read the paragraphs in section 8.5.2 (pages 195-196 of Quinn and Keough) to gain some background into the Kruskal-Wallis test. Based on what you read, describe a potential problem with your use of the Kruskal-Wallis test with this variable.

**Exercise 2- Comparisons among means**

When we conduct ANOVA, the emphasis is on overall differences among groups or treatments, and not on whether individual groups differ from other groups. As scientists, however, we are often very interested in determining which means are significantly different. Two kinds of means comparisons exist.

*Planned comparisons or contrasts* are made between means that we expect, for some reason that is a part of the experiment or study, should be different. We usually construct a contrast that compares *groups of group means* rather than each individual pair of means.

*Unplanned* comparisons are pairwise tests of significant differences among means, usually conducted without knowing in advance which comparisons might be of interest. One could conduct individual *t* tests on each pair of means, but many statisticians believe that this procedure introduces the multiple comparisons problem of repeated significance testing (i.e. making many statistical tests increases the likelihood that two means will differ to a certain degree simply by chance). We will therefore use the *Tukey’s HSD test*, which adjusts the significance level by the number of comparisons made. This test is discussed in Quinn and Keough (pages 199-200) and in your JMP manual (pages 212-213).

In our case, we are only comparing three species, so three pairwise comparisons among species means are possible (MYR vs PHY), (MYR vs PEN), (PEN vs PHY). The multiple comparisons problem is much greater if you are comparing more groups of means.

**Unplanned comparisons**

1. Begin a table with eight rows (one for each variable) and 10 columns, listed in the following order: sample size for *Salix myrsinifolia*, mean ± standard error for *S. myrsinifolia*, sample size for *S. pentandra*, mean ± standard error for *S. pentandra*, sample size for *S. phylicifolia*, mean ± standard error for *S. phylicifolia*, a column for F, degrees of freedom in the numerator, degrees of freedom in the denominator, and the P value. You will use the last four columns when you make your planned comparisons (see below). Check with your instructor if you are not sure about the format of this table. I
suggest that you enter your means and standard errors to 2-3 significant digits (e.g. 62.5 ± 0.59). Remember to use your transformed data on this table if you found that a variable conformed to ANOVA assumptions better after transformation.

2. For each variable, select ‘Analyze’ and ‘Fit Y by X.’ Click on the red triangle to obtain ‘Means and Stddev.’ Then click on it again to select ‘Compare Means’ and ‘All Pairs, Tukey HSD.’ You will see a new window appear, to the right of the graph, and output below that indicates (with positive values) which means are significantly different. Compare your output to the output shown on pages 183 and 184 of your JMP manual. Fill in the means and standard errors on your table. Then write, in lower case superscript, next to each mean, a letter. If there are no significant differences among the means, use the letter ‘a’ for all three superscripts. If one species has higher values than the others, use superscript ‘a’ for the higher mean and superscript ‘b’ for the two lower means. Check with your instructor if you are not clear how to interpret the output or to decide which superscripts to use.

Planned comparisons

This data is well suited for a planned comparison. As you may have noticed, the willows *S. myrsinifolia* and *S. pentandra* both contain substantial amounts of salicylates, while the willow *S. phylicifolia* contains little or no salicylates. The beetle I studied in Europe, *Phratora vitellinae*, never fed on *S. phylicifolia* in nature. We can therefore compare the beetle’s two natural host plants (MYR and PEN) to the species it does not use in nature (PHY). The results are of interest because they would show whether the two species used by the beetle share characteristics (high nutritive quality, high beetle growth and survival).

3. To conduct a planned comparison, you must use the ‘Analyze’ and ‘Fit Model’ platform. Construct your model using the variable ‘LeafWater’ and press ‘Run Model.’ To the right of the ANOVA output, you should see a leverage plot with the label ‘HostSpec.’ This window shows values for the least-squares means, which are estimated in the ANOVA calculations. I prefer to report these in my own data analyses because they are adjusted for differences in sample size. In the present case, you should notice that the least squares means are pretty similar to the actual means in the far right column.

4. Click on the red triangle under ‘HostSpec’ and select ‘LSmeans plot.’ This shows you a graph of the least squares means and their standard errors and is a very useful feature in JMP.

5. Click on the red triangle under ‘HostSpec’ again and select ‘LSmeans contrast.’ You will see a new window open at the bottom of the output. Click in the plus box next to ‘SalixMyr’ and then click in the plus box next to ‘SalixPentandra.’ Now click on the minus box next to ‘SalixPhylic.’ This sets up your contrast comparing the overall mean of PEN and MYR to the mean of PHY. Note that the two sets of contrasts sum to zero (as described just below formula 8.8, page 198 of Quinn and Keough).

6. Click on ‘Done.’ Note that you are now given an F value, numerator degrees of freedom, denominator degrees of freedom, and a P value. Enter these values onto your table.

7. Repeat steps 3-6 for the other variables and fill in the table.
Interpreting your results

8. Write a paragraph presenting and interpreting the results on your table. As you did last week, organize your paragraph by the kinds of variables tested (plant nutritional quality, plant chemistry, beetle performance). Indicate which variables differed significantly between species, and which differed little or not at all. Then summarize the information about the planned comparison between the two species used by the beetle and the other willow species. Use the results to answer the following question: do host species used by the beetle tend to be similar with respect to plant nutritional characteristics and beetle performance, or not?

Tasks

1. Construct an ANOVA data and confirm how some values in the table are calculated directly from others, using an example from your JMP manual (page 176).

2. Examine a series of variables to determine whether they conform to ANOVA assumptions. Record the results of the ANOVA in a table, for those variables that did conform. Try to transform other variables to bring them into conformance and enter results from any transformed variables into your table.

3. Save residuals from two ANOVAs and compare their distributions to normal expectations using the Shapiro-Wilk test.

4. Conduct and interpret results from a non-parametric ANOVA on remaining variables.

5. Conduct unplanned comparisons tests among species means and use these to fill in a table of means with superscripts indicating which means are significantly different from others.

6. Conduct planned comparisons tests between the two species used by a beetle as natural host plants (S. myrsinifolia and S. pentandra) and the remaining species S. phylicifolia. Write a paragraph interpreting your results.