Rock Pool Statistical Analysis

Now that you have made your maps, and examined the spatial distribution of rock pools and the species that live in them, it is time to examine some of the statistical patterns in the data.

Using the Excel file that you downloaded from ArcGIS, you will examine the relationship between pool size and species diversity. You will compare two approaches for conducting this analysis: a one-way ANOVA with pool size as categories and a liner regression with pool size as a continuous numeric variable. We will be using “Surface\_Area” as our size metric.

**Step 1.** Create your pool categories (Small, Medium, Large, XLarge)

Make a new column in Excel for Size category and use the following formula to assign a pool category based on Surface Area (Small = 0 – 1,999, Medium = 2,000 – 3,9999, Large = 4,000 – 9,999, XLarge = >10,000 cm2). You can do this by selecting ALL the data and sorting by Surface Area and manually entering and copying the labels into a new column. Or, you can use an “IF” formula.

In this example, the Surface Area data is in column G, but make sure to adjust the formula if your data is in a different column:

=IF(G2>10000,"XLarge",IF(G2>4000,"Large",IF(G2>2000,"Medium","Small")))

Step 2. Copy and paste your data from Excel into JMP. Remember to use the “**Paste with Column Names**” command to make it easy to get your data into JMP with the correct labels.

Conduct two separate analyses to examine the effect of pool size on species richness. Select the appropriate variables to conduct: 1) One-way ANOVA testing for differences in biodiversity between pool sizes and 2) a Linear Regression testing for the effect of pool size on biodiversity. Species richness is your measure biodiversity and your dependent variable.

In addition to these two analyses, look at the available data fields and select an additional third comparison to test. Do not sift through every comparison to try and find a significant result (this is data dredging!). Think about an interesting question, conduct an appropriate statistical test (you can use any of the statistical tests we’ve covered), and report whether you obtained a statistically significant result.

Assignment: Results Section + Literature Comparison

Your task is to compose a complete results section for your rock pool analysis. You should have three maps and three figures to include with your results text. **Include a figure legend, NOT a figure title. Label your axes. You should include appropriate and correctly formatted descriptive and statistical results for the three statistical analyses you performed. For comparisons of means, include some representation of the variation in your data (standard error or standard deviation). In the parenthetical reporting of the relevant statistics, be sure to also reference the correct figures and maps that illustrates your results.**

**In addition to a standard results section, include an additional short paragraph comparing your results with the results found in TWO relevant peer-reviewed scientific studies. Include the citations for these studies at the end of the assignment, as well as using appropriate in text citations.**

ANOVA and Regression in JMP

Conduct a one-way ANOVA: Compare the means of more than two groups

Conduct your ANOVA analysis using the same procedure as the two-sample t-Test (“Fix Y by X”). Remember, a one-way ANOVA is very similar computationally to a two-sample t-Test. The only difference is that you use an F statistic in ANOVA when your independent variable has three or more groups instead of a T statistic with a two sample t-Test. In JMP, you set-up an ANOVA the exact same way as a two-sample t-Test and JMP chooses the test statistic based on the number of groups in your data for you. Other statistical programs require that you DO know this difference and code it correctly.

Conduct an analysis to test for pairwise differences between your groups by selecting “Compare Means” and “All pairs, Tukey HSD”.

Conduct a Simple Linear Regression

Use “Fit Y by X” to set-up your analysis and place your predictor variable in the X box and the response variable in the Y box. Click “OK” to obtain a scatterplot of your data.

Use the red triangle to select “Fit Line” to obtain your regression results.

Your R2 value estimates the proportion of variation in the response that can be attributed to the model rather than to random error. An R2 closer to 1 indicates a better fit to the data than does an R2 closer to 0. An R2 near 0 indicates that the model is not a much better predictor of the response than the response mean. Rsquare Adj is the R2 value adjusted by the degrees of freedom to account for the number of parameters in the model. You would need this value if you were comparing analyses with different sample sizes.

The Analysis of Variance report provides the calculations for comparing the fitted model to a model where all predicted values equal the response mean. Perhaps more straightforward is the Parameter Estimates. The Parameter Estimates report shows the estimates of the model parameters and, for each parameter, gives a t test for the hypothesis that the relationship between your two parameters equals zero. Notice that your p-value is the sane for either test! Use the Parameter Estimates output to report the results from your regression.

Grading (50 point lab assignment)

\_\_\_\_\_/ 10 Did you clearly and correctly state the descriptive and statistical results in words?

\_\_\_\_\_/ 5 Did you include the correct parenthetical information to report the statistics?

\_\_\_\_\_/ 10 Did you select the appropriate format and display for your figures? Do your figures have all the necessary components?

\_\_\_\_\_/ 10 Did you select the appropriate format and display for your maps? Do your maps have all the necessary components?

\_\_\_\_\_/ 5 Did you provide a descriptive legend for each figure/map?

\_\_\_\_\_/ 10 Did you compared your results with two published studies, with correct citations?