Research Methods – Statistical Analysis Take Home Practice

Here are some examples for you to work through. For each example,

- make an appropriate figure to visualize the data
- identify the logical (biological) hypothesis,
- translate the logical hypothesis into a prediction (a statistical hypothesis),
- conduct the appropriate analysis, (hint, are the predictor variables continuous, categorical, or both)
- interpret the analysis and present the results using the format we have discussed in class (e.g. effect size, df, p-value, calculated value of test statistic, etc.)
- Does the test allow you to reject or confirm your predictions?

Housekeeping Issues:

- I have put all the data for these exercises into a single Excel workbook. You will have to import the data into R.
- Recall that R typically requires data in a specific format (long format), so it may be necessary to stack the data columns to comply with the formatting requirement.
- Don't forget to check the assumptions of each test.

Examples:

1. In fall 2004, students in the 2 p.m. section of my Biological Data Analysis class had an average height of 66.6 inches, while the average height in the 5 p.m. section was 64.6 inches. Are the average heights of the two sections significantly different? (datafile = heights)

2. Bergmann's Rule is a commonly investigated biogeographical hypothesis to explain geographical variation in adult body size as a function of environmental temperature. As a simple test of the rule, researchers often measure the body size of different populations across a latitudinal gradient (latitude used as a proxy for environmental temperature). Our herpetology class collected a dataset for body size (plastron length) of adult female painted turtles (*Chrysemys picta*) during a spring break field trip covering more than 3,000 road miles from northern New York to south Alabama. Were our efforts able to provide support for a significant Bergmann cline in painted turtles? What can be said about the precision of the relationship? (datafile = Bergmann Turtles).

3. Medley and Clements (1998) sampled a number of stations (between four and seven) in each of six streams known to be polluted by heavy metals in the Rocky Mountains of Colorado. At each site they recorded zinc concentrations and species diversity of the diatom community (higher numbers mean more diverse community). What is the effect of zinc level on the diversity of diatoms in the community? What do the pairwise comparisons among zinc categories tell us? (datafile = diatom diversity)

4. Throughout much of the southeast, two sympatric treefrogs, *Hyla chrysoscelis* and *Hyla versicolor* are extremely difficult to tell apart because they appear very similar in size, color, and behavior. Some have hypothesized that these species may be distinguished by their mating calls. However, call frequency can also be influenced by body size (larger males have more experience calling) and by temperature, as individuals typically prefer to call early in the year at cool temperatures. How would we analyze this potentially complex relationship? Can we determine differences in call rate after accounting for body size? Draw a figure (graph) of what you might expect these relationships to look like. Can we determine differences in call rate after accounting for temperature? What do we do if there is a significant interaction between species and body size or species and temperature? With no significant interaction? (datafile = frog call)

5. A common observation in ecology is that species diversity decreases as you get further from the equator. To see whether this pattern could be seen on a small scale, I used data from the Audubon Society's Christmas Bird Count, in which birders try to count all the birds in a 15-mile diameter area during one winter day. I looked at the total number of species seen in each area on the Delmarva Peninsula during the 2005 count. Latitude and number of bird species are the two measurement variables; location is the hidden nominal variable. What can we say from these observation? (datafile = Christmas Bird Count)

6. An experiment was set up to measure the degree of bacterial contamination on quarter coins ($0.25) collected from three different businesses, a butcher's shop, a sandwich shop, and a newsagent. A sample of four coins was collected at random from the cash register at each location. The number of bacterial colonies present on each coin was estimated by dilution plating. (datafile = contaminated coins)

7. In a Vanderbilt hematology lab they are developing new drugs to help patients with hemophilia. They have tested the efficacy of a new compound (Drug G) against the most common currently prescribed drug (Drug B). What can we say about the efficacy of the two different drugs? (data are blood-clotting times (minutes); datafile = hemophilia)

8. Farmer John owns a pig feedlot and is interested in testing the differences among four of the most commonly used commercial feed types. These feeds differ only in their ratio of protein to fat and he is interested in determining whether the different feeds produce different rates of weight gain at his feedlot. Do we have evidence to suggest that the pig feeds vary in their ability to fatten-up pigs? Is one kind of feed better at producing fat pigs than another? (datafile = feed)

9. A research biologist at the Tennessee Wildlife Resources Agency is interested in developing a predictive model to describe the relationship between age and body size (measured as wing length) in Bachman's Sparrow (*Peucaea aestivalis*). This species is only rarely found in Tennessee, but the researcher was able to find 13 nests (42 birds total) over the course of a one month survey. He measured wing

length once for each nestling bird and was able to determine the hatching date (and thus age) by observations of each nest. Since individual nestlings cannot be considered true replicates (why?) he used the mean wing length for each nest in the analysis. What is the relationship between wing length and age in Bachman's Sparrow?

10. We observe that Clark's Nutcrackers (granivorous seed caching birds) uses forest micro-habitats (trees, shrubs, ground microhabitats) for foraging in the spring differently than compared to the proportion of those same micro-habitats used for foraging in the autumn. The null hypothesis would be that birds feed in these three forest micro-habitats in the same proportion in the spring as in the autumn. Is that hypothesis supported?

Source of Examples:

Armstrong, R.A. and A. Hilton. 2004. The use of analysis of variance (ANOVA) in applied microbiology. Microbiologist. 2004:18-21.

Brower, J.E., J.H. Zar, and C.N. von Ende. 1998. Field and Laboratory Methods for General Ecology. 4th Edition. McGraw-Hill.

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.

Medley, C.N. and W.H. Clements. 1998. Responses of diatom communities to heavy metals in streams: the influence of longitudinal variation. Ecological Applications. 8(3):631-644.

Zar, J.H. 1999. Biostatistical Analysis (4th ed.). Prentice-Hall.