STANDARD PARAMETRIC ANALYSIS
BI 311
Homework 06

Introduction:

You have already been introduced to the fundamentals of inferential statistics and practiced Chi-squared tests of independence useful for comparing frequency counts. However, often you have two data sets made of continuous measurements (not counts) and you would like to test for differences or correlations between the data sets. Parametric statistics are powerful analytical tools that allow such testing as long as data meet parametric assumptions such as the assumption that data are normally distributed. You have already learned how to test for normality in a previous homework. Now, we will practice some of the fundamental types of parametric tests: t-test, ANOVA and regression analysis.

Skills:

1. Learn how to use EXCEL to estimate a t-statistic and conduct a t-test.
2. Learn how to use EXCEL to conduct an analysis of variance (ANOVA).
3. Learn how to use EXCEL to conduct a simple-linear regression analysis.

Materials:

- A Handbook of Biological Investigation, Ambrose et al. 2007
- Access to WORD and EXCEL
- EXCEL file named Scrub Lizard Demographics

EXERCISE 1- USING EXCEL FOR A t-TEST

Background:

A t-test is used whenever there are two data sets (groups/populations) representing two levels of an independent variable. We will use a simple t-test to test for significant difference between female and male home range size for the Florida scrub lizard. In this case, we have a data set for females and for males. The independent variable is sex and the dependent variable is home range size. We want to know if sex has a significant effect on home range size.

Laboratory Procedure:

1. Download the *file Scrub Lizard Demographics* available on the moodle site and select the worksheet (tab at the bottom) labeled *Home Range*. This is the same data we analyzed for normality in last week's homework and you should recall that we found the raw values for home range (m^2) were significantly skewed for both female and male scrub lizards. However, after ln-transforming the values, we removed the skewed distribution. Therefore, we will want to test for a difference between male and female lizards using the ln-transformed values and not the raw values.
2. Another assumption of parametric analysis is that the variance is similar between the groups being analyzed (in this case female versus male home range size). We can easily exam the differences in variance between the two groups. Within the worksheet labeled *Home Range*, calculate descriptive statistics and exam the variance estimates for ln-transformed data of female and male home range size. If the difference between the smallest and largest variance estimate is greater than 20:1, we should assume unequal variance between the groups. Otherwise, we will assume the variance is equal.
3. Now go to *data analysis* and select *t-test for 2 samples assuming equal variance* or *t-test assuming unequal variance* depending on the result of your variance test. (Another t-test appears on the list that we will not use today. A *paired t-test* is used when the "samples" are measurements of the same subject at different points in time. For example, if we measured each lizards home range twice and wanted to know if there was a difference between the first and second measurement.) Highlight all the data in the female column for *variable one range* and then highlight all the data in the male column for *variable 2 range*. Specify that we are testing for zero difference between the means by typing a 0 and select ok. A new worksheet will be created with the statistical results.
4. Create a table (as you did in a previous homework) in a WORD file and report the factor being tested (independent variable = sex), t-statistic, degrees of freedom and the p-value for your test. Note that there are two p-values: one for a one-tailed test and one for a two-tailed test. You will want to use the p-value for the two-tailed test because we want to know if there is <u>any</u> significant difference regardless of whether it was males or females that have a larger size. If your two-tailed p-value is less than 0.05, the 2 statistical populations (sexes) are statistically different: meaning that sex has a significant effect on home range size. What is your conclusion? Which sex has a larger/smaller home range?

EXERCISE 2- USING EXCEL FOR ANALYSIS OF VARIANCE (ANOVA)

Background:

ANOVAs are useful for testing for differences between 3 or more data sets and also for testing across 2 or more independent variables. We will first use a simple One-Way ANOVA to test for differences in survivorship between scrub lizards in different age categories: hatchlings, juveniles and adults. We have only one independent variable in this case (age) but it has 3 levels resulting in 3 different data sets. We would like to know if there is a difference in survival between hatchlings, juveniles and adults.

We will then follow this analysis with an example of a Two-way ANOVA. If we have survivorship data grouped by age (hatchlings, juveniles and adults) and by sex (male versus females) we can use a Two-Way ANOVA to simultaneously test for age and sex effects. Additionally, we can test for interaction effects. An interaction occurs when one of the independent variables has a significant effect but at only one level of another independent variable. For example, we might find that survivorship is significantly affected by age but only for females and not for males.

Laboratory Procedure:

1. Open the *Scrub Lizard Demographics* file available on moodle and open the worksheet named *Survivorship*. Because ANOVA is a type of parametric test, we always need to test parametric assumptions of normality and equal variance as we did above in the t-test example. To save time, I've already completed these tests and you can assume all parametric assumptions have been met.
2. Within the worksheet labeled *Survivorship* are survivorship data for hatchlings, juveniles and adults for 10 different patches. Go to the *Data* menu then to *data analysis* and select *ANOVA: Single Factor*. For the *input* range, select columns B, C and D including the first row and then select the labels in the first row option. Now select *OK* and EXCEL will calculate a summary table and the results of the ANOVA. Submit the ANOVA table for your homework. Note whether the p-value is less than 0.05 (careful: sometimes this is given in scientific notation). Is there a significant age effect on survival?
3. Note: this analysis simply tells you that there is a difference between hatchlings, juveniles and adults with respect to survival but doesn't tell you which groups are different. For example, is there a difference between all 3 ages or are only hatchlings different from the other two. To find out, you can complete a post-hoc analysis using either post-hoc tests or graphical analysis. Some sophisticated software will automatically give you a post-hoc test such as a Tukey HSD. EXCEL does not have this capability so I recommend a graphical analysis. Calculate the mean and standard errors for each of the age groups and then make a column graph (as you did in a previous homework) of age versus mean survivorship. After adding your error bars, note which groups have error bars that overlap or not. If error bars overlap, the groups are unlikely to be significantly different and vice versa. Paste your graph in the WORD file and explain which groups are significantly different. Alternatively, you can calculate descriptive statistics for each group (hatchlings, juveniles, adults) and compare confidence intervals for each pair of groups (e.g. hatchlings versus juveniles, hatchlings versus adults, and then juveniles versus adults). If confidence intervals for the mean do not overlap, it is likely the two groups are different with respect to survivorship.
4. Now we will practice a Two-Way ANOVA. Go to the *Data* menu and select *data analysis* and then select *2 factor ANOVA with replication*. When you highlight your input data you will have to highlight the entire table, including the column labels. Enter *5* for the number of rows per sample (i.e. replicates for each sex) and click *ok*. The results will print out as a new table. The output is a bit confusing if you don't change some labels. Scroll down to the bottom table and under *Source of variation* replace *sample* with *sex*. Then replace *column* with *age*. Now you can interpret the results. If the p-value in the sex row is less than 0.05, then sex significantly affects survivorship. Likewise, if the p-value in the age row is less than 0.05, then age significantly affects survivorship (but we already knew this from the first analysis). Finally, if the p-value in the *interaction* row is less than 0.05, then there is an interaction between the independent variables. Create a table in your WORD file and report the factor being tested (sex, age and interaction between sex and age), F-Statistic, degrees of freedom and p-value for each test. Is there a sex effect? If so, which sex has higher survivorship? Is there an interaction effect between the two independent variables?

EXERCISE 3- USING EXCEL TO PERFORM CORRELATION AND REGRESSION ANALYSIS

Background:

A t-test or an ANOVA is fine if you have independent variables that can be categorized nominally (e.g. male versus female or hatchlings versus adults). However, sometimes both our independent and dependent variables are continuous and we want to know if one varies significantly with the other. Correlation analysis allows you to test for associations between continuous variables without assuming that one is necessarily causing the effect in the other. Regression assumes that one variable is directly responsible for the variation in the other. We will use regression analysis to test for relationships between patch size and scrub lizard abundance and recruitment.

Laboratory Procedure:

1. In the *Scrub Lizard Demographics* file is a worksheet labeled *Abundance/recruitment*. Data for 8 different patches are provided included the patch size, the log (ln) of patch size, scrub lizard abundance and female recruitment.

2. Again, correlation/regression being parametric analyses you should test each of the continuous variables for parametric assumptions. To save time, I have completed this step for you. Abundance and recruitment met the assumptions of parametric analysis. Because there is a large difference between the smallest and largest patch, I had to log transform patch size to meet parametric assumptions. Therefore, use the ln-transformed data for patch size.
3. From the *Data* menu, select *Data Analysis* and then *Regression*. For the *Input Y range* select all the data in the abundance column (this will be our dependent variable) and for the *Input X range* select all the data in the Ln Size column (this is our independent variable). Select *OK* and several tables will be displayed. The first table is a summary of how well the independent variable explains the dependent variable. The important number to note is R-squared. The larger this number the better the independent variable explains the dependent variable. For our purposes, you can ignore the ANOVA table and skip to the third table. The important row in the third table is the *x-variable* row. The important numbers in this row are the coefficient and the p-value. The p-value will tell you if you have a significant association: if p-value < 0.05 then you have a significant effect of the independent variable on the dependent variable. The coefficient will tell you whether the independent variable is positively or negatively associated with the dependent variable and the strength of the relationship. For example, a large positive coefficient reveals a significant positive association. Is there a significant association between Ln Patch Size and the abundance of scrub lizards? If so, make an x-y scatter plot of Ln Patch Size versus Abundance and paste it in your homework along with the R-squared, coefficient and p-value.
4. Repeat this analysis for Ln Patch Size versus Recruitment. Is there a significant association? Positive or negative? Print a plot showing any significant relationship.

HOMEWORK SUBMISSION

Each person should create a WORD file named *Homework 6*, answer the following questions, and submit their file to me via email.

1. Report the results of your t-test for differences in male versus female home range. Create a table in the WORD file and report the factor being tested (independent variable = sex), t-statistic, degrees of freedom and the p-value for your test. What are your conclusions?
2. Report the results of your single factor and two-factor ANOVA's testing for the effects of age and sex on survivorship. Paste your column graph of age versus mean survivorship. Submit the table describing your single factor and two factor ANOVA's. What are your conclusions about the effects of age and gender on survivorship? Was there an interaction?
3. Report the results of your regression analysis testing for the effects of patch size on abundance and recruitment. Create a table that lists the dependent variable, the R-squared value, the coefficient and the p-value for each test (you can put both in one table). Create a scatter plot for tests that are significant (hint: check out the results in Hokit and Branch, 2003, Journal of Herpetology… the paper you read in class).