

Homework 1

Due January 24, 2018

Question 1

The mean height of the *Sagittaria lancifolia* plant (which grows in the Florida Everglades) is to be compared at four different levels of phosphorus contamination in the soil. Twenty plants are monitored at each of the four soil conditions (for a total of 80 plants). Suppose the resulting data on the *Sagittaria lancifolia* plant heights is strongly skewed to the left. What is a reasonable way to analyze the data (circle the best answer):

- a) Re-do the experiment using less phosphorus contamination in the soil.
- b) Re-do the experiment using a test plant other than *Sagittaria lancifolia*.
- c) Use the Kruskal-Wallis nonparametric ANOVA test.
- d) Since the data is skewed left, the normality assumption is reasonable and one can use the ANOVA F -test.
- e) Compute the correlations between the plant heights at the different sites.
- f) Do a paired comparison t -test.

Question 2

Concentrations of nitrate were compared and contrasted among four major land covers: urban, forested, cropland, and pasture. Ten soil samples were obtained at each type of land cover. An ANOVA was performed to test if the mean nitrate concentrations differed among the four types of land cover. The data produced the following ANOVA table from R:

```
      Df Sum Sq Mean Sq F value    Pr(>F)
cover    3   21.85      7.28   0.0113 *
Residuals 16   62.77      3.92
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- a) Fill in the values for the degrees of freedom (DF), the mean squares (Mean Sq), and the F -test statistic (F value) in the above table.
- b) Would we have rejected the null hypothesis using $\alpha = 0.05$? (Circle One) YES or NO
- c) Would we have rejected the null hypothesis using $\alpha = 0.01$? (Circle One) YES or NO
- d) Based on the p -value, state the conclusion of the ANOVA F -test **in the context of this problem**.
- e) A Tukey multiple comparison procedure was conducted using 95% confidence by forming confidence intervals for the differences of the mean nitrate levels at the four sites. Letting U=urban, F=forested, C=cropland, and P=pasture, the Tukey results were summarized as follows: C-P U-F, where the hyphen indicates no significant difference. From this illustration, write a sentence or two describing the results of the Tukey multiple comparison procedure.

Question 3

A study was done on the effects of thermal pollution on clams. Clams were collected at three sites: an intake site to a plant, a discharge site, and a site near Interstate 55. The R program `clams.R` in the `code` folder

loads the data and will compute statistics regarding the heights of the clams (note that there is also data on width and length of the clams too). The goal of this problem is to determine if the clams differ in terms of mean heights at the three sites. Do the following parts:

- a) To test if the mean heights of the clams are equal at the three sites, define the appropriate parameters and state H_0 and H_1 for this problem.
- b) Run the `clams.R` R program and verify that the F -test statistic is the ratio of the appropriate mean squares from the ANOVA table.
- c) If H_0 is true, what value should F take on average if the experiment were to be repeated over and over?
- d) What are the numerator and denominator degrees of freedom for the F -test?
- e) The R output gives the p -value from the F -test. Would we have rejected the null hypothesis using a significance level $\alpha = 0.05$? How about $\alpha = 0.10$?
- f) Does it make sense to do multiple comparisons looking at differences in pairs of mean heights for this problem? Explain.