Statistics 411/511
Homework 7
Due Tuesday, November 23 by midnight

- **Instructions:** Please see the end of the syllabus for guidelines. Upload your homework to Gradescope via Canvas (access specific homework assignments from the "Assignments" link at the left of the Canvas course page). Your file must be a pdf document. **There will be a one-point deduction if you don't assign pages** (see this Gradescope help video).
- Do the computational part of the homework shortly after completing the week's lab activity.
- The problems are assigned from the **third edition** of the textbook. If you have another edition, consult the copy on one-hour reserve at the library website for the homework problems.
- **Academic Integrity** You are encouraged to *discuss* the homework with other students, but what you turn in must be your own work in your own words. **DO NOT** copy someone else's homework. You may share ideas and R code, but do not share R output or written language. The syllabus contains details and links to OSU's Student Conduct Code and procedure for reporting suspected academic misconduct.

1. This homework deals with the amoeba data from Homework 6. Read in the data using `read.csv()` but add an argument `stringsAsFactors=TRUE`. This will force R to interpret `Treatment` as a factor variable, which will be important for the `relevel()` function. Submit your R code and a `summary()` of the resulting data frame.

2. Do a preliminary analysis by completing the following steps.
   (a) Produce side-by-side boxplots of yield for each treatment. Submit your code and plot.
   (b) Produce an ANOVA table. You did this in Question 2(b) of Homework 6. Submit your code and output.
   (c) Produce a plot of the residuals vs. fitted values. Submit your plot, and comment on the plausibility of the equal variance and normality assumptions based on this plot. See page 13 of Outline 5 for R code to produce the plot.

3. Suppose the researchers had preplanned to estimate all pairwise comparisons among population means.
   (a) Use the Tukey-Kramer procedure to calculate simultaneous 95% confidence intervals for all ten pairwise comparisons. Include your R code and output.
   (b) According to the Tukey-Kramer results, which pairs of population means are found to be different?

4. Suppose the researchers had preplanned to compare None to each of the other four treatments.
   (a) Use Dunnett's method to estimate these four pairwise comparisons at a 95% familywise confidence level. Include your R code and output.

(Problem 4 continued on next page)

(b) Write a statistical conclusion for the Dunnett's results. This statistical conclusion should include a formatted table (not computer output) of the confidence limits. The statistical conclusion should refer to the table and state which pairs of population means are found to differ.

5. Suppose the researchers had preplanned to make the following four comparisons:

    1. None vs. the average of the other four treatments
    2. Form10 vs. FormHeat
    3. Heat10 vs. HeatForm
    4. FormHeat vs. HeatForm

    Use the Bonferroni procedure to estimate these comparisons with a familywise confidence level of 95%. Include your R code. State or highlight the resulting intervals (i.e., please don't make the grader have to search for them in your R output). It's not necessary to format a table of the Bonferroni intervals for this question. Suggestion: include comments in your code so you can make sense of it later.

6. Suppose after looking at side-by-side boxplots, the researchers decided to compare the population mean for None with the average population mean for the other four treatments.

    (a) Use Scheffé's method to estimate this comparison with a confidence level of 95%. Include your R code. State the resulting confidence interval.

    (b) Compare your Scheffé confidence interval to the Bonferroni confidence interval for the same quantity in question 5. Which is a more precise estimate?