

BTRY 6020 Homework VIII

NAME: student name

NETID: student NetID

DUE DATE: May 10, by 8:40 am

Question 1.

A study was conducted to investigate the effect of fertilization on the yield of a commercial variety of tomato. In this study, a completely randomized design was used as the experimental plot was homogeneous both in humidity and nutrient content. The treatments of interest in this study consisted of a control (E=no fertilizer used), very slow release (A), a slow release fertilizer (B), a moderate release fertilizer (C) and a fast release fertilizer (D). These treatment levels were each assigned at random to 20 plots with 15 plants per plot. Prior to planting, the researchers noticed big differences in plant heights. There was no record of when the seeds were planted, so nobody knew if the plants were the same age or not. They therefore decided to record information on mean plant height on each plot. The response variable of interest was the total weight of tomatoes harvested per plot. Data for this study can be downloaded from the course web site. The file name is Hwk8Q1DatSp17.

Answer the following questions.

- A) Give a model for the analysis of covariance, explain each term in the model and formulate appropriate assumptions.
- B) Plot the response variable weight against the covariate height using levels of the treatments as different plotting symbols. What relationship is there between these two variables?
- C) Based on ordinary one-way ANOVA perform a significance test for the equality of the five treatment means. State hypotheses, test statistic, p-value, and your conclusions.
- D) Determine if the interaction between treatment and initial plant height is significant. State hypotheses, test statistic, p-value, and your conclusions.
- E) Regardless of your answer to Part D above, perform analysis of covariance and make and summarize your results.
- F) Based on the model in Part A, make use of standardized residual plots to assess validity of the assumptions of independence, equal variance, and normality.
- G) What multiple comparison method that we've used would be more appropriate to compare each treatment to the control treatment? Use such a method at an overall error rate $\alpha = 0.05$ and state carefully your conclusions.

Question 2.

A field trial is run to test the productivity of three different varieties of strawberries in an experimental field station in New York State. Four equally sized fields are available for use, and one-third of each field is planted in one variety of strawberry, the density of plants kept the same throughout the trial. Each $1/3$ field is then randomly assigned one of the three varieties in such a way that each field has all three varieties planted within it. The Yield of strawberries (kg) over a two week period is then recorded for each $1/3$ field.

- A) Give a model statement for this experiment. Define each term and any constraints or conditions that are attached to it.
- B) Test to see if there are any differences between the yields of the various strawberries. State hypotheses, test statistic, p-value, and conclusions.
- C) Use Tukey's HSD to find the varieties of strawberries which produce different yields. Which would you recommend to be used in New York State (based solely on this yield criteria).
- D) What are the values of the two components of variance in this study? What proportion of the variance of yield is attributable to the differences in fields?
- E) Use `library(lme4)` and `library(LmerTest)` to test to see if the random effects are statistically significant in this model. (Note: There is some debate amongst statisticians that this procedure is appropriate when using REML. Some say it is OK to use REML with this test IF THE FIXED EFFECTS ARE THE SAME IN BOTH THE FULL AND REDUCED MODELS, which how it is stated in our text.)
- F) State which of our observations are correlated in this situation.