STAT 512 HW6

Reading: Ott & Longnecker Chapter 15
See Canvas calendar for the due date.
36 points Total, 2 points per question unless otherwise noted.

1. A study was done to investigate the effectiveness of five methods for the irrigation of blueberry shrubs. Ten farms were included in the study. Each of the five treatments was evaluated at each of the ten farms (with irrigation treatments randomly assigned to plots). The response variable is weight of the harvested fruit. The data is available from Canvas as "**Irrigation.csv**". Note: Be sure to define Farm as.factor!
   A. Calculate the sample size, simple mean and SE for each method (averaging over farms). Include the resulting summary table in your assignment.
   B. Create a bar chart (with SE bars) to summarize the data. Include the resulting graph in your assignment. Assuming your summary table from above is called SumStats (with columns **Method, mean** and **se**), you can use code something like this:
   ```
   Library(ggplot2)
   ggplot(SumStats, aes(x = Method, y = mean)) +
     geom_bar(stat = "identity") + geom_errorbar(aes(ymin =
   mean-se, ymax = mean+se), width = 0.2)
   ```
   C. Fit the RCB model. Inspect the diagnostic plots (Resids vs Fitted and Normal QQplot of Resids), and comment on what you see. Do the assumptions appear to be satisfied? Note: You do not have to include the diagnostic plot in your assignment, just comment on each graph. (**4 pts**)
   D. Continuing with the RCB model from the previous question, include the Type3 ANOVA table in your assignment.
   E. Can we conclude that there is a difference between the irrigation methods? Justify your response with a test statistic and p-value.
   F. Make a conclusion about the effectiveness of the blocking in this example. Justify your response with a test statistic and p-value.
   G. The investigators are interested in which irrigation methods are significantly different from each other. Use `emmeans()` function from the `emmeans` package to get Tukey-adjusted p-values for comparing treatments. Then use this information to create a "cld" display, where methods that are NOT significantly different from each other are given the same number grouping. Hint: Use code something like this:
   ```
   library(emmeans)
   library(multcompView)
   emout <- emmeans(Model1, pairwise ~ Method)
   emout
   cld(emout$emmeans)
   ```
   H. Are the simple means (part A) and emmeans (part G) the same for this analysis? What about the simple SE's (part A) versus SE's returned by emmeans (part G)?
   I. Run the analysis as a one-way ANOVA using just Method in the model. (In practice I would not do this, but try it here for illustration.) Include the ANOVA table in your assignment. How does dfResid compare to the RCB model? How does MSResid compare to the RCB model? (**4 pts**) Hint: Recall that MSResid = SSResid/dfResid.

2. A fertilizer trial on a range grass (blue grama) was conducted in a randomized complete block design. Five fertilizer treatments were randomly assigned to the plots in each of five blocks, but two observations have missing values. The response variable (Y) represents phosphorous. The data is available from Canvas as "**GrassMiss.csv**". Note: Be sure to define Block as.factor!

   A. Calculate the simple mean for each trt (averaging over blocks). Include the resulting summary table in your assignment. Hint: Because of the NA values, it is easiest to use aggregate() here.
   B. Fit the RCB model and include the Type3 ANOVA table in your assignment.
   C. Calculate the emmeans and corresponding confidence intervals for each trt and include them in your assignment. Note that the SE is larger (and CIs are wider) for trts that have missing values.
   D. Are the simple means (part A) and emmeans (part C) the same for this analysis?
   E. Use the coefficient estimates (from the `summary()` output) to compute predicted values for the **two** missing observations. Show your work for full credit. (Note that you can verify these using the `predict()` function.) (**4 pts**)
   F. Verify that the emmean for N50wP is the average of the five predicted values (one from each block) for N50wP. Show your work for full credit.