STAR 511 HW #6 (38 points)

Questions 1 through 8 (Corn Yield): An agricultural study was done to compare the mean yield for 4 varieties of corn (A, B, C, D). There are 8 observations (or reps) for each of the 4 varieties for a total of n = 32 observations. Use alpha = 0.05. The data is available as "CornYield.csv".

1. Construct side-by-side boxplots of the data. From visual inspection, do any groups appear to have produced similar mean yields? Do any appear to have produced different mean yields? (**3 pts**)
2. Do we have statistical evidence of population-level differences in mean yield for the varieties? Conduct a one-way ANOVA analysis.
   A. State the null and alternative hypotheses. (**2 pts**)
   B. Run an appropriate analysis and report the ANOVA table. (**2 pts**)
   C. State the p-value, the statistical decision ("reject" or "fail to reject" H_0), and write a sentence or two explaining what we can conclude, if anything, about corn yields across the four varieties, using our data. (**3 pts**)

3. Use the `plot()` function to generate the diagnostic plots from the model from Q2.
   A. Report both the residuals vs. fitted plot and the QQ plot.)
   B. Briefly discuss what you see in the residuals vs fitted plot: what assumption is being checked, and do you see any indication it's been violated? (**2 pts**)
   C. Briefly discuss what you see in the QQ plot: what assumption is being checked, and do you see any indication it's been violated? (**2 pts**)

4. Run (unadjusted) pairwise comparisons of means. Use `emmeans()` and `pairs()` with `adjust = "none"`. Based on your results, which group means differ significantly and which do not? (**3 pts**)
5. Now use the Tukey adjustment to compare means across groups, and report the R output. Again, state which group means differ significantly and which do not. (**2 pts**)
6. Report a plot of confidence intervals for unadjusted pairwise comparisons. Here is code to do this, assuming your object is called `HW6Q6_placeholder_noadjust` (you probably don't want to use this name; it's just an example): (**2 pts**)

```
plot(pairs(HW6Q6_placeholder_noadjust),horizontal=F)
```

7. Make another plot like the one from Q6 above, this time of Tukey adjusted pairwise comparisons. When comparing this plot to the one in Q6, what aspect of it indicates that a multiple comparisons adjustment has been performed? (**2 pts**)
8. The plots you reported in Q6 and Q7 both show kinds of "95%" confidence intervals, where the "95%" refers to some kind of "success" or "coverage" rate for the procedure used to make the intervals. Explain the difference between what "95%" refers to in questions 6 and 7. (**2 pts**)

Questions 8 through 13 (Power Plants): The data from O&L problem 8.23 **(**ex8-23.txt) concerns reliability of nuclear power plant generators. The data provides Emergency Diesel Generator (EDG) values, giving the number of times the EDGs successfully worked, for t = 7 power plants (A-G).

Notes:

- The original data are in "wide" format, meaning that each group (in this case, each power plant) gets its own column of measured values of the outcome variable (in this case, EDG). You can use code like the following to transpose from wide to long, where "long" format data have all measured values of the outcome variable (EDG) in a single column, followed by a column of values indicating Plant ("A", "B", "C",…).

  This code assumes (1) the original data is called InData after importing and (2) column names are A, B, C….G:

```
library(tidyverse)
Reliability <-
  InData %>%
  pivot_longer(A:G, names_to = "Plant", values_to = "EDG",
values_drop_na = TRUE) %>%
  mutate(Plant = as_factor(Plant))
str(Reliability)
```

- If you don't want to use code, you are welcome to open the data in Excel and use a "cut/paste" approach to put your data into long form: create a new column called "Plant", whose values will be one of the letters "A" through "G", and stack all the EDG values in a single column.

9. Construct a side-by-side boxplot showing EDG values across all Plants (should be on one graph). (**2 pts**)
10. Report the one-way ANOVA table for a model that compares EDG values across Plants. (**2 pts**)
11. Report a plot that assesses the assumption of equal variances, and a plot that assesses the assumption of normal residuals. In this case, both assumptions are violated. For each plot, briefly explain how the plot indicates that an assumption has been violated. (**4 pts**)
12. Perform a Kruskal-Wallace test:
    a. Report the R output from the test (**1 pt**)
    b. Write a sentence or two explaining what the result of the test tells you (**2 pts**)
    c. Suppose someone looks at this analysis and says "you can't do a hypothesis test here, because the assumptions are obviously violated!". Explain why performing the Kruskal-Wallace test is justifiable here, even though there are clear signs of non-normality and non-constant variance. (**2 pts**)