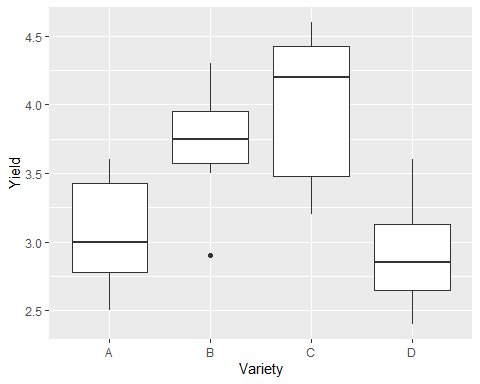
STAR511: HW 6

Megan Sears

# Q1



Groups A and D seem to have similar mean yields. Group C seems to have produced a different mean yield than C, D, and A. However, Group B mean yield is likely most similar to group C.

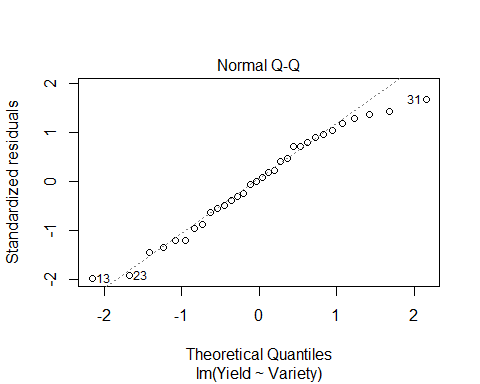
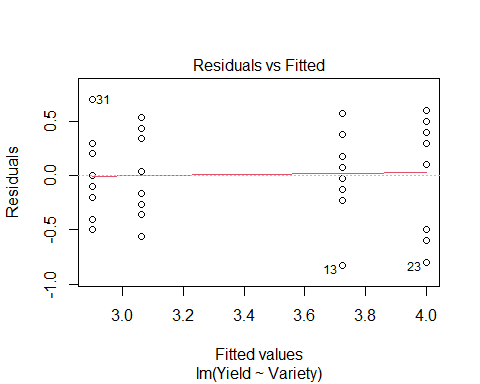
# Q2

A. The null hypothesis is that the mean yields for the varieties are equal, and the alternative hypothesis is that one or more difference exists between the mean yield for the varieties.

## Analysis of Variance Table  
##   
## Response: Yield  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Variety 3 6.6209 2.20698 11.047 5.85e-05 \*\*\*  
## Residuals 28 5.5938 0.19978   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

C. The p-value is 0.0000585. Here, we will reject the null hypothesis since the p-value is less than alpha. There is strong evidence indicating differences among the population-level mean yields for varieties.

# Q3



B. The residuals versus fitted plot is assessing the variance or equality of variance. If there is equality of variance, then in the plot we will see equal scatter among the points. This shows primarily equal variance in the residual diagnostic plot.

C. This plot is used to check the normality of the residuals. Here, the points follow closely along the line indicating the distribution is normal.

# Q4

## contrast estimate SE df t.ratio p.value  
## A - B -0.662 0.223 28 -2.964 0.0061  
## A - C -0.938 0.223 28 -4.195 0.0002  
## A - D 0.163 0.223 28 0.727 0.4732  
## B - C -0.275 0.223 28 -1.231 0.2287  
## B - D 0.825 0.223 28 3.692 0.0010  
## C - D 1.100 0.223 28 4.922 <.0001

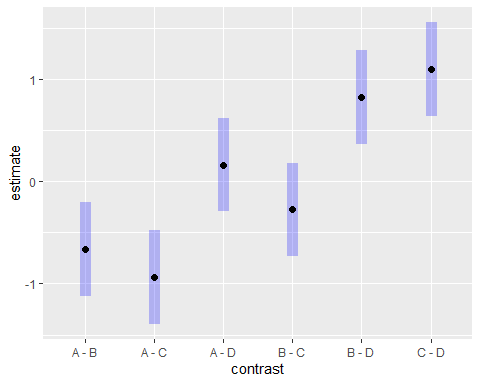
Groups B&C and A&D have p-values greater than 0.05, which indicates the null for those groups should not be rejected and the means are not significantly different. The other groups (A&B, A&C, B&D, C&D) have p-values less than 0.05, which indicates the null for those should be rejected and the means are significantly different.

# Q5

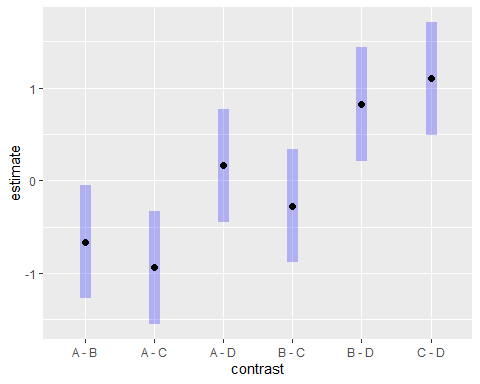
## contrast estimate SE df t.ratio p.value  
## A - B -0.662 0.223 28 -2.964 0.0295  
## A - C -0.938 0.223 28 -4.195 0.0013  
## A - D 0.163 0.223 28 0.727 0.8854  
## B - C -0.275 0.223 28 -1.231 0.6132  
## B - D 0.825 0.223 28 3.692 0.0050  
## C - D 1.100 0.223 28 4.922 0.0002  
##   
## P value adjustment: tukey method for comparing a family of 4 estimates

This produced the same conclusions as above. Groups B&C and A&D have p-values greater than 0.05, which indicates the null for those groups should not be rejected and the means are not significantly different. The other groups (A&B, A&C, B&D, C&D) have p-values less than 0.05, which indicates the null for those should be rejected and the means are significantly different.

# Q6



# Q7

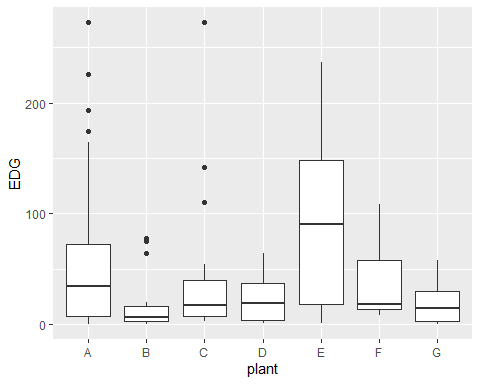


When comparing the 2 plots, I am able to tell that the plot in Q7 has had a multiple comparisons adjustment because it has wider confidence intervals (length of blue bars) than the plot in Q6.

# Q8

In Q6, the unadjusted CIs, represents the interval that covers the mean 95% of the time under repeated sampling for each individual group. In Q7, the Tukey CIs, represents the family-wise or collection of the groups CI. Since the chances of making a type I error for a series of comparisons is greater than individual comparisons, Tukey’s methods adjusts the confidence level for each individual group so that the resulting family-wise confidence level is equal to the value specified. In other words, the individual error rates are lowered to meet the family error rate, which makes the CIs wider.

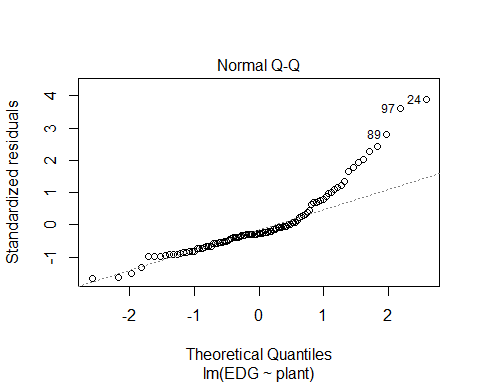
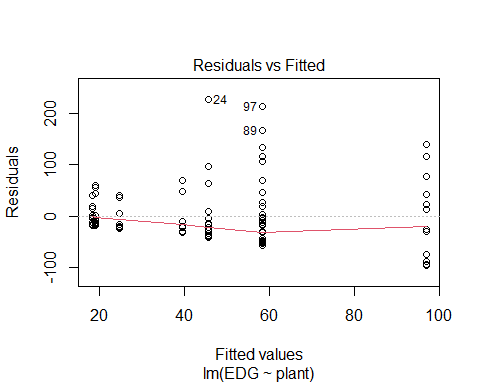
# Q9



# Q10

## Analysis of Variance Table  
##   
## Response: EDG  
## Df Sum Sq Mean Sq F value Pr(>F)   
## plant 6 58745 9790.9 2.6761 0.01912 \*  
## Residuals 96 351233 3658.7   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Q11



The assumption of equality of variance has been violated since the scatter in the plot is not equal. For the lower fitted values, the residual range is smaller. Whereas, values 60 and greater have a larger residual range.

The normality of residuals is violated in the second plot since the points are not on or near the line. The residuals are right skewed.

# Q12

##   
## Kruskal-Wallis rank sum test  
##   
## data: EDG by plant  
## Kruskal-Wallis chi-squared = 12.537, df = 6, p-value = 0.051

B. The p-value is greater than 0.05, which indicates we should fail to reject the null hypothesis. The null hypothesis is the medians are all equal and since we failed to reject the null, the medians are not significantly different.

C. We can do a hypotheses test here even though the assumption are violated since the Kruskal-Wallis test is a non-parametric alternative to the one-way ANOVA F-test. This test does not require the assumption of normality and equality of variances.

# Appendix

#Retain (and do not edit) this code chunk!!!  
library(knitr)  
knitr::opts\_chunk$set(echo = FALSE)  
knitr::opts\_chunk$set(message = FALSE)  
  
library(tidyverse)  
library(here)  
  
#Q1  
corn <- read\_csv(here('./HW6/CornYield.csv')) %>%  
 mutate(Variety = as.factor(Variety))  
  
ggplot(corn, aes(x=Variety, y=Yield)) +  
 geom\_boxplot()  
   
#Q2  
corn\_anov <- lm(Yield ~ Variety, data = corn)  
  
anova(corn\_anov)  
  
#Q3  
plot(corn\_anov, 1)  
plot(corn\_anov, 2)  
  
library('emmeans')  
  
#Q4  
lmfit <- lm(Yield ~ Variety, data = corn)  
emout <- emmeans(lmfit, ~ Variety)  
pairs(emout, adjust = 'none')  
  
#Q5  
pairs(emout)  
  
#Q6  
  
plot(pairs(emout, adjust = 'none'), horizontal = F)  
  
#Q7  
plot(pairs(emout), horizontal = F)  
  
#Q9  
pp <- read\_csv(here('./HW6/ex8-23.txt'), quote = "'")  
  
pp <- pp %>%  
 pivot\_longer(everything(), names\_to = 'plant', values\_to = 'EDG') %>%  
 mutate(plant = as\_factor(plant)) %>%  
 na.omit()  
  
ggplot(pp, aes(x = plant, y = EDG)) +  
 geom\_boxplot()  
  
  
fit <- lm(EDG ~ plant, data = pp)  
anova(fit)  
  
#Q11  
plot(fit, 1)  
plot(fit, 2)  
  
#Q12  
kruskal.test(EDG ~ plant, data = pp)