Assign7

Natalie Schmer

36 points total, 2 points per problem part unless otherwise noted.

## Q1 Power Plants

library(tidyverse)

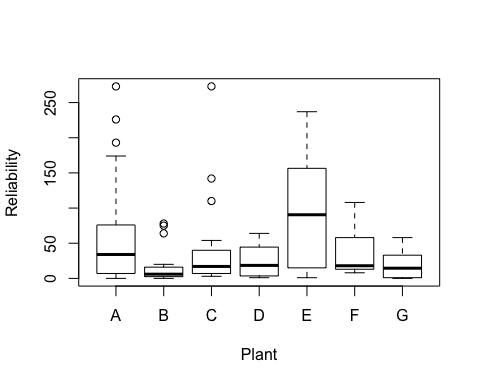
## Warning: package 'tibble' was built under R version 3.6.2

InData <- read.csv("/Users/natalieschmer/Desktop/GitHub/stats\_511/data/ASCII-comma/CH08/ex8-23.TXT", quote = "'")  
str(InData)

## 'data.frame': 103 obs. of 9 variables:  
## $ A : int 28 50 193 55 4 7 174 76 10 0 ...  
## $ B : int 2 11 75 6 1 12 4 6 64 3 ...  
## $ C : int 142 110 3 273 54 32 3 40 23 30 ...  
## $ D : int 64 29 1 3 8 29 4 60 NA NA ...  
## $ E : int 139 21 214 67 174 1 9 2 119 237 ...  
## $ F : int 18 108 9 8 17 88 28 NA NA NA ...  
## $ G : int 0 6 0 16 1 58 13 36 33 19 ...  
## $ Plant : Factor w/ 7 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Reliability: int 28 50 193 55 4 7 174 76 10 0 ...

### 1A. Boxplots

boxplot(Reliability ~ Plant, data = InData)



### 1B. ANOVA (Original Scale)

Fit\_1b = lm(Reliability ~ Plant, data = InData)  
(Ftest\_1b = anova(Fit\_1b))

## Analysis of Variance Table  
##   
## Response: Reliability  
## 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

#ANOVA Table values  
Ftest\_1b$`Sum Sq`

## [1] 58745.34 351233.07

Ftest\_1b$`Mean Sq`

## [1] 9790.889 3658.678

Ftest\_1b$Df

## [1] 6 96

Ftest\_1b$`F value`

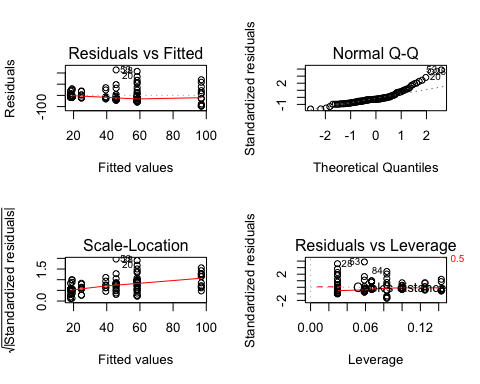
## [1] 2.676073 NA

Ftest\_1b$`Pr(>F)`

## [1] 0.01912123 NA

### 1C. (**4pts**) Diagnostics (Original Scale)

#Set up plot space and plot  
par(mfrow= c(2, 2))  
plot(Fit\_1b)



#Levene's test  
car::leveneTest(Reliability ~ Plant, data = InData, )

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 6 2.3122 0.03963 \*  
## 96   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Shapiro - Wilk  
(shapiro.test(InData$Reliability))

##   
## Shapiro-Wilk normality test  
##   
## data: InData$Reliability  
## W = 0.73347, p-value = 2.2e-12

**The conditions for a one way anova are: a random sample, there are independent observations, there are normally distributed residuals, and there is the equality of variances. Based on the above tests, it is evident that this data is not normal, since the SW test of normality gives a p-value much less than 0.5, the F- statistic is 2.67, and the p = 0.03, and since F > p-value the null hypothesisis rejected that the ratio of variances are equal, and the plots, specifically the Q-Q plot, show non-normal data.**

### 1D. ANOVA (Square Root Transform)

#square root transform EDG   
InData$srt\_trans <- sqrt(InData$Reliability)  
  
Fit\_1d = lm(srt\_trans ~ Plant, data = InData)  
(Ftest\_1d = anova(Fit\_1d))

## Analysis of Variance Table  
##   
## Response: srt\_trans  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Plant 6 252.49 42.082 2.6817 0.0189 \*  
## Residuals 96 1506.46 15.692   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#ANOVA Table values  
Ftest\_1d$`Sum Sq`

## [1] 252.4933 1506.4597

Ftest\_1d$`Mean Sq`

## [1] 42.08222 15.69229

Ftest\_1d$Df

## [1] 6 96

Ftest\_1d$`F value`

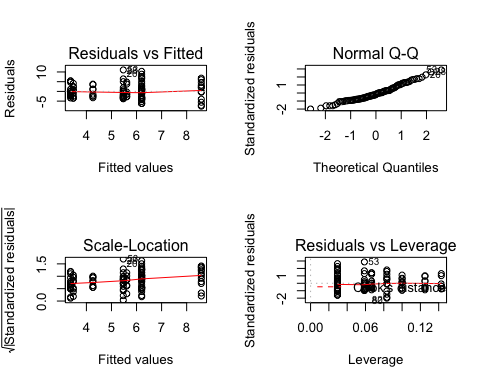
## [1] 2.681713 NA

Ftest\_1d$`Pr(>F)`

## [1] 0.01890474 NA

### 1E. (**4pts**) Diagnostics (Square Root Transform)

#Set up plot space and plot  
par(mfrow= c(2, 2))  
plot(Fit\_1d)



#Levene's test  
car::leveneTest(srt\_trans ~ Plant, data = InData, )

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 6 1.6464 0.1428  
## 96

#Shapiro - Wilk  
(shapiro.test(InData$srt\_trans))

##   
## Shapiro-Wilk normality test  
##   
## data: InData$srt\_trans  
## W = 0.91961, p-value = 1.02e-05

**Even with doing a square root transform, the data still does not appear to satisfy the requirements for a one way anova. The Q-Q plot looks better, and the residuals vs fitted closer to horizontal, but the shapiro p value is still much less than 0.05 and F > p.**

### 1F. Kruskal-Wallis

kruskal.test(Reliability ~ Plant, data = InData)

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

**p = 0.051, since the p-value is technically greater than 0.05, we fail to reject the null hypothesis that the variance in reliability between the plants are significantly different from eachother, it appears the distributions are identical.**

## Q2 Weight Loss

library(tidyverse)  
  
weightloss <- read.csv("/Users/natalieschmer/Desktop/GitHub/stats\_511/data/WeightLoss.csv")  
  
str(weightloss)

## 'data.frame': 50 obs. of 2 variables:  
## $ Trt : Factor w/ 5 levels "A1","A2","A3",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Loss: num 8.7 9.3 8.2 8.3 9 9.4 9.2 12.2 8.5 9.9 ...

WtLoss <- weightloss %>%  
 gather(key = "Trt", value = "Loss") %>%  
 mutate(Trt = as\_factor(Trt)) %>%  
 mutate(Trt = fct\_relevel(Trt, "S"))  
str(WtLoss)

## 'data.frame': 50 obs. of 2 variables:  
## $ Trt : Factor w/ 5 levels "S","A1","A2",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Loss: num 8.7 9.3 8.2 8.3 9 9.4 9.2 12.2 8.5 9.9 ...

### 2A. (**4pts**) Summary Statistics

WtLoss %>%   
 group\_by(Trt) %>%   
 summarise(n = n(),  
 mean = mean(Loss),  
 sd = sd(Loss),  
 se = sd / sqrt(n))

## # A tibble: 5 x 5  
## Trt n mean sd se  
## <fct> <int> <dbl> <dbl> <dbl>  
## 1 S 10 9.27 1.16 0.366  
## 2 A1 10 12.0 0.829 0.262  
## 3 A2 10 11.0 1.12 0.355  
## 4 A3 10 10.3 1.03 0.325  
## 5 A4 10 12.2 0.756 0.239

### 2B. ANOVA

Note: Plots not required but will be shown ins solutions for convenience.

Fit\_2b = lm(Loss ~ Trt, data = WtLoss)  
(Ftest\_2b = anova(Fit\_2b))

## Analysis of Variance Table  
##   
## Response: Loss  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Trt 4 61.618 15.4045 15.681 4.164e-08 \*\*\*  
## Residuals 45 44.207 0.9824   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#ANOVA Table values  
Ftest\_2b$`Sum Sq`

## [1] 61.618 44.207

Ftest\_2b$`Mean Sq`

## [1] 15.4045000 0.9823778

Ftest\_2b$Df

## [1] 4 45

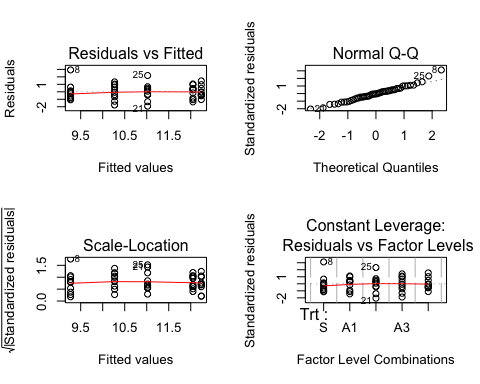
Ftest\_2b$`F value`

## [1] 15.68083 NA

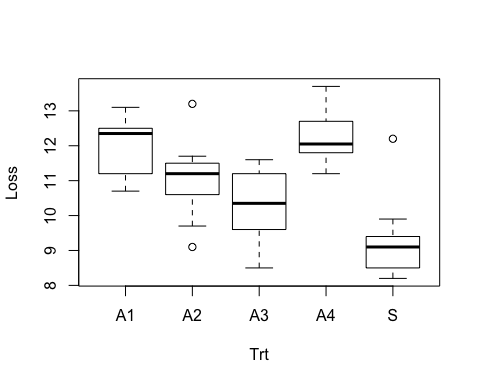
Ftest\_2b$`Pr(>F)`

## [1] 4.16447e-08 NA

par(mfrow= c(2,2))  
plot(Fit\_2b)



boxplot(Loss ~ Trt, weightloss)



### 2C. Unadjusted Pairwise Comparisons

(loss\_2c<- emmeans::emmeans(Fit\_2b, pairwise ~ Trt, adjust = "none"))

## $emmeans  
## Trt emmean SE df lower.CL upper.CL  
## S 9.27 0.313 45 8.64 9.9  
## A1 12.05 0.313 45 11.42 12.7  
## A2 11.02 0.313 45 10.39 11.7  
## A3 10.27 0.313 45 9.64 10.9  
## A4 12.24 0.313 45 11.61 12.9  
##   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df t.ratio p.value  
## S - A1 -2.78 0.443 45 -6.272 <.0001   
## S - A2 -1.75 0.443 45 -3.948 0.0003   
## S - A3 -1.00 0.443 45 -2.256 0.0290   
## S - A4 -2.97 0.443 45 -6.700 <.0001   
## A1 - A2 1.03 0.443 45 2.324 0.0247   
## A1 - A3 1.78 0.443 45 4.016 0.0002   
## A1 - A4 -0.19 0.443 45 -0.429 0.6702   
## A2 - A3 0.75 0.443 45 1.692 0.0976   
## A2 - A4 -1.22 0.443 45 -2.752 0.0085   
## A3 - A4 -1.97 0.443 45 -4.444 0.0001

### 2D. Tukey adjusted Pairwise Comparisons

(loss\_2d<- emmeans::emmeans(Fit\_2b, pairwise ~ Trt))

## $emmeans  
## Trt emmean SE df lower.CL upper.CL  
## S 9.27 0.313 45 8.64 9.9  
## A1 12.05 0.313 45 11.42 12.7  
## A2 11.02 0.313 45 10.39 11.7  
## A3 10.27 0.313 45 9.64 10.9  
## A4 12.24 0.313 45 11.61 12.9  
##   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df t.ratio p.value  
## S - A1 -2.78 0.443 45 -6.272 <.0001   
## S - A2 -1.75 0.443 45 -3.948 0.0024   
## S - A3 -1.00 0.443 45 -2.256 0.1784   
## S - A4 -2.97 0.443 45 -6.700 <.0001   
## A1 - A2 1.03 0.443 45 2.324 0.1563   
## A1 - A3 1.78 0.443 45 4.016 0.0020   
## A1 - A4 -0.19 0.443 45 -0.429 0.9927   
## A2 - A3 0.75 0.443 45 1.692 0.4490   
## A2 - A4 -1.22 0.443 45 -2.752 0.0618   
## A3 - A4 -1.97 0.443 45 -4.444 0.0005   
##   
## P value adjustment: tukey method for comparing a family of 5 estimates

### 2E. Without adjusting for multiple testing compared to Tukey

**Without adjustment, there were 8 significant comparisons, and with adjustment there were 5 significant comparisons.**

### 2F. HSD value

qtukey(0.95, 5, 45)

## [1] 4.018417

### 2G. CLD display

emmeans::CLD(loss\_2d$emmeans)

## Trt emmean SE df lower.CL upper.CL .group  
## S 9.27 0.313 45 8.64 9.9 1   
## A3 10.27 0.313 45 9.64 10.9 12   
## A2 11.02 0.313 45 10.39 11.7 23   
## A1 12.05 0.313 45 11.42 12.7 3   
## A4 12.24 0.313 45 11.61 12.9 3   
##   
## Confidence level used: 0.95   
## P value adjustment: tukey method for comparing a family of 5 estimates   
## significance level used: alpha = 0.05

### 2H. (**4pts**) Dunnett adjusted comparisons

emmeans::emmeans(Fit\_2b, dunnett ~ Trt)

## $emmeans  
## Trt emmean SE df lower.CL upper.CL  
## S 9.27 0.313 45 8.64 9.9  
## A1 12.05 0.313 45 11.42 12.7  
## A2 11.02 0.313 45 10.39 11.7  
## A3 10.27 0.313 45 9.64 10.9  
## A4 12.24 0.313 45 11.61 12.9  
##   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df t.ratio p.value  
## A1 - S 2.78 0.443 45 6.272 <.0001   
## A2 - S 1.75 0.443 45 3.948 0.0010   
## A3 - S 1.00 0.443 45 2.256 0.0961   
## A4 - S 2.97 0.443 45 6.700 <.0001   
##   
## P value adjustment: dunnettx method for 4 tests

**Treatments A1, A2, and A4 appear to result in a significantly higher weight loss than the standard at 0.05 level.**