HW7 KEY

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

Q1 Power Plants

```
library(tidyverse)
InData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH08/ex8-23.txt",quote=" ' ")
str(InData)
```

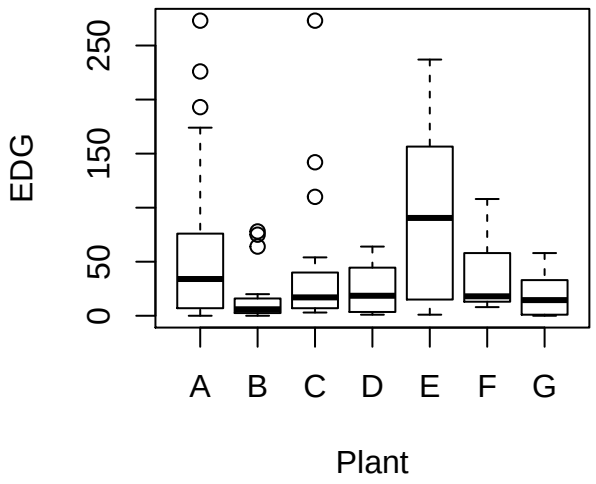
```
## 'data.frame':    34 obs. of  7 variables:
##  $ A: int  28 50 193 55 4 7 174 76 10 0 ...
##  $ B: int  2 11 75 6 1 12 4 6 64 3 ...
##  $ C: int  142 110 3 273 54 32 3 40 23 30 ...
##  $ D: int  64 29 1 3 8 29 4 60 NA NA ...
##  $ E: int  139 21 214 67 174 1 9 2 119 237 ...
##  $ F: int  18 108 9 8 17 88 28 NA NA NA ...
##  $ G: int  0 6 0 16 1 58 13 36 33 19 ...
```

```
Reliability <- gather(InData, key = "Plant", value = "EDG")
Reliability$Plant <- as.factor(Reliability$Plant)
str(Reliability)
```

```
## 'data.frame':    238 obs. of  2 variables:
##  $ Plant: Factor w/ 7 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ EDG  : int  28 50 193 55 4 7 174 76 10 0 ...
```

1A. Boxplots

```
boxplot(EDG ~ Plant, data = Reliability)
```



1B. ANOVA (Original Scale)

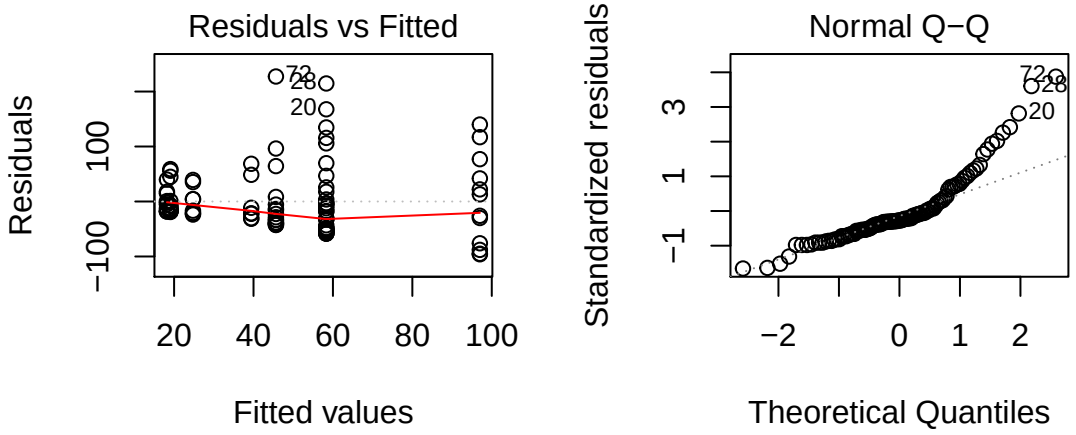
```
Model1 <- lm(EDG ~ Plant, data = Reliability)
anova(Model1)
```

```
## 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
```

1C. (**4pts**) Diagnostics (Original Scale)
Plot of residuals vs fitted values shows megaphone shape. Levene's test p-value = 0.039.
Based on this information, assumption of equal variance is NOT met.

QQplot shows strong curvature. Shapiro-Wilk test p-value < 0.001.
Based on this information, assumption of normality is NOT met.
Note: Plots not required but shown here for convenience.

```
library(car)
par(mfrow=c(1,2))
plot(Model1, which = 1:2)
```




```
leveneTest(EDG ~ Plant, data = Reliability)
```

```
## 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.test(Model1$residuals)
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  Model1$residuals
## W = 0.85649, p-value = 1.444e-08
```

1D. ANOVA (Square Root Transform)

```
Reliability$sqrtEDG <- sqrt(Reliability$EDG)
Model2 <- lm(sqrtEDG ~ Plant, data = Reliability)
anova(Model2)
```

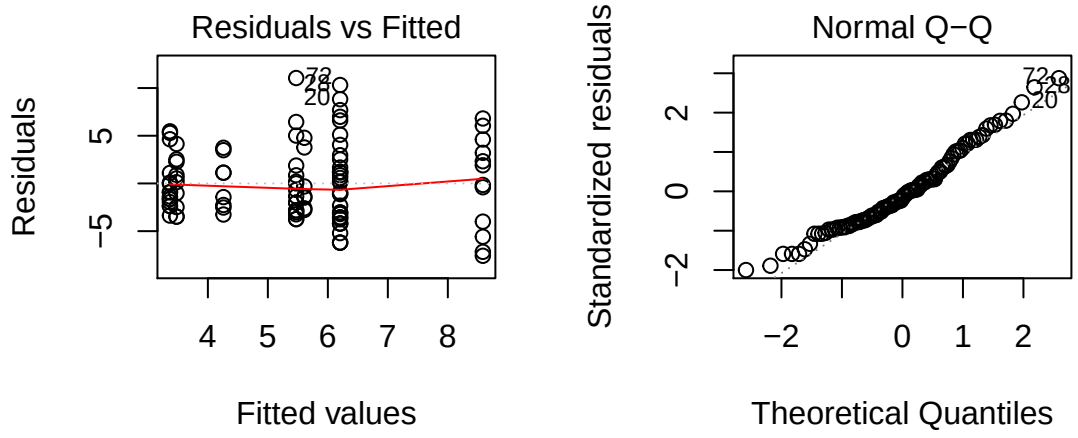
```
## Analysis of Variance Table
##
## Response: sqrtEDG
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Plant      6  252.49  42.082  2.6817 0.0189 *
## Residuals 96 1506.46  15.692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1E. (**4pts**) Diagnostics (Square Root Transform)
Plot of residuals vs fitted values looks better (but not perfect). Levene's test p-value = 0.1428.
Mixed evidence on equality of variances.

QQplot looks much better, but Shapiro-Wilk test p-value = 0.0164.
Again, mixed evidence on normality.

Note: Plots not required but shown here for convenience.

```
par(mfrow=c(1,2))
plot(Model2, which = 1:2)
```




```
leveneTest(sqrtEDG ~ Plant, data = Reliability)
```

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

```
shapiro.test(Model2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Model2$residuals
## W = 0.96913, p-value = 0.01645
```

1F. Kruskal-Wallis p-value = 0.051.
Fail to Reject H0 (but just barely!).

```
kruskal.test(EDG ~ Plant, data = Reliability)
```

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

1G. (**4pts**) Dunn's comparisons
Comparing other plants to G, only plant E shows evidence of a difference from plant G (at the 0.05 level).

```
library(dunn.test)
dunn.test(x = Reliability$EDG, g = Reliability$Plant)
```

```
##   Kruskal-Wallis rank sum test
##
## data: x and group
## Kruskal-Wallis chi-squared = 12.5372, df = 6, p-value = 0.05
##
##
##                             Comparison of x by group
##                                   (No adjustment)
## Col Mean-|
## Row Mean |          A          B          C          D          E          F
## ---------+----------------------------------------------------------------------
##        B |   2.286906
##          |    0.0111*
##          |
##        C |   0.455880  -1.618775
##          |     0.3242     0.0527
##          |
##        D |   0.994053  -0.726924   0.595221
##          |     0.1601     0.2336     0.2758
##          |
##        E |  -1.385870  -3.031778  -1.593366  -1.875307
##          |     0.0829    0.0012*     0.0555     0.0304
##          |
##        F |  -0.118471  -1.656043  -0.411029  -0.849750   0.875049
##          |     0.4528     0.0489     0.3405     0.1977     0.1908
##          |
##        G |   1.596615  -0.329450   1.101424   0.387375   2.428228   1.265278
##          |     0.0552     0.3709     0.1354     0.3492    0.0076*     0.1029
##
## alpha = 0.05
## Reject Ho if p <= alpha/2
```

Q2 Weight Loss

```
library(tidyverse)
InData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH09/ex9-13.txt",quote=" ' ")
str(InData)
```

```
## 'data.frame':    10 obs. of  5 variables:
##  $ A1: num  12.4 10.7 11.9 11 12.4 12.3 13 12.5 11.2 13.1
##  $ A2: num  9.1 11.5 11.3 9.7 13.2 10.7 10.6 11.3 11.1 11.7
##  $ A3: num  8.5 11.6 10.2 10.9 9 9.6 9.9 11.3 10.5 11.2
##  $ A4: num  12.7 13.2 11.8 11.9 12.2 11.2 13.7 11.8 12.2 11.7
##  $ S : num  8.7 9.3 8.2 8.3 9 9.4 9.2 12.2 8.5 9.9
```

```
WtLoss <- InData %>%
          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",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Loss: num  12.4 10.7 11.9 11 12.4 12.3 13 12.5 11.2 13.1 ...
```

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

```
SumStats <- WtLoss %>%
            group_by(Trt) %>%
            summarise(n = n(),
                    mean = mean(Loss),
                    sd = sd(Loss),
                    se = sd/sqrt(n))
SumStats
```

```
## # 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
```

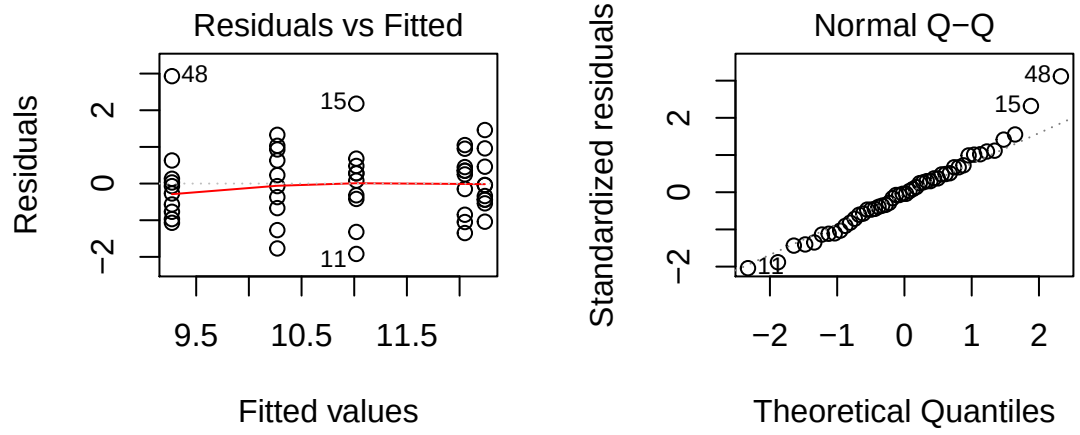
‘‘‘

2B. ANOVA Note: Plots not required but shown here for convenience.

```
Model <- lm(Loss ~ Trt, data = WtLoss)
anova(Model)
```

```
## 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
```

```r
par(mfrow=c(1,2))
plot(Model, which = 1:2)
```



2C. Unadjusted Pairwise Comparisons

```r
library(emmeans)
emout <- emmeans(Model, ~ Trt)
pairs(emout, adjust = "none")
```

```
##  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

```r
pairs(emout)
```

```
##  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, **8** comparisons have p-values < 0.05.
After Tukey adjustment, **5** comparisons have p-values < 0.05.

2F. HSD value

```
qtukey(0.95, 5, 45)*sqrt(0.9824)*sqrt(1/10)
```

```
## [1] 1.259503
```

2G. CLD display

```
CLD(emout)
```

```
##  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
A1, A2, A4 show evidence of greater weight loss than S (at the 0.05 level).

```
emout2 <- emmeans(Model, dunnett ~ Trt)
emout2$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
```