HW6 KEY

40 points total, 2 points per problem part unless otherwise noted.
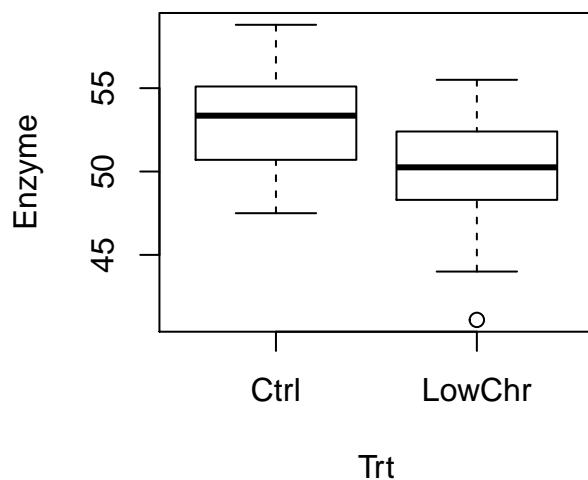
Q1 Rat Liver

```r
RatLiver <- read.csv("c:/hess/STAT511_FA11/HW_2019/HW6/RatLiver.csv")
str(RatLiver)
```

```
## 'data.frame':    24 obs. of  2 variables:
##  $ Trt   : Factor w/ 2 levels "Ctrl","LowChr": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Enzyme: num  44 48.5 50.7 45 53 52.7 51.8 49.8 48.3 55.5 ...
```

1A. Boxplots

```r
boxplot(Enzyme ~ Trt, data = RatLiver)
```


1B. (**4pts**) F-test (Variances)

```r
Test1B <- var.test(Enzyme ~ Trt, data = RatLiver)
Test1B
```

```
##
##  F test to compare two variances
##
## data:  Enzyme by Trt
## F = 0.78978, num df = 9, denom df = 13, p-value = 0.7373
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2384571 3.0253182
## sample estimates:
## ratio of variances
```

```
##          0.7897775
```

H0: $\sigma_1^2 = \sigma_2^2$
TS: F = 0.79
p-value = 0.737
Fail to Reject H0; cannot conclude the variances are different.

1C. Levene's Test (Variances)

```
library(car)
Test1C <- leveneTest(Enzyme ~ Trt, data = RatLiver, center = "median")
Test1C
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##       Df F value Pr(>F)
## group  1   0.176 0.6789
##       22
```

p-value = 0.6789
Fail to Reject H0; cannot conclude that the variances are different.

1D. The conclusions from both tests are the same. We do not have evidence that the variances are different. Use pooled variance t-test.

1E. (**4pts**) t-test (Means)

```
Test1E <- t.test(Enzyme ~ Trt, data = RatLiver, var.equal = TRUE)
Test1E
```

```
##
##   Two Sample t-test
##
## data:  Enzyme by Trt
## t = 2.1709, df = 22, p-value = 0.041
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.1505995 6.5894005
## sample estimates:
##    mean in group Ctrl mean in group LowChr
##               52.87                  49.50
```

H0: $\mu_1 = \mu_2$
TS: t = 2.171
p-value = 0.041
Reject H0; conclude there is a difference between the means.

1F. F-test (Means)

```
Test1F <- lm(Enzyme ~ Trt, data = RatLiver)
anova(Test1F)
```

```
## Analysis of Variance Table
##
## Response: Enzyme
##            Df  Sum Sq Mean Sq F value Pr(>F)
## Trt         1  66.249  66.249  4.7127  0.041 *
## Residuals 22 309.261  14.057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q2 Corn Yield

```
Corn<-read.csv("c:/hess/STAT511_FA11/HW_2019/HW6/CornYield.csv")
str(Corn)
```

```
## 'data.frame':    32 obs. of  2 variables:
##  $ Variety: Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 2 2 ...
##  $ Yield  : num  2.5 3.6 2.8 2.7 3.1 3.4 2.9 3.5 3.6 3.9 ...
```
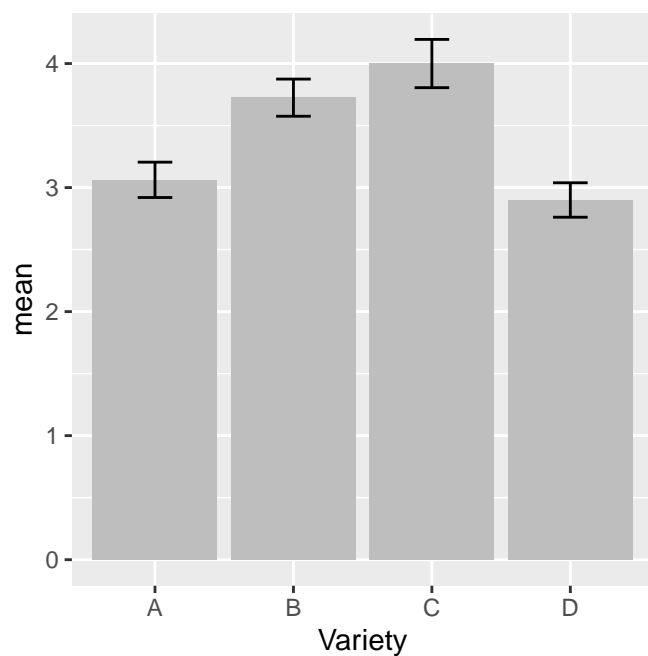
2A. (**4pts**) Bar Chart (Summary Stats not required)

```
library(tidyverse)
SumStats <- summarise(group_by(Corn, Variety),
                n    = n(),
                mean = mean(Yield),
                sd   = sd(Yield),
                se   = sd / sqrt(n) )
SumStats
```

```
## # A tibble: 4 x 5
##   Variety     n  mean    sd    se
##   <fct>   <int> <dbl> <dbl> <dbl>
## 1 A           8  3.06 0.403 0.143
## 2 B           8  3.72 0.423 0.150
## 3 C           8  4    0.550 0.195
## 4 D           8  2.9  0.393 0.139
```

```
ggplot(aes(x = Variety, y = mean), data = SumStats) +
  geom_col(fill = "grey") +
  geom_errorbar(aes(ymin = mean-se, ymax = mean+se), width = .25)
```


"'

2B. ANOVA (**4pts**)

```r
Model <- lm(Yield ~ Variety, data = Corn)
anova(Model)
```

```
## 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
```

H0: $\mu_A = \mu_B = \mu_C = \mu_D$
TS: F = 11.047
p-value < 0.001
Reject H0; conclude not all the means are the same.

2C. (**4pts**) Pairwise Comparisons

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

```
##  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
```

2D. LSD value

```r
qt(0.975,28)*sqrt(0.1998)*sqrt(2/8)
```

```
## [1] 0.4578087
```

2E. (**4pts**) CLD/Lines display

```r
CLD(emout, adjust = "none")
```

```
##  Variety emmean    SE df lower.CL upper.CL .group
##  D         2.90 0.158 28     2.58     3.22  1
##  A         3.06 0.158 28     2.74     3.39  1
##  B         3.73 0.158 28     3.40     4.05   2
##  C         4.00 0.158 28     3.68     4.32   2
##
## Confidence level used: 0.95
## significance level used: alpha = 0.05
```

2F.
There is NOT evidence that Variety D is different from Variety A.
There is NOT evidence that Variety B is different from Variety C.
There is evidence of differences for all other comparisons.
(Indicating that varieties B and C have higher mean yield than varieties A and D.)

2G. (**4pts**) Diagnostic Plots
Plot of Residuals vs Fitted shows equal scatter supporting equal variance.

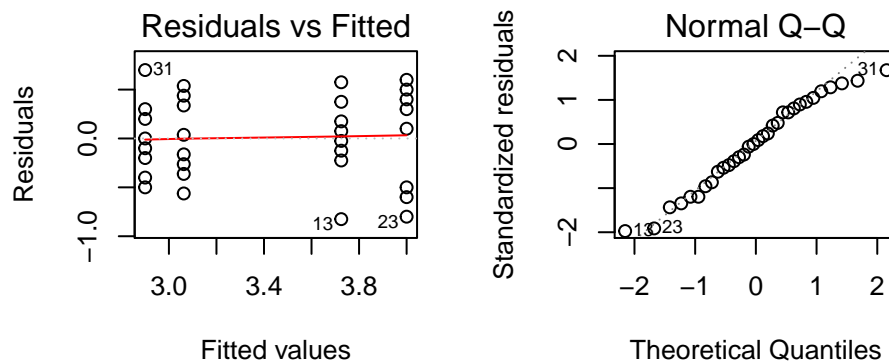QQplot of Residuals approximately linear supporting normality.

Extra Output (Not Required)

```r
par(mfrow = c(2,2))
plot(Model, c(1:2))
hist(Model$residuals)
library(car)
leveneTest(Yield ~ Variety, data = Corn)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.5647 0.6428
##       28
```

```r
shapiro.test(Model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Model$residuals
## W = 0.96814, p-value = 0.4497
```