HW6 KEY

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

Irrigation #1 (Balanced RCB)

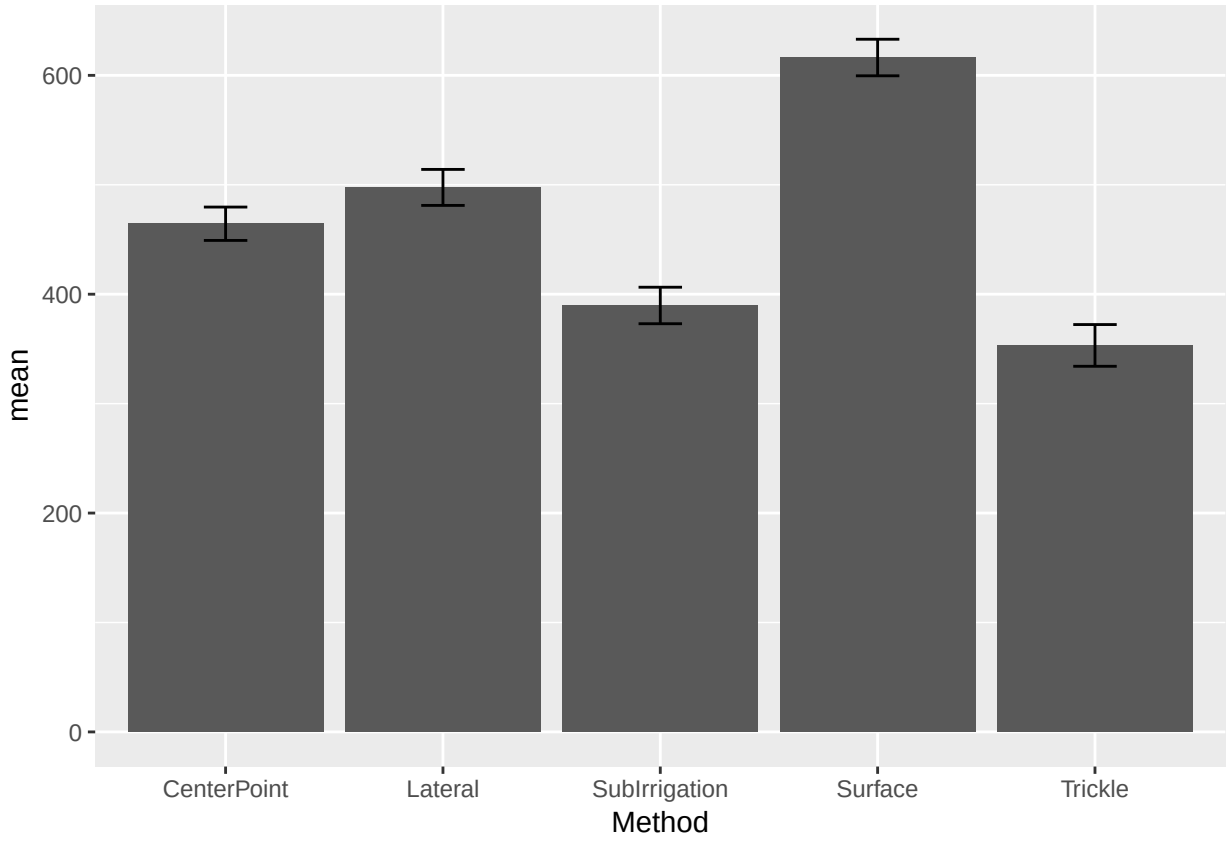
A. Summary Statistics

```
library(dplyr)
library(ggplot2)
library(car)
library(emmeans)
library(multcompView)
Irrigation <- read.csv("C:/hess/STAT512/HW_2018/HW6/Irrigation.csv")
#str(Irrigation)
Irrigation$Farm <- as.factor(Irrigation$Farm)
SumStats <- summarize(group_by(Irrigation, Method),
                  n = n(),
                  mean = mean(Weight),
                  sd = sd(Weight),
                  se = sd/sqrt(n))
SumStats
```

```
## # A tibble: 5 x 5
##   Method            n  mean    sd    se
##   <fct>         <int> <dbl> <dbl> <dbl>
## 1 CenterPoint      10  464.  48.2  15.3
## 2 Lateral          10  498.  52.0  16.4
## 3 SubIrrigation    10  390.  52.7  16.7
## 4 Surface          10  616.  52.8  16.7
## 5 Trickle          10  353.  60.3  19.1
```

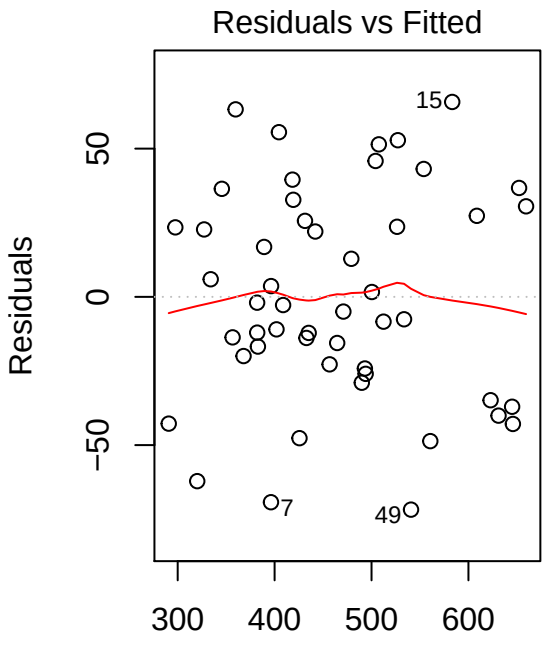
B. Bar Chart

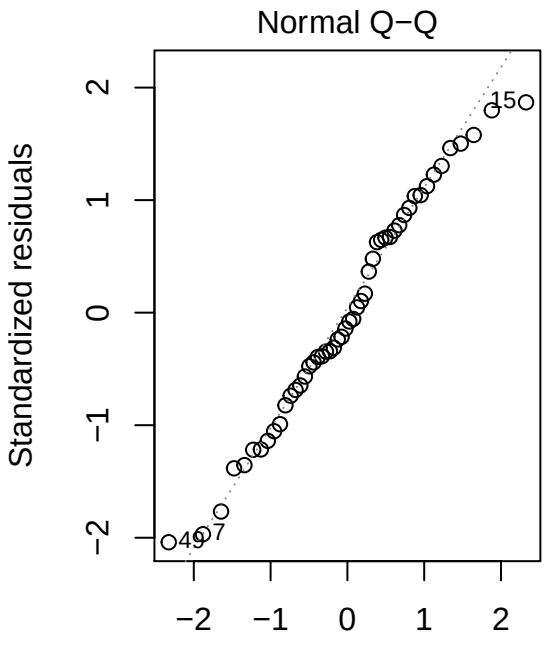
```
ggplot(SumStats, aes(x = Method, y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = mean-se, ymax=mean+se), width = 0.2)
```



C. (**4 pts**) Full credit for anything that looks reasonable. Diagnostic plots look good. Assumptions appear to be satisfied. Resids vs Fitted shows equal scatter (supporting assumption of equal variance). QQplot of residuals roughly linear (supporting assumption of normality). Plots not required for credit, but shown here for completeness.

```
Model1 <- lm(Weight ~ Method + Farm, data = Irrigation)
par(mfrow=c(1,2))
plot(Model1, which = c(1,2))
```





D. Type3 ANOVA table

```
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Weight
##             Sum Sq Df  F value    Pr(>F)
## (Intercept) 577042  1 335.7767 < 2.2e-16 ***
## Method      421213  4  61.2751 1.434e-15 ***
## Farm         66312  9   4.2874 0.0007685 ***
## Residuals    61867 36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E. We can conclude there is a difference between means for the methods. F = 61.28, p-value < 0.0001.

F. We can conclude that the blocking was effective. F = 4.29, p-value = 0.0008.

G. CLD Display

```
emout <- emmeans(Model1, pairwise ~ Method)
cld(emout$emmeans)
```

```
##  Method         emmean   SE df lower.CL upper.CL .group
##  Trickle           353 13.1 36      327      380  1
##  SubIrrigation     390 13.1 36      363      416  1
##  CenterPoint       464 13.1 36      438      491   2
##  Lateral           498 13.1 36      471      524   2
```

```
##  Surface          616 13.1 36      590      643    3
##
## Results are averaged over the levels of: Farm
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 5 estimates
## significance level used: alpha = 0.05
```

H. The simple means and lsmeans **are the same** for this analysis due to balance (no missing data). Even with balance, the simple and model based SE's **will not be the same**.

Note: A "model-based" SE assuming a common variance and accounting for blocking is returned by emmeans: SE = sigmahat/sqrt(n) = sqrt(MSResid)/sqrt(n) = sqrt(61867/36)/sqrt(10) = 13.109

I. (**4 pts**) One-way ANOVA
dfResid is higher for the one-way ANOVA (45 vs 36).
MSResid is higher for the one-way ANOVA (2848 vs 1719).

```
Model2 <- lm(Weight ~ Method, data = Irrigation)
anova(Model2)
```

```
## Analysis of Variance Table
##
## Response: Weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Method     4 421213  105303  36.969 1.096e-13 ***
## Residuals 45 128179    2848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Grass #2 (Unbalanced RCB)

A. Simple Means

```
library(car)
library(emmeans)
Grass <- read.csv("C:/hess/STAT512/HW_2018/HW6/GrassMiss.csv")
#str(Grass)
Grass$Block <- as.factor(Grass$Block)
aggregate(Y ~ Trt, data = Grass, FUN = mean)
```

```
##      Trt      Y
## 1   Ctrl 2.0450
## 2   N100 1.8780
## 3 N100wP 2.3340
## 4    N50 2.0420
## 5  N50wP 2.4525
```

B. Type3 ANOVA Table

```
Model1 <- lm(Y ~ Trt + Block, data = Grass)
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Y
##              Sum Sq Df   F value    Pr(>F)
## (Intercept) 10.1078  1 1566.5568 8.996e-16 ***
```

```
## Trt          0.9651  4   37.3924 2.490e-07 ***
## Block        0.0333  4    1.2911    0.3204
## Residuals    0.0903 14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C. Emmeans with CIs

```
emmeans(Model1, ~ Trt)
```

```
##  Trt    emmean     SE df lower.CL upper.CL
##  Ctrl     2.05 0.0412 14     1.97     2.14
##  N100     1.88 0.0359 14     1.80     1.96
##  N100wP   2.33 0.0359 14     2.26     2.41
##  N50      2.04 0.0359 14     1.96     2.12
##  N50wP    2.45 0.0412 14     2.36     2.54
##
## Results are averaged over the levels of: Block
## Confidence level used: 0.95
```

D. No, simple means and emmeans are the NOT same for this analysis due to missing data.
NOTE: This question not graded. For this question the values work out to be very close (after rounding, they are the same to the second decimal place).

E. (**4 pts**) Predicted values for missing observations
Note: Need to show work for full credit.
Block 3, N50wP: $2.428 = 2.02059 + 0.394 + 0.01365$
Block 5, Ctrl: $2.0882 = 2.02059 + 0 + 0.06765$

```
summary(Model1)
```

```
##
## Call:
## lm(formula = Y ~ Trt + Block, data = Grass)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13859 -0.03977 -0.01894  0.03624  0.16541
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.02059    0.05105  39.580 9.00e-16 ***
## TrtN100     -0.17565    0.05464  -3.215 0.006238 **
## TrtN100wP    0.28035    0.05464   5.131 0.000153 ***
## TrtN50      -0.01165    0.05464  -0.213 0.834281
## TrtN50wP     0.39400    0.05866   6.716 9.84e-06 ***
## Block2      -0.00400    0.05080  -0.079 0.938357
## Block3       0.01365    0.05464   0.250 0.806405
## Block4       0.08800    0.05080   1.732 0.105202
## Block5       0.06765    0.05464   1.238 0.236075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08033 on 14 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:   0.92,  Adjusted R-squared:  0.8744
## F-statistic: 20.14 on 8 and 14 DF,  p-value: 2.018e-06
```

```
Temp <- data.frame(Grass, Yhat = predict(Model1, newdata = Grass))
Temp
```

```
##    Block    Trt    Y     Yhat
## 1      1   Ctrl 2.03 2.020588
## 2      1    N50 1.99 2.008941
## 3      1   N100 1.93 1.844941
## 4      1  N50wP 2.38 2.414588
## 5      1 N100wP 2.26 2.300941
## 6      2   Ctrl 2.09 2.016588
## 7      2    N50 2.04 2.004941
## 8      2   N100 1.79 1.840941
## 9      2  N50wP 2.42 2.410588
## 10     2 N100wP 2.23 2.296941
## 11     3   Ctrl 1.99 2.034235
## 12     3    N50 2.04 2.022588
## 13     3   N100 1.72 1.858588
## 14     3  N50wP   NA 2.428235
## 15     3 N100wP 2.48 2.314588
## 16     4   Ctrl 2.07 2.108588
## 17     4    N50 2.04 2.096941
## 18     4   N100 2.00 1.932941
## 19     4  N50wP 2.56 2.502588
## 20     4 N100wP 2.36 2.388941
## 21     5   Ctrl   NA 2.088235
## 22     5    N50 2.10 2.076588
## 23     5   N100 1.95 1.912588
## 24     5  N50wP 2.45 2.482235
## 25     5 N100wP 2.34 2.368588
```

F. Emmean for N50wP
Note: Need to show work for full credit.
$2.4476 = (2.414588 + 2.410588 + 2.428235 + 2.502588 + 2.482235)/5$