Assignment #6

1 Irrigation of Blueberries

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
library(emmeans)
berries <- read.csv("~/Dropbox/STAT512/Assigns/Assign6/Irrigation.csv")
str(berries)
```

```
## 'data.frame':    50 obs. of  3 variables:
##  $ Method: Factor w/ 5 levels "CenterPoint",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ Farm  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Weight: int  350 370 460 452 343 340 327 378 419 458 ...
```

```r
#berries
berries$Farm <- as.factor(berries$Farm)
```

A

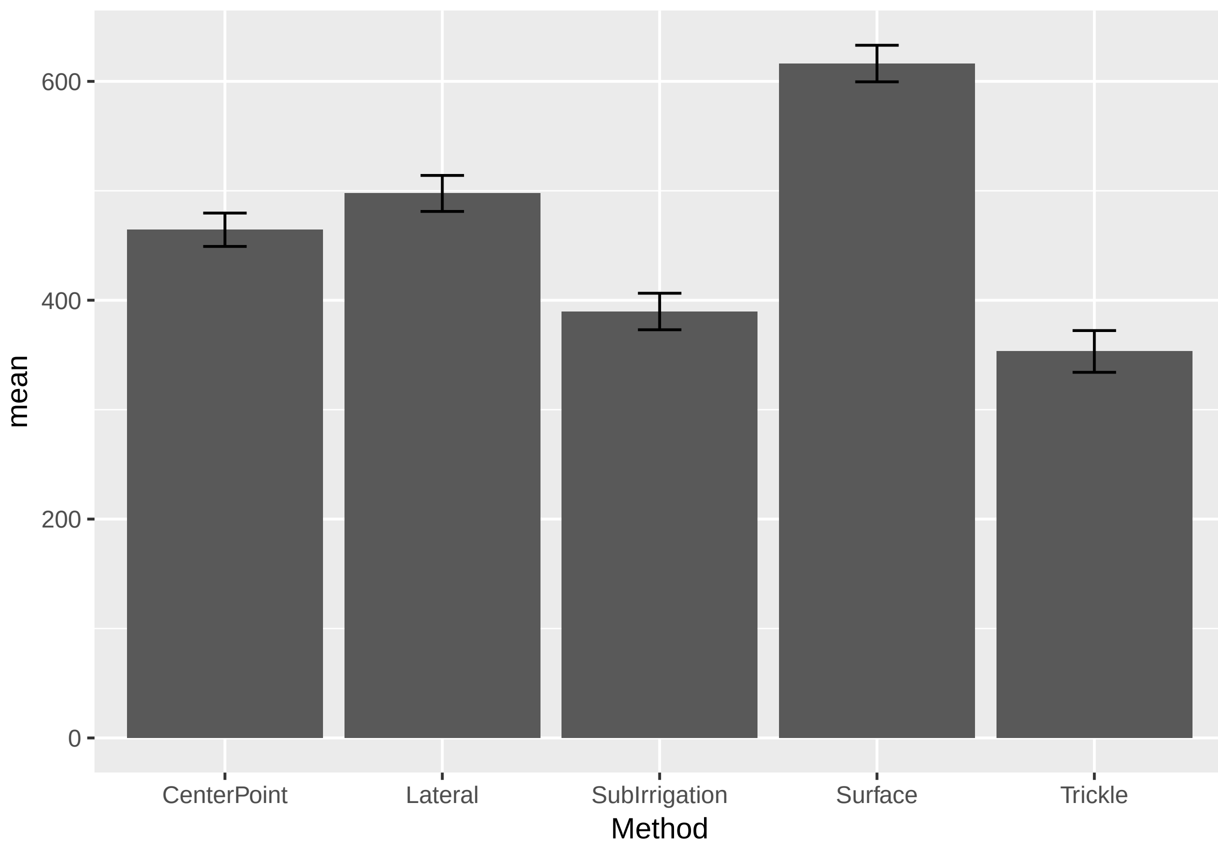
```r
#with round function
SumStats <- summarize(group_by(berries, Method),
                      n = n(),
                      mean = round(mean(Weight),1),
                      sd = round(sd(Weight),3),
                      se = round(sd/sqrt(n),3))
SumStats
```

```
## # 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

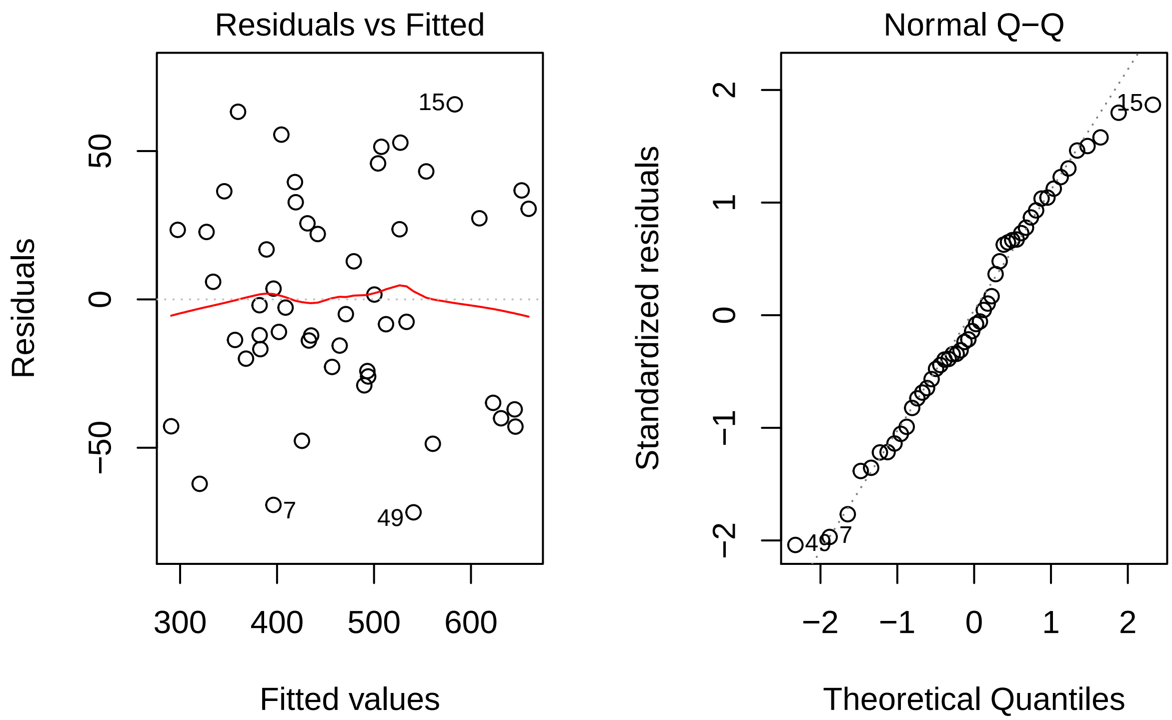
```
ggplot(SumStats, aes(x = Method, y = mean)) +
 geom_bar(stat = "identity") + geom_errorbar(aes(ymin = mean-se, ymax=mean+se), width = 0.2)
```



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 = berries)
par(mfrow=c(1,2))
plot(Model1, which = c(1,2))
```



D

```r
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
```

```r
anova(Model1)
```

```
## Analysis of Variance Table
##
## Response: Weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Method     4 421213  105303 61.2751 1.434e-15 ***
## Farm       9  66312    7368  4.2874 0.0007685 ***
## Residuals 36  61867    1719
## ---
## 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

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

```
## $emmeans
##  Method       emmean       SE df lower.CL upper.CL
##  CenterPoint   464.4 13.10927 36 437.8132 490.9868
##  Lateral       497.6 13.10927 36 471.0132 524.1868
##  SubIrrigation 389.7 13.10927 36 363.1132 416.2868
##  Surface       616.3 13.10927 36 589.7132 642.8868
##  Trickle       353.2 13.10927 36 326.6132 379.7868
##
## Results are averaged over the levels of: Farm
## Confidence level used: 0.95
##
## $contrasts
##  contrast                    estimate       SE df t.ratio p.value
##  CenterPoint - Lateral          -33.2 18.53931 36  -1.791  0.3944
##  CenterPoint - SubIrrigation     74.7 18.53931 36   4.029  0.0024
##  CenterPoint - Surface         -151.9 18.53931 36  -8.193  <.0001
##  CenterPoint - Trickle          111.2 18.53931 36   5.998  <.0001
##  Lateral - SubIrrigation        107.9 18.53931 36   5.820  <.0001
##  Lateral - Surface             -118.7 18.53931 36  -6.403  <.0001
##  Lateral - Trickle              144.4 18.53931 36   7.789  <.0001
##  SubIrrigation - Surface       -226.6 18.53931 36 -12.223  <.0001
##  SubIrrigation - Trickle         36.5 18.53931 36   1.969  0.3014
##  Surface - Trickle              263.1 18.53931 36  14.191  <.0001
##
## Results are averaged over the levels of: Farm
## P value adjustment: tukey method for comparing a family of 5 estimates
```

```r
CLD(emout)
```

```
## Warning in CLD.emm_list(emout): `CLD()` called with a list of 2 objects.
## Only the first one was used.
```

```
##  Method       emmean       SE df lower.CL upper.CL .group
##  Trickle       353.2 13.10927 36 326.6132 379.7868  1
##  SubIrrigation 389.7 13.10927 36 363.1132 416.2868  1
##  CenterPoint   464.4 13.10927 36 437.8132 490.9868   2
##  Lateral       497.6 13.10927 36 471.0132 524.1868   2
```

```
##  Surface        616.3 13.10927 36 589.7132 642.8868    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 = sigma/sqrt(n) = sqrt(MSResid)/sqrt(n) = sqrt(61867/36)/sqrt(10) = 13.109

I

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 = berries)
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
```

2 Missing Data with Grasses

```
Grass <- read.csv("~/Dropbox/STAT512/Assigns/Assign6/GrassMiss.csv")
Grass <- read.csv("~/Dropbox/STAT512/Assigns/Assign6/GrassMiss.csv")
str(Grass)
```

```
## 'data.frame':    25 obs. of  3 variables:
##  $ Block: int  1 1 1 1 1 2 2 2 2 2 ...
##  $ Trt  : Factor w/ 5 levels "Ctrl","N100",..: 1 4 2 5 3 1 4 2 5 3 ...
##  $ Y    : num  2.03 1.99 1.93 2.38 2.26 2.09 2.04 1.79 2.42 2.23 ...
```

```
#Grass
Grass$Block <- as.factor(Grass$Block)
```

A

```r
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
```

```r
#aggregate(Y ~ Block, data = Grass, FUN = mean)
#mean(Grass$Y, na.rm = T)
```

B

```r
#this is default R parameterization, more on this with factorial designs
options(contrasts=c("contr.treatment","contr.poly"))
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
```

```r
#plot(Model1)
```

C

The emmeans for the control and for N50wP are different because each of those treatment groups have a missing value.

```r
emm_options(opt.digits = F)
emmeans(Model1, ~ Trt)
```

```
##  Trt      emmean         SE df lower.CL upper.CL
##  Ctrl   2.053647 0.04117405 14 1.965337 2.141957
##  N100   1.878000 0.03592289 14 1.800953 1.955047
##  N100wP 2.334000 0.03592289 14 2.256953 2.411047
##  N50    2.042000 0.03592289 14 1.964953 2.119047
##  N50wP  2.447647 0.04117405 14 2.359338 2.535957
##
## 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.

```
emmeans(Model1, pairwise ~ Trt)$contrasts
```

```
##  contrast        estimate         SE df t.ratio p.value
##  Ctrl - N100     0.1756471 0.05464207 14   3.215  0.0418
##  Ctrl - N100wP  -0.2803529 0.05464207 14  -5.131  0.0012
##  Ctrl - N50      0.0116471 0.05464207 14   0.213  0.9995
##  Ctrl - N50wP   -0.3940000 0.05866183 14  -6.716  0.0001
##  N100 - N100wP  -0.4560000 0.05080263 14  -8.976  <.0001
##  N100 - N50     -0.1640000 0.05080263 14  -3.228  0.0408
##  N100 - N50wP   -0.5696471 0.05464207 14 -10.425  <.0001
##  N100wP - N50    0.2920000 0.05080263 14   5.748  0.0004
##  N100wP - N50wP -0.1136471 0.05464207 14  -2.080  0.2813
##  N50 - N50wP    -0.4056471 0.05464207 14  -7.424  <.0001
##
## Results are averaged over the levels of: Block
## P value adjustment: tukey method for comparing a family of 5 estimates
```

E

Hand Calculate
Block 3, N50wP:
$2.428 = 2.02059 + 0.394 + 0.01365$ Block 5, Ctrl:
$2.0882 = 2.02059 + 0 + 0.06765$

Coeffs from Model

```
#summary(Model1)
coeffs <- summary(Model1)$coeff
#emmean TrtN50wP Block 3
coeffs[1]+coeffs[5]+coeffs[7]
```

```
## [1] 2.428235
```

```
#emmean Control Block 5
coeffs[1]+coeffs[9]
```

```
## [1] 2.088235
```

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

F.

LSmean for N50wP Hand Calculate $2.4476 = (2.414588 + 2.410588 + 2.428235 + 2.502588 + 2.482235)/5$

```
aggregate(Yhat ~ Trt, data = Temp, FUN = mean)[5,2]
```

```
## [1] 2.447647
```