STAT 512 – Assignment 6

Vignesh M. Pagadala
Vignesh.Pagadala@ColoState.Edu

1. A study was done to investigate the effectiveness of five methods for the irrigation of blueberry shrubs. Ten farms were included in the study. Each of the five treatments was evaluated at each of the ten farms (with irrigation treatments randomly assigned to plots). The response variable is weight of the harvested fruit. The data is available from Canvas as "Irrigation.csv". Note: Be sure to define Farm as.factor!

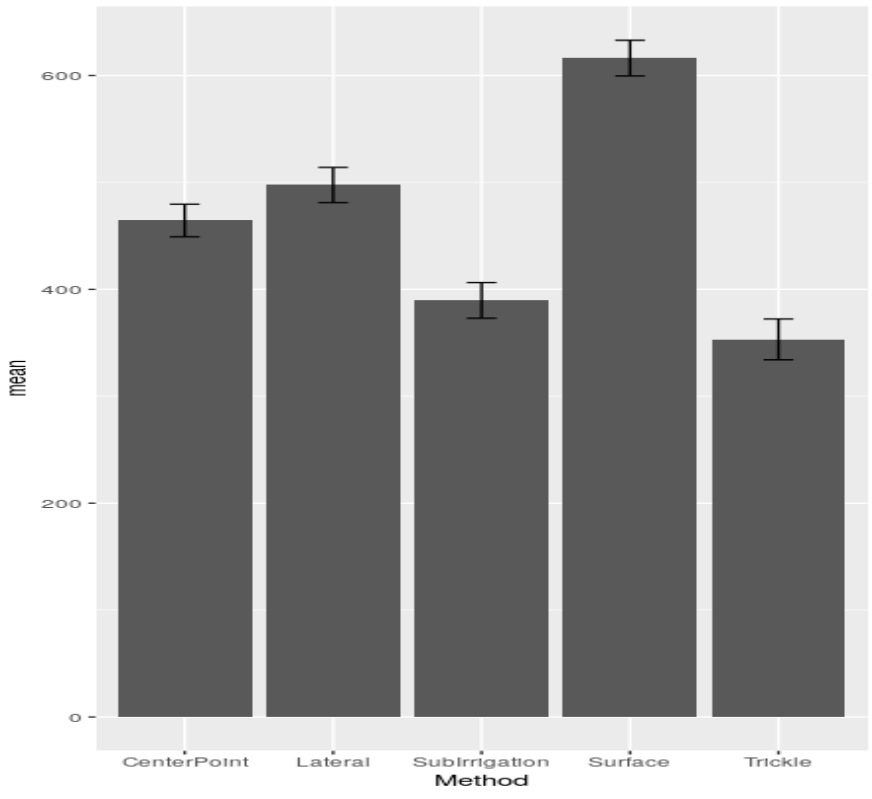
A. Calculate the sample size, simple mean and SE for each method (averaging over farms). Include the resulting summary table in your assignment.
ANSWER:

```
# 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
```

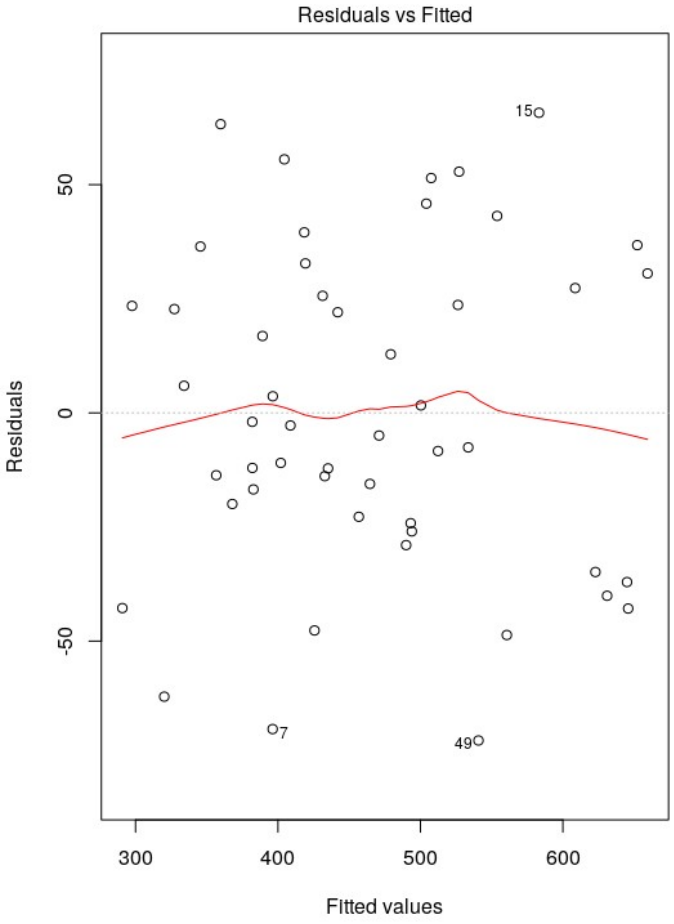
B. Create a bar chart (with SE bars) to summarize the data. Include the resulting graph in your assignment. Assuming your summary table from above is called SumStats (with columns Method, mean and se), you can use code something like this:

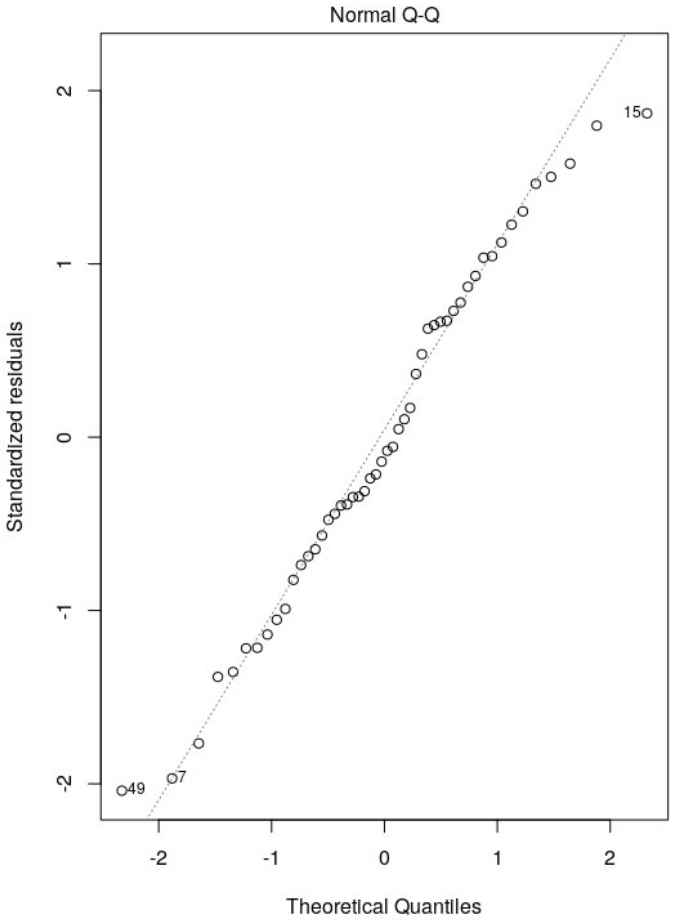
```
Library(ggplot2)
ggplot(SumStats, aes(x = Method, y = mean)) +
geom_bar(stat = "identity") +
geom_errorbar(aes(ymin =
mean-se, ymax = mean+se),
width = 0.2)
```

ANSWER:



C. Fit the RCB model. Inspect the diagnostic plots (Resids vs Fitted and Normal QQplot of Resids), and comment on what you see. Do the assumptions appear to be satisfied? Note: You do not have to include the diagnostic plot in your assignment, just comment on each graph. (4 pts)
ANSWER:





From the above plots, we can conclude that the assumptions are indeed satisfied. The Resid vs Fitted has around equal scatter, and the Q-Q plot is a clear straight line.

D. Continuing with the RCB model from the previous question, include the Type3 ANOVA table in your assignment.
ANSWER:

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
```

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. Can we conclude that there is a difference between the irrigation methods? Justify your response with a test statistic and p-value.
ANSWER:
Yes, we have a test statistic of 61.2751 and a very significant p-value.

F. Make a conclusion about the effectiveness of the blocking in this example. Justify your response with a test statistic and p-value.
ANSWER:
Yes, the blocking was effective, since we have a p-value of 0.0007685 (significant) and test statistic 4.2874.

G. The investigators are interested in which irrigation methods are significantly different from each other. Use emmeans() function from the emmeans package to get Tukey-adjusted p-values for comparing treatments. Then use this information to create a “cld” display, where methods that are NOT significantly different from each other are given the same number grouping. Hint: Use code something like this:
library(emmeans)
library(multcompView)
emout <- emmeans(Model1, pairwise ~ Method)
emout
cld(emout)
ANSWER:

$emmeans

Method	emmean	SE	df	lower.CL	upper.CL
CenterPoint	464	13.1	36	438	491
Lateral	498	13.1	36	471	524
SubIrrigation	390	13.1	36	363	416
Surface	616	13.1	36	590	643
Trickle	353	13.1	36	327	380

```
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	18.5	36	-1.791	0.3944
CenterPoint - SubIrrigation	74.7	18.5	36	4.029	0.0024
CenterPoint - Surface	-151.9	18.5	36	-8.193	<.0001
CenterPoint - Trickle	111.2	18.5	36	5.998	<.0001
Lateral - SubIrrigation	107.9	18.5	36	5.820	<.0001
Lateral - Surface	-118.7	18.5	36	-6.403	<.0001
Lateral - Trickle	144.4	18.5	36	7.789	<.0001

```
 SubIrrigation - Surface       -226.6 18.5 36 -12.223 <.0001
 SubIrrigation - Trickle         36.5 18.5 36   1.969 0.3014
 Surface - Trickle              263.1 18.5 36  14.191 <.0001

Results are averaged over the levels of: Farm
P value adjustment: tukey method for comparing a family of 5 estimates
```

```
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. Are the simple means (part A) and emmeans (part G) the same for this analysis? What about the simple SE's (part A) versus SE's returned by emmeans (part G)?
ANSWER:
Yes they are the same. The simple SEs and the SEs returned by emmeans are not the same.

I. Run the analysis as a one-way ANOVA using just Method in the model. (In practice I would not do this, but try it here for illustration.) Include the ANOVA table in your assignment. How does dfResid compare to the RCB model? How does MSResid compare to the RCB model? (4 pts) Hint: Recall that MSResid = SSResid/dfResid.
ANSWER:

```
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
```

Both DFResid and MSResid are higher compared to the RCB model.

2.
A fertilizer trial on a range grass (blue grama) was conducted in a randomized complete block design. Five fertilizer treatments were randomly assigned to the plots in each of five blocks, but two observations have missing values. The response variable (Y) represents phosphorous. The data is available from Canvas as "GrassMiss.csv". Note: Be sure to define Block as.factor!

A. Calculate the simple mean for each trt (averaging over blocks). Include the resulting summary table in your assignment. Hint: Because of the NA values, it is easiest to use

aggregate() here.
ANSWER:

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

B. Fit the RCB model and include the Type3 ANOVA table in your assignment.
ANSWER:

```
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. Calculate the emmeans and corresponding confidence intervals for each trt and include them in your assignment. Note that the SE is larger (and CIs are wider) for trts that have missing values.
ANSWER:

```
 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. Are the simple means (part A) and emmeans (part C) the same for this analysis?
ANSWER:
No the simple and emmeans are not the same.

E. Use the coefficient estimates (from the summary() output) to compute predicted values for the two missing observations. Show your work for full credit. (Note that you can verify these using the predict() function.) (4 pts)
ANSWER:

Block 3, N50wP:
2.02059 + 0.394 + 0.01365 = **2.428**

Block 5, Ctrl:
2.02059 + 0 + 0.06765 = **2.0882**

F. Verify that the lsmean for N50wP is the average of the five predicted values (one from each block) for N50wP. Show your work for full credit.
ANSWER:

For N50wP:
(2.414588 + 2.410588 + 2.428235 + 2.502588 + 2.482235)/5 = **2.4476**, which is the same value as the computed lsmean.