Simonson_HW7

Martin Simonson

March 8, 2019

1.) Fertilized Lettuce

The data in *lettuce.csv* come from a randomized experiment on the response of lettuce to additions of nitrogen fertilizer. The experiment was conducted on irrigated plots in Arizona. Plots were randomly assigned to one of four fertilizer levels: 0, 50, 100 or 150 lb. N /acre. The response is the number of lettuce heads harvested from each plot. There were four replicates of each treatment. Don't worry about assumptions and analyze the yield data without using any transformation.

a)

Test the null hypothesis of no difference among treatments. Report the test statistic and p-value.

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value  Pr(>F)
## N          3 6461.2 2153.73   4.206 0.02998 *
## Residuals 12 6144.8  512.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Answer:** The F-statistic is 4.206, with a p-value of 0.02998 that has 3 and 12 degrees of freedom. Therefore, we have weak evidence to reject the null hypothesis that there is no difference in yield among treatments.

b)

In part (a) above, you have probably rejected the null hypothesis that the means for all four treatments are equal. However, the only meaningful way to decide which of the differences are significant is to either construct a confidence interval for each difference in means, or to conduct tests of hypotheses for every pair of group means. Using the Tukey-Kramer correction with a familywise confidence level of 95%, determine which treatment means are significantly different. Provide the adjusted p-values and a connected letter diagram.

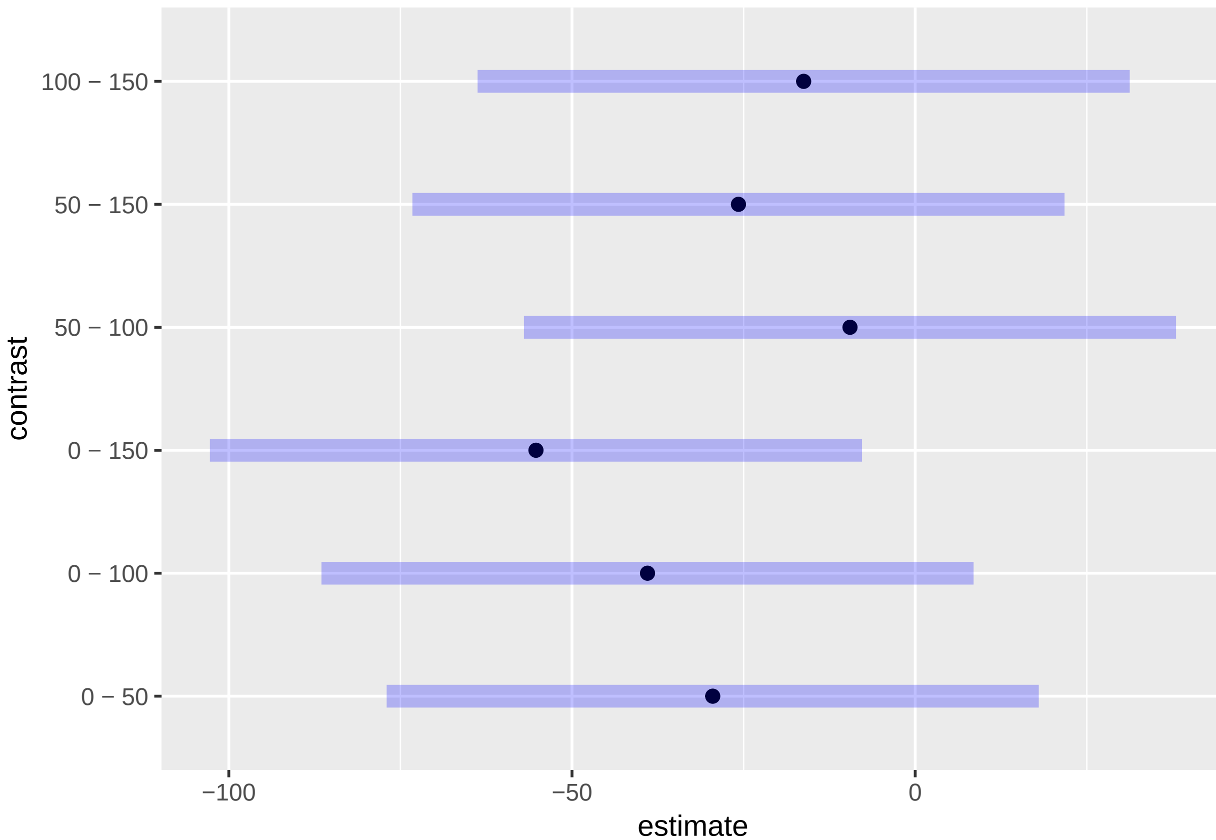
```
## Loading required package: emmeans
```

```
##  N   emmean   SE df lower.CL upper.CL
##  0      108 11.3 12     82.8      132
##  50     137 11.3 12    112.3      162
##  100    146 11.3 12    121.8      171
##  150    163 11.3 12    138.1      187
##
## Confidence level used: 0.95
```

```
##  contrast  estimate SE df t.ratio p.value
##  0 - 50       -29.5 16 12 -1.844  0.3013
##  0 - 100      -39.0 16 12 -2.437  0.1224
##  0 - 150      -55.2 16 12 -3.453  0.0215
```

```
##  50 - 100      -9.5 16 12 -0.594  0.9321
##  50 - 150     -25.8 16 12 -1.609  0.4098
##  100 - 150    -16.2 16 12 -1.016  0.7438
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

```
##  N    emmean   SE df lower.CL upper.CL .group
##  0       108 11.3 12     82.8      132  A
##  50      137 11.3 12    112.3      162  AB
##  100     146 11.3 12    121.8      171  AB
##  150     163 11.3 12    138.1      187   B
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```



- **Answer:** The only treatment means with a difference were 0 and 150 pounds of nitrogen per acre. The adjusted p-value is 0.0215, indicating some evidence to reject the null hypothesis that the mean number of lettuce heads harvested between 0 additional nitrogen and 150 lbs of nitrogen per acre.

c)

Based only on your work above (i.e. no additional work required), do you believe these data provide evidence that adding more N fertilizer increases lettuce yield? Briefly explain why or why not?

- **Answer:** I think the scale of yield is important here. Based on the pairwise 95% confidence interval for the difference in mean response (lettuce heads), there is some evidence for a difference between 0 lbs. of nitrogen added per acre and 150 lbs. of nitrogen per acre. The cost of additional nitrogen fertilizer may outweigh the increased yield of lettuce.

d)

Construct a linear trend contrast to test the null hypothesis of no linear trend in yield due to additional N fertilizer. Report the test statistic, p-value and conclusion for this test.

```
contrast(lettuce.emm, list("Linear Trend Contrast" = c(-75,-25,25,75)))
```

```
##  contrast              estimate   SE df t.ratio p.value
##  Linear Trend Contrast     4381 1265 12 3.463   0.0047
```

2.) Rat Diets

For an evaluation of diets used for routine maintenance of laboratory rats, researchers used a completely randomized design (same number of units randomly assigned to each treatment) to allocate weanling male rats to five different diets (20 rats total). After four weeks, specimens of blood were collected and various biochemical variables were measured. We consider the results for blood urea concentration (mg/dl). The data is given in *RatDiet.csv*.

Suppose diet A is a control and the researchers were only interested in the performance of Diet A compared to the other diets. Construct a family of confidence intervals for the pairwise differences between Diet A and the other four diets (i.e. there are only 4 confidence intervals to construct). Use a method for determining significance that results in a familywise confidence level of at least 95%, for just this set of pairwise differences. Based on these intervals, which diets produce significantly different mean blood urea concentrations from Diet A?

```
cc<-contrast(rats.emm, list("A-B" = c(1,-1,0,0,0),
                            "A-c" = c(1,0,-1,0,0),
                            "A-D" = c(1,0,0,-1,0),
                            "A-E" = c(1,0,0,0,-1)),
         adjust = "bonferroni")

cc
```

```
##  contrast estimate   SE df t.ratio p.value
##  A-B         -0.75 2.69 15 -0.279  1.0000
##  A-c          7.25 2.69 15  2.694  0.0666
##  A-D         10.25 2.69 15  3.809  0.0068
##  A-E         -8.50 2.69 15 -3.159  0.0260
##
## P value adjustment: bonferroni method for 4 tests
```

```
confint(cc,level=0.95)
```

```
##  contrast estimate   SE df lower.CL upper.CL
##  A-B         -0.75 2.69 15   -8.383    6.883
##  A-c          7.25 2.69 15   -0.383   14.883
##  A-D         10.25 2.69 15    2.617   17.883
##  A-E         -8.50 2.69 15  -16.133   -0.867
##
## Confidence level used: 0.95
## Conf-level adjustment: bonferroni method for 4 estimates
```

- **Answer:** Diet D is the only treatment with evidence of a different in urea concentration wehn compared to diet A. The 95% bonferroni adjusted confidence interval for the difference in mean urea concentration between the treatments does not contain zero.

3.) Rat Diet, Revisited

Amphetamine is a drug that suppresses appetite. In a study of this effect, a pharmacologist randomly allocated 24 rats to three treatment groups to receive an injection of amphetamine at one of two dosage levels (2.5 mg/kg and 5 mg/kg), or an injection of saline solution. She measured the amount of food consumed by each animal in the 3-hour period following the injection. The results (gm of food consumed per kg body weight) are available in *amphetamine.csv*.

a)

Is there evidence that the amount of food consumed by animals is different among the three groups? Provide a test statistic, approximate p-value, and a conclusion, in context.

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Consumed
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment  2 8134.2  4067.1  30.082 6.839e-07 ***
## Residuals 21 2839.2   135.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Answer:** The F-statistic is is 30.082 and a p-value less than 0.0001 with 2 and 21 degrees of freedom. There is strong evidence that the mean amount of food consumed by animals is not equal between treatments.

b)

Test the claim that mean food consumption for the group receiving the saline solution is larger than the average of the mean food consumption for the two groups receiving amphetamine dosages. [Use a two-sided alternative hypothesis even if the claim is one-sided]

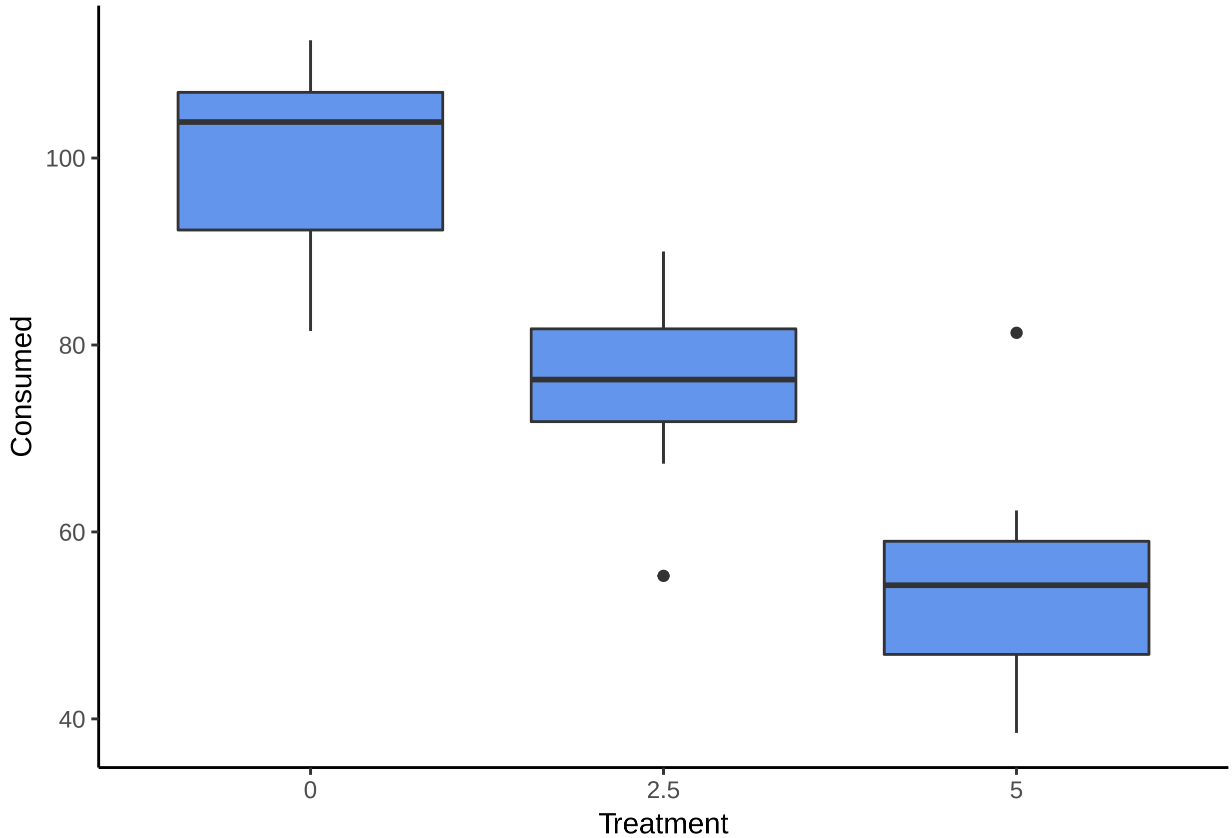
```
rats2.cc<-contrast(rats2.emm,list("Saline-drug"=c(1,-.5,-.5)))
rats2.cc
```

```
##  contrast    estimate   SE df t.ratio p.value
##  Saline-drug     34.8 5.03 21 6.904   <.0001
```

c)

Sketch a plot of the estimated mean food consumption for each group versus amphetamine dosage levels (0, 2.5, 5). Describe in a few words the relationship between the means (if any) you observe in this plot.

```
## Loading required package: ggplot2
```



- **Answer:** The plot of the means show that the mean consumption rate decreases as amphetamine dosage increases.

d)

Use a linear trend contrast to test the null hypothesis of no linear trend in food consumption. Report the test statistic and p-value and a conclusion within the context of the data.

```
rats2.cc2<-contrast(rats2.emm,list("Linear Trend Contrast"=c(23.18,-1.31,-21.86)))
rats2.cc2
```

```
##  contrast              estimate  SE df t.ratio p.value
##  Linear Trend Contrast     1018 131 21 7.762   <.0001
```

- **Answer:** The t-ratio is 7.762 with 21 degrees of freedom. The p-value less than 0.0001 suggests strong evidence against the null hypothesis and data suggest there is a linear decrease in food consumption when amphetamine dosage increases.

4) Soybean Yields

An experiment was conducted to compare soybean yields of plots treated with one of three herbicides (A, B, or C) or a control (D). A total of 32 soybean plots were used in the experiment; 8 plots were randomly assigned to each of the four treatments. Statistics summarizing the yields (in bushels per acre) are provided in the following table.

Treatment	Number of Plots	Mean Plot Yield	Variance of Yield (SD^2)
A	8	60.3	9.8
B	8	62.1	10.1
C	8	57.6	9.2
D	8	52.0	9.7

Test the following Contrasts

a) Herbicide A on Soybead Yield (A-D)

Let μ_1 and μ_4 denote the mean yield of soybeans treated with herbicide A and control D, respectively. Therefore,

- $H_0 : \mu_1 = \mu_4$ vs. $H_A : \mu_1 \neq \mu_4$

Mean Squared Error (MSE), Error, T-statistic, p-value, and 95% Confidence Interval:

- $MSE = \frac{(n_1-1)S_1+(n_2-1)S_2+(n_3-1)S_3+(n_4-1)S_4}{(n_1-1)+(n_2-1)+(n_3-1)+(n_4-1)}$
- $S_p = \sqrt{MSE}$

```
MSE<- ((7*9.8)+(7*10.1)+(7*9.2)+(7*9.7)) /
      (7*4)
MSE # 9.7 bushels/acre
```

```
## [1] 9.7
```

```
Sp<-sqrt(MSE)
Sp # 3.114482 bushels / acre
```

```
## [1] 3.114482
```

- $\gamma_1 = \mu_1 + 0 * \mu_2 + 0 * \mu_3 - \mu_4$

```
gamma.1<-60.3-52
```

- $SE = S_p * \sqrt{\frac{coeff_1^2}{n} + \frac{coeff_2^2}{n} + \frac{coeff_3^2}{n} + \frac{coeff_4^2}{n}}$

```
SE<-Sp * sqrt(2/8)
SE
```

```
## [1] 1.557241
```

- $T = \frac{\gamma_1}{SE}$

```
T.stat<-gamma.1/SE
T.stat
```

```
## [1] 5.329939
```

```
p.value <- 1-pt(T.stat,28)
p.value
```

```
## [1] 5.629969e-06
```

- 95% CI $= \gamma_1 \pm t^* * SE$

```
gamma.1-(2.048*SE)
```

```
## [1] 5.11077
```

```
gamma.1+(2.048*SE)
```

```
## [1] 11.48923
```

- **Answer:** With a t-statistic of 5.33 and 28 denominator degrees of freedom, we have an observed p-value less than 0.0001. Therefore, we have strong evidence to reject the null hypothesis that herbicide A and control D have no difference in mean soybean yields. We are 95% confident that the mean yeild of soybeans will increase by 5.11 5o 11.49 bushels per acre with herbicide A when compared to the control D.

b) Herbicide B vs. Herbicide C

Let μ_2 and μ_3 denote the mean yield of soybean plots treated with herbicide B and herbicide C, respectively. Therefore, we want to test:

- $H_0 : \mu_2 = \mu_3$ vs. $H_A : \mu_2 \neq \mu_3$

Gamma, t-test, p-value, and 95% confidence interval:

- $\gamma_2 = 0 * \mu_1 + \mu_2 - \mu_3 + 0 * \mu_4$

```
gamma.2<-(62.1-57.6) #4.5
```

- $T = \frac{\gamma_1}{SE}$

```
T.stat<-gamma.2/SE
T.stat
```

```
## [1] 2.889726
```

```
p.value <- 1-pt(T.stat,28)
p.value
```

```
## [1] 0.003682697
```

- 95% CI $= \gamma_2 \pm t^* * SE$

```
gamma.2-(2.048*SE)
```

```
## [1] 1.31077
```

```
gamma.2+(2.048*SE)
```

```
## [1] 7.68923
```

- **Answer:** We have an observed t-statistic of 2.89 with 28 denominator degrees of freedom and p-value of 0.00368, which suggests strong evidence against the null hypothesis that there is no true difference in mean soybean yields between herbicide B and herbicide C. We are 95% confidenc that the mean soybean ield for herbicide B is 1.31 to 7.69 bushels per acre more than when herbicide C is used.

c) Average effect of herbicides on soybead yield: (A+B+C)/3-D

Let μ_5 denote the mean yield of soybean plots treated with the average combined effects of all three fertilizers A-C, and let μ_4 denote the mean yield of soybean plots without treatment, D. Therefore, we want to test:

- $H_0 : \mu_5 = \mu_4$ vs. $H_A : \mu_5 \neq \mu_4$

Gamma, SE, T-statistic, P-value, and 95% CI:

- $\gamma_3 = (1/3) * \mu_1 + (1/3) * \mu_2 + (1/3) * \mu_3 - \mu_4$

```
gamma.3<-(((1/3)*60.3)+
            ((1/3)*62.1)+
            ((1/3)*57.6)-
            52)
# 8
```

- $SE = S_p * \sqrt{\frac{coeff_1^2}{n} + \frac{coeff_2^2}{n} + \frac{coeff_3^2}{n} + \frac{coeff_4^2}{n}}$

```
coeff<-(1/3)^2
SE<-Sp * sqrt((coeff/8)+(coeff/8)+(coeff/8)+(1/8))
SE
```

```
## [1] 1.271482
```

- $T = \frac{\gamma_1}{SE}$

```
T.stat<-gamma.3/SE
T.stat
```

```
## [1] 6.29187
```

```
p.value <- 1-pt(T.stat,28)
p.value
```

```
## [1] 4.187444e-07
```

- 95% CI $= \gamma_2 \pm t^* * SE$

```
gamma.3-(2.048*SE)
```

```
## [1] 5.396005
```

```
gamma.3+(2.048*SE)
```

```
## [1] 10.604
```

- **Answer:** We observed a t-statistic of 6.29187 with 28 denominator degrees of freedom, and a p-value less than 0.0001, suggesting very strong evidence to reject the null hypothesis and conclude that that the mean soybean yield for all herbicide treatments is not equal to the control treatment. We are 95% confident that the combined use of herbicides A, B, and C will increase from mean yield by between 5.396 and 10.6 bushels per acre compared to no herbicide treatment.

5.) Root Metabolism

A plant physiologist investigated the effect of flooding on root metabolism in two tree species: flood-tolerant river birch and the intolerant European birch. Four seedlings of each species were flooded for one day and four were used as controls. The concentration of adenosine triphosphate (ATP) in the roots of each plant was measured. The data (nmol ATP per mg tissue) are shown in the table below, and are available in *ATP.csv*.

a)

Test the null hypothesis that the ATP concentration for the four groups is the same. Provide a test statistic, p-value, and a conclusion.

- We are testing H_0 : All group means are equal vs. H_A : All group means are not equal.

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: ATP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Group      3 4.5530 1.51766  38.391 1.984e-06 ***
## Residuals 12 0.4744 0.03953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Answer:** With an f-statistic of 38.391 and a p-value less than 0.0001 there is very strong evidence that the mean ATP concentrations are different between groups.

b)

Show the contrasts to measure each of the following and test the contrasts. (remember to define the μ-notations)

Let μ_1 denote the mean nmol ATP per mg of root tissue in European Birch in a reference site and let μ_2 denote the mean nmol ATP per mg of root tissue in European Birch trees in a flood setting. Further, let μ_3 denote the mean nmol ATP per mg of root tissue in River Birch trees in a reference setting and let μ_4 denote the mean nmol ATP per mg root tissue in River Birch trees in a flooded site.

```
atp.emm
```

```
##  Group           emmean     SE df lower.CL upper.CL
##  EuropeanControl  1.200 0.0994 12   0.9834    1.417
##  EuropeanFlooded  0.292 0.0994 12   0.0759    0.509
##  RiverControl     1.785 0.0994 12   1.5684    2.002
##  RiverFlooded     1.190 0.0994 12   0.9734    1.407
##
## Confidence level used: 0.95
```

i) The effect of flooding in European Birch

- $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$

```
contrast(atp.emm, list("European Control - European Flood"=c(-1,1,0,0)))
```

```
##  contrast                          estimate    SE df t.ratio p.value
##  European Control - European Flood   -0.907 0.141 12 -6.455  <.0001
```

- **Answer:** With a t-ratio of -6.455 with 12 degrees of freedom and a p-value less than 0.0001 we have very strong evidence to reject the null hypothesis and conclude there is a difference in mean nmol ATP per mg root tissue between Eropean Birch trees in a control and flood site.

ii) Difference between river birch and European birch with repect to flood effect.

- $H_0 : \mu_2 = \mu_4$ vs. $H_A : \mu_2 \neq \mu_4$

```
contrast(atp.emm, list("European Flood - River Flood" = c(0,1,0,-1)))
```

```
##  contrast                     estimate    SE df t.ratio p.value
##  European Flood - River Flood   -0.897 0.141 12 -6.384  <.0001
```

- **Answer:** With a t-ratio of -6.384 on 12 degrees of freedom and a p-value of less than 0.0001, we observe very strong evidence against the null hypothesis and can conclude that the mean nmol ATP per mg root tissue is not equal between species.

iii) Comparing the average of the flooded and control river birch ATP to the average of the flooded and control european birch mean ATP levels.

```
contrast(atp.emm, list("Mean European Birch - Mean River Birch"=c(0.5,0.5,-0.5,-0.5)))
```

```
##  contrast                               estimate     SE df t.ratio p.value
##  Mean European Birch - Mean River Birch   -0.741 0.0994 12 -7.456  <.0001
```

- **Answer:** With a t-ratio of -7.456 on 12 degrees of freedom and a p-value less than 0.0001 we have very strong evidence to reject the null hypothesis and conclude there is a difference in the mean nmol ATP per mg root tissue bewteen flooded sites and reference sites.

c)

Obtain Bonferroni adjusted confidence intervals for the above contrasts, with a familywise confidence of 95%.

```
##  contrast                               estimate     SE df t.ratio p.value
##  European Control - European Flood        -0.907 0.1406 12 -6.455  0.0001
##  European Flood - River Flood             -0.897 0.1406 12 -6.384  0.0001
##  Mean European Birch - Mean River Birch   -0.741 0.0994 12 -7.456  <.0001
##
## P value adjustment: bonferroni method for 3 tests
```

```
##  contrast                               estimate     SE df lower.CL
##  European Control - European Flood        -0.907 0.1406 12    -1.30
##  European Flood - River Flood             -0.897 0.1406 12    -1.29
##  Mean European Birch - Mean River Birch   -0.741 0.0994 12    -1.02
##  upper.CL
##    -0.517
##    -0.507
##    -0.465
##
## Confidence level used: 0.95
## Conf-level adjustment: bonferroni method for 3 estimates
```

6.) Beetles, Revisited

[Beetles problem from HW6] Test the hypothesis that the means for yellow, blue and green are the same. That is, test the hypothesis that $\mu_{yellow} = \mu_{blue} = \mu_{green}$.

```
cld(beetles.emm, alpha = 0.05, Letters = LETTERS)
```

```
##  Color  emmean   SE df lower.CL upper.CL .group
##  Gray     14.6 1.99 35     10.5     18.6  A
##  Blue     15.3 2.10 35     11.0     19.5  A
##  Green    33.8 1.99 35     29.8     37.9   B
##  Yellow   47.2 1.99 35     43.1     51.2    C
##
## Confidence level used: 0.95
```

```
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

```
beetles.cc<-contrast(beetles.emm, list(c(1,0,-1,0),
                                       c(1,0,0,-1),
                                       c(0,0,1,-1)))
test(beetles.cc,joint=T)
```

```
##  df1 df2 F.ratio p.value note
##    2  35  61.065 <.0001   d
##
## d: df1 reduced due to linear dependence
```

- **Answer:** The F-statistic was 61.065 on 2 and 35 degrees of freedom, with a p-value less than 0.0001. Therefore, we have strong evidence to reject the null hypothesis and conclude there is a difference in mean number of beetles between green, blue, and yellow board colors. In addition, the connected letter diagram indicates different letters for each of those colors, which is more evidence for a difference in a familywise comparison.