BTRY 3020/STSCI 3200: Homework VIII

NAME: Kevin Klaben

NETID: kek228

DUE DATE: 8:40 am Tuesday April 30, 2019

Question 1.

An experiment was conducted to determine the effects of four medicines in combination with levels sodium intake on blood pressure. The medications included three new medications that were assumed to reduce blood pressure using a new, unique biological approach, while the fourth was the most commonly used medication at that time. The first three medications were the new ones, and are labelled M1, M2, and M3, while the last, currently used medication was labelled M4. Sodium levels were S1, where no additional sodium (salt) was added to the patient's food, while the other was S2, where patients used sodium (salt) normally.

Each combination of medication and sodium level was randomly assigned ten patients, and each patient followed their medication-sodium regiment for a month, to allow time for the medications to take effect. Blood pressure was then measured on each patient once a day for the next ten days. This resulted in a total of 800 records of blood pressure.

A) What are the experimental factors in this study? What is the experimental design?

The experimental factors sodium intake and medicine. The design for this experiment is a completely randomized design in which patients were randomly evenly split amoungst the 8 treatment groups.

B) What are the experimental treatments in this study? Explain in one sentence. How many are there?

There a total of 8 treatment groups, as there are four levels of medication and 2 levels of sodium intake. Patients are randomly assigned a medicine to take and a sodium intake amoutn resulting in 8 total treatment groups.

C) What are the experimental units in this study? Explain in one sentence.

The experimental units are the people with high blood pressure being assigned to treatment groups.

D) How many replicates are there for each treatment? Explain in one sentence.

There are 10 replicates for each treatment, as there are ten patients assigned to each treatment.

E) What are the sampling units in this study? Explain in one sentence.

The sampling units are blood pressure measures taken of each person in each treatment group 10 times resulting in 800 total samples.

F) Are pseudoreplicates present in this study? Explain in one sentence. How would these be used in the analysis of treatment effects?

Yes, pseudoreplication is present, the fact that we are taking 10 data points from each patient is pseudoreplication as it yield an artificially high number of observations whihc will not indicate the same strength of confidence in our conclusion as would 800 different patients each having their blood pressure taken once. In the analysis, we could take this into account by instead of using each sample taken as a separate observation, instead take the average of ten observations per patients and analyze suing these averages.

G) If you wanted to compare each of the new medicines with the currently used one, which technique would you use? What would allow you to make this comparison while ignoring the sodium levels?

This can be done by using dunnett's method, as it can be used to compare all other treatments with a control. We would need to see that different sodium levels had no significant impact on the blood pressures, then we would be able to ignore the sodium levels.

H) Let μ_{ij} be the mean reduction in blood pressure for the ith medicine at the jth sodium level, i = 1, 2, 3, 4; j = 1, 2. Construct a contrast that compares the medicines using the new biological approach to the currently used medicine.

The contrast would be $(\mu_{1,1}+\mu_{2,1}+\mu_{3,1}+\mu_{1,2}+\mu_{2,2}+\mu_{3,2})/6-(\mu_{4,1}+\mu_{4,2})/2$.

Question 2.

An experiment was run to study how long mung bean seeds should be soaked prior to planting in order to promote early growth of bean sprouts. The experiment was run using a completely randomized design. Soaking levels used in this experiment were as follows: A= low, B= medium, C = high, and D = very high. For each treatment level, 17 beans were used and the mean shoot length (Y in mm) was measured 48 hours following soaking. Data appears in the file Hwk8Q2DatSp19.xlsx.

A) Perform analysis of variance to test the hypothesis that the four treatments' means are equal. State carefully your conclusions.

```
library(readxl)
BeanDat <- read_excel("Hwk8Q2DatSp19.xlsx")
BeanDat$Treatment <- as.factor(BeanDat$Treatment)
Bean.aov <- aov(Length ~ Treatment, data = BeanDat)
summary(Bean.aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    3 2501.3   833.8   75.92 <2e-16 ***
## Residuals   64  702.8    11.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The overall F-Test in ANOVA is testing Ho:μ1 =μ2 = …=μ4 vs Ha:NotHo As we saw with the regression, the F-Test has test statistic 75.92 on 3 and 64 df, with an extremely small p-value, we reject the null hypothesis and conclude that at least one mean is not equal to zero.

B) Give a statistical model appropriate for describing the response variable in this study and explain each term in the model.

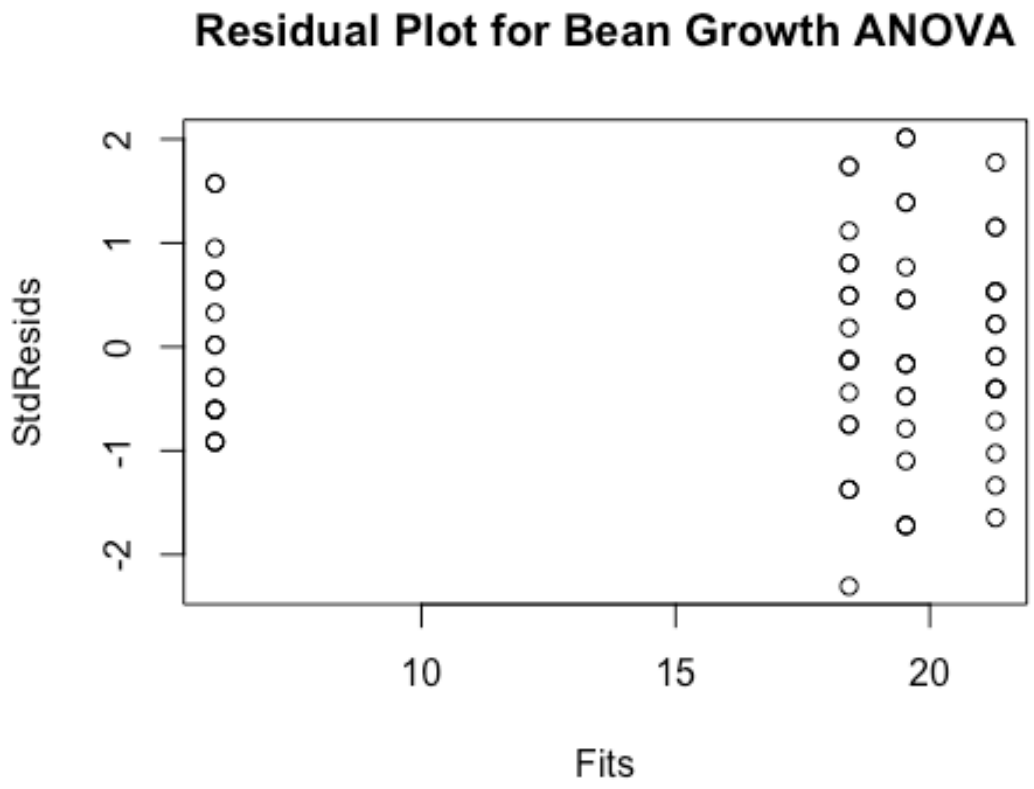
the model would be: $y_{ij}=\mu_i+\epsilon_{ij}$ where y_{ij} is normally distributed $N(\mu_i,\sigma^2)$ μi is the theoretical mean of all observations at factor level i ϵij is the jth error term of observation y_{ij}

C) Assess the validity of assumptions underlying analysis of variance in this study.

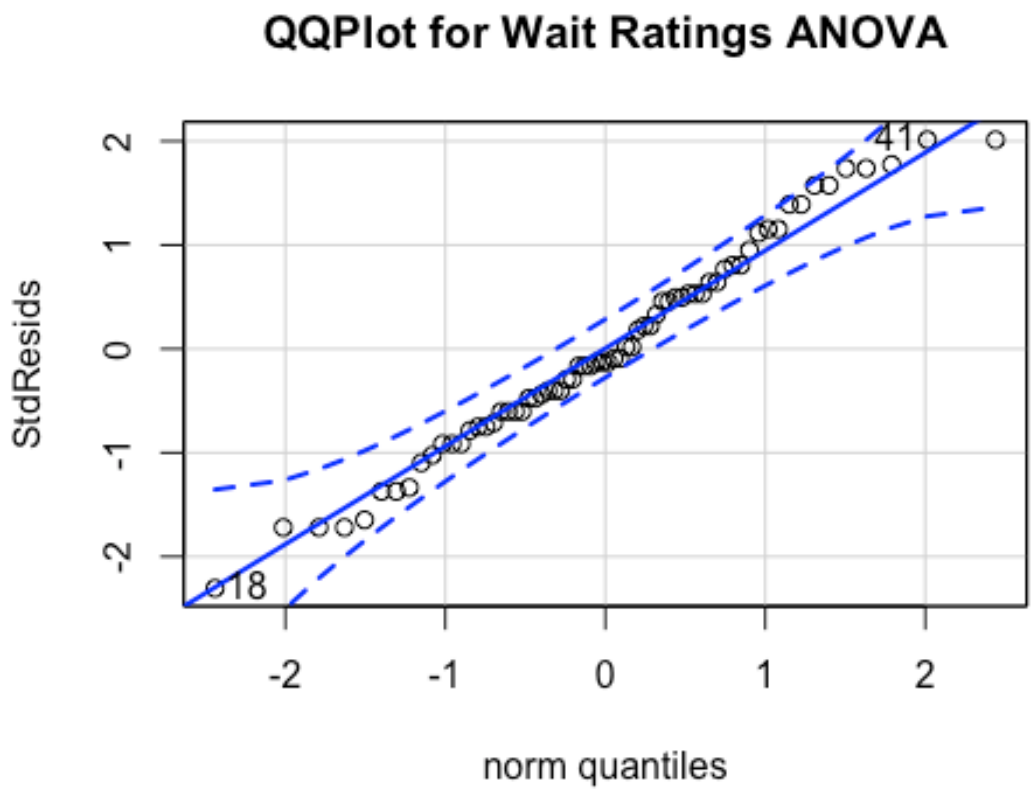
```
Bean.lm <- lm(Length ~ Treatment, data=BeanDat)
library(car)
```

```
## Loading required package: carData
```

```
StdResids<-rstandard(Bean.lm)
Fits<-fitted.values(Bean.lm)
plot(Fits, StdResids, main="Residual Plot for Bean Growth ANOVA")
```




```
qqPlot(StdResids, main="QQPlot for Wait Ratings ANOVA")
```




```
## [1] 18 41
```

Assumptions: 1) Independence: guaranteed by random assignment of treatments to beans (experimental units); 2) Normality: qqPlot looks good enough (linear models' procedures are robust against non-normality); 3) Linearity: Not really an assumption since all predictors are categorical 4) Constant variance: Looks fine in residual plot 5) Outliers not driving conclusions: No outliers apparent in either diagnostic plot.

i) Compare all pairs of means, using Bonferoni's method and make an interpretation of your results. Use $\alpha_{overall}$ = .05.

```
with(BeanDat,pairwise.t.test(x=Length, g=Treatment, p.adjust="bonferroni"))
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  Length and Treatment
##
##   A       B     C
## B 1.4e-15 -     -
## C < 2e-16 1.000 -
## D < 2e-16 0.082 0.753
##
## P value adjustment method: bonferroni
```

We see very low p-values for comparing low to medium, low to high, and low to very high, thus these are the signifcant differences in thier means in our results. ii) What are the advantages and disadvantages of using this method of pairwise comparisons?

Bonferroni controls $\alpha_{overall}$ precisely but it is not as powerful as tukey's for all pairwise comparisons.

i) Compare all pairs of means, using Tukey's method and make an interpretation of your results. Use $\alpha_{overall}$ = .05.

```
TukeyHSD(Bean.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Length ~ Treatment, data = BeanDat)
##
## $Treatment
##          diff        lwr       upr     p adj
## B-A 12.470588  9.4723100 15.468866 0.0000000
## C-A 13.588235 10.5899571 16.586513 0.0000000
## D-A 15.352941 12.3546630 18.351219 0.0000000
## C-B  1.117647 -1.8806311  4.115925 0.7594325
## D-B  2.882353 -0.1159253  5.880631 0.0638738
## D-C  1.764706 -1.2335723  4.762984 0.4128068
```

We see very low p-values for comparing low to medium, low to high, and low to very high, thus these are the signifcant differences in thier means in our results. ii) What are the advantages and disadvantages of using this method of pairwise comparisons?

Most powerful procedure for comparing means pairwise while controlling for $\alpha_{overall}$ absolutely.

i) Compare all pairs of means, using Fisher's Protected LSD and make an interpretation of your results. Use $\alpha_{overall}$ = .05.

```
with(BeanDat, pairwise.t.test(x=Length, g=Treatment, p.adjust="none"))
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  Length and Treatment
##
##   A       B     C
## B 2.4e-16 -     -
## C < 2e-16 0.329 -
## D < 2e-16 0.014 0.125
##
## P value adjustment method: none
```

We see very low p-values for comparing low to medium, low to high, and low to very high as well as medium to very high, thus these are the signifcant differences in thier means in

our results. ii) What are the advantages and disadvantages of using this method of pairwise comparisons? Fisher's is easy to use and more powerful than Tukey's and scheffe's methods while it does have the draw back of only approximately controlling $\alpha_{overall}$ and it can't be used when the overall F-test fails to reject.

G) Consider the first two levels (low, medium) as "short" soaking periods and the two higher levels (High, very high) as "long" soaking periods. You want to determine the difference in mean sprout length between the short and long soaking periods.

```
aggregate(Length ~ Treatment, data=BeanDat, mean)
```

```
##   Treatment    Length
## 1         A  5.941176
## 2         B 18.411765
## 3         C 19.529412
## 4         D 21.294118
```

```
i) Give a 90% confidence interval for the difference in mean sprout length
between short and long soaking periods.
```

L=$(\mu A+\mu B)/2$-$(\mu C+\mu D)/2$

```
(5.94+18.41)/2-(19.53+21.29)/2
```

```
## [1] -8.235
```

```
summary(Bean.aov)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    3 2501.3   833.8   75.92 <2e-16 ***
## Residuals   64  702.8    11.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sqrt(11.0*(1/4+1/4+1/4+1/4)/17)
```

```
## [1] 0.8043997
```

```
-8.235+1.669*(0.804)
```

```
## [1] -6.893124
```

```
-8.235-1.669*(0.804)
```

```
## [1] -9.576876
```

L(hat)=(5.94+18.41)/2-(19.53+21.29)/2=-8.235
SE(L(hat))=Sqrt(11.0*(1/4+1/4+1/4+1/4)/17)= 0.804* CI=-8.235+/-(0.804)$t_{64,0.05}$=-8.235+/-1.669x(0.804)=(-6.89,-9.58) We are 90% confident that the average mung bean sprout is 6.89 to 9.58 millimeters shorter when soaking for shorter periods than for longer periods.
ii) Test to see if the long soaking periods produce higher mean sprout length than the short periods, using α = .05. State hypotheses, test statistic, p-value, and conclusions. Ho:

$(\mu A+\mu B)/2-(\mu C+\mu D)/2=0$ Ha: $(\mu A+\mu B)/2-(\mu C+\mu D)/2<0$ L=$(\mu A+\mu B)/2-(\mu C+\mu D)/2$ TS=-8.235-0/0.804=-10.24 p=P(t_{64}<-10.24)=2.054723e-15

```
(-8.235-0)/0.804
```

```
## [1] -10.24254
```

```
pt(-10.24,64)
```

```
## [1] 2.054723e-15
```

As p=2.054723e-15<alpha=0.05 we reject the null hypothesis and conclude that long soaking periods produce higher mean lengths than shorter soaking periods. H) Using the values corresponding to the levels of the treatments: A = 12 hours, B= 18 hours, C = 24 hours, and D = 30 hours,

```
i) Fit a polynomial regression in hours to this data; report what you get and
how you got there (show all steps and tests).
```

```
library(readxl)
BeanDat2 <- read_excel("Hwk8Q2part2.xlsx")
Bean3.lm=lm(Length~Treatment2+I(Treatment2^2)+I(Treatment2^3)+I(Treatment2^4)
, data=BeanDat2)
summary(Bean3.lm)
```

```
##
## Call:
## lm(formula = Length ~ Treatment2 + I(Treatment2^2) + I(Treatment2^3) +
##     I(Treatment2^4), data = BeanDat2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4118 -2.0294 -0.4118  2.0588  6.4706
##
## Coefficients: (1 not defined because of singularities)
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.011e+02  2.188e+01  -4.619 1.91e-05 ***
## Treatment2       1.548e+01  3.497e+00   4.426 3.82e-05 ***
## I(Treatment2^2) -6.577e-01  1.751e-01  -3.756 0.000375 ***
## I(Treatment2^3)  9.259e-03  2.773e-03   3.339 0.001407 **
## I(Treatment2^4)         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.314 on 64 degrees of freedom
## Multiple R-squared:  0.7806, Adjusted R-squared:  0.7704
## F-statistic: 75.92 on 3 and 64 DF,  p-value: < 2.2e-16
```

```
Bean2.lm=lm(Length~Treatment2+I(Treatment2^2)+I(Treatment2^3), data=BeanDat2)
summary(Bean2.lm)
```

```
##
## Call:
## lm(formula = Length ~ Treatment2 + I(Treatment2^2) + I(Treatment2^3),
##     data = BeanDat2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4118 -2.0294 -0.4118  2.0588  6.4706
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.011e+02  2.188e+01  -4.619 1.91e-05 ***
## Treatment2       1.548e+01  3.497e+00   4.426 3.82e-05 ***
## I(Treatment2^2) -6.577e-01  1.751e-01  -3.756 0.000375 ***
## I(Treatment2^3)  9.259e-03  2.773e-03   3.339 0.001407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.314 on 64 degrees of freedom
## Multiple R-squared:  0.7806, Adjusted R-squared:  0.7704
## F-statistic: 75.92 on 3 and 64 DF,  p-value: < 2.2e-16
```

Our final polynomial has up to cubic terms as this is the first in which all terms are significant. Y=-101.1+15.48(x)-0.6577(x^2)+0.009259(x^3)

```
ii) Compare the MSE you got from using the treatments as categorical
predictors and the polynoimial predictors; assess how much explained
variation you lost by forcing the means to follow the regression "line"
compared to letting them "float".
```

```
anova(Bean.aov)
```

```
## Analysis of Variance Table
##
## Response: Length
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Treatment  3 2501.29  833.76  75.924 < 2.2e-16 ***
## Residuals 64  702.82   10.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Bean2.lm)
```

```
## Analysis of Variance Table
##
## Response: Length
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## Treatment2       1 1891.78 1891.78 172.268 < 2.2e-16 ***
## I(Treatment2^2)  1  487.12  487.12  44.358 7.317e-09 ***
## I(Treatment2^3)  1  122.40  122.40  11.146  0.001407 **
## Residuals       64  702.82   10.98
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The MSE's are the same between the two models as the residual sum of squares are equal between the two models. Thus, as the residual sum of squares are equal, we conclude that no explained variation was lost as a result of forcing the regression "line".

```
iii) What mean sprout length could you expect for 15 hours of soaking (use a
95% interval). Could you have gotten this by using the treatments as
categorical predictors?
```

pt est+/- $t_{0.025,64}$ (SE) =

```
dat=data.frame(Treatment2=15, Treatment2Sq=15^2, Treatment2Cub=15^3)
predict(Bean2.lm, dat, se.fit=TRUE, interval="confidence")

## $fit
##        fit      lwr      upr
## 1 14.34559 12.67842 16.01276
##
## $se.fit
## [1] 0.8345325
##
## $df
## [1] 64
##
## $residual.scale
## [1] 3.313852
```

We can be 95% confident that the expected bean sprount length for 15 hours of soaking will be on average between 12.6784 and 16.0128 millimeters.

No, you can't know where 15 hour would have fit in among catagorical predictors and thus we would have no way of making this calcuation.

Question 3.

For newly planted strawberries, the development of flower clusters decreases the plant vigor. It is common practice to remove the flower stalks by hand, but this is a laborious and time-consuming procedure. To investigate the effect of flower clusters on the plant vigor, an experiment consisting of four treatments was conducted. This experiment was completely randomized and consisted of the following treatments: A = Control (no flower removal), B = Hand removal, C = Regulator G1, and D = Regulator G2 (note that G1 and G2 are hormone-based regulators). A plot of 10 plants was treated and the average number of runners per mother plant, a measure of vigor, was recorded on each plot.

The layout of the experiment and the measures of vigor are provided below for each plot.

C. 3.6 (plot 1) A. 1.4 (plot 6) A. 0.8 (plot 11) B. 5.2 (plot 16)

C. 2.4 (plot 2) D. 7.3 (plot 7) B. 6.8 (plot 12) C. 1.8 (plot 17)

A. 0.6 (plot 3) C. 4.6 (plot 8) B. 3.0 (plot 13) D.6.2 (plot 18)

D. 3.8 (plot 4) D. 4.1 (plot 9) A. 1.2 (plot 14) B. 5.0 (plot 19)

B. 6.0 (plot 5) B. 4.0 (plot 10) A. 0.5 (plot 15) A. 1.5 (plot 20)

Note: This data set is not provided, so you need to create it.

A) Construct a set of 3 contrasts that are suggested by the treatment structure in this experiment to be orthogonal.

removal vs no removal:

hand removal vs regulator removal:

g1 regulator vs G2 regulator:

```
library(readxl)
PlantDat <- read_excel("plantvigordat.xlsx")
PlantDat$treatment <- factor(PlantDat$treatment)
contrvnr=c(1,-1/3,-1/3, -1/3)
conthvreg=c(0,1,-.5, -.5)
contg1vg2=c(0,0,1, -1)
cons=cbind(contrvnr,conthvreg, contg1vg2)

rownames(cons)<= c("A","B", "C","D")

## logical(0)
```

The sum of the contrasts sums of squares is 62.585 and the treatment sums of squares is 65.33. This is

B) Verbally define each of the three contrasts above.

Our first contrast is removal vs no removal. Thus we will contrast A with B,C, and D.

Our second contrast is hand removal vs regulator removal. Thus we will contrast B with C and D.

Our third contrast is G1 regulator vs G2 regulator. Thus, we will contrast C with D.

C) Using the contrasts in a, assess the statistical significance of each contrast based on p-values from an appropriate test.

```
t(cons) %*% cons

##           contrvnr conthvreg contg1vg2
## contrvnr  1.333333       0.0         0
## conthvreg 0.000000       1.5         0
## contg1vg2 0.000000       0.0         2
```

```
contrasts(PlantDat$treatment) = cons
#Now run your ANOVA
plant.lm<-
lm(vigor~C(treatment,contrvnr,1)+C(treatment,conthvreg,1)+C(treatment,contg1v
g2,1), data=PlantDat)
Anova(plant.lm, type="III")

## Anova Table (Type III tests)
##
## Response: vigor
##                              Sum Sq Df  F value    Pr(>F)
## (Intercept)                 250.563  1 171.6920 5.697e-10 ***
## C(treatment, contrvnr, 1)    50.401  1  34.5361 2.340e-05 ***
## C(treatment, conthvreg, 1)    2.059  1   1.4111   0.25221
## C(treatment, contg1vg2, 1)   10.125  1   6.9379   0.01805 *
## Residuals                    23.350 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aov.cons <- aov(vigor ~ treatment, data = PlantDat)
#Good Luck following this next funky command to get the proper ANOVA table
with the orthogonal contrasts included!
#Note the second list names your contrasts in the order you entered them into
Cons
```

As we can see from the anova table, only the signifcant contrasts are removal vs non-removal with a p-value on an F-test of 2.340e-05, and similarly, the contrast of the G1 regulator vs G2 regulator with a p-value of 0.01805.

D) Demonstrate that the three contrast sums of squares do not add up to the treatment sum of squares (there is more than one way to do this). Are you surprised by your results? Why or why not? Are these contrasts orthogonal? Why or why not?

```
t(cons) %*% cons

##           contrvnr conthvreg contg1vg2
## contrvnr  1.333333       0.0         0
## conthvreg 0.000000       1.5         0
## contg1vg2 0.000000       0.0         2

contrasts(PlantDat$treatment) = cons
#Now run your ANOVA
plant.lm<-
lm(vigor~C(treatment,contrvnr,1)+C(treatment,conthvreg,1)+C(treatment,contg1v
g2,1), data=PlantDat)
Anova(plant.lm, type="III")

## Anova Table (Type III tests)
##
## Response: vigor
##                              Sum Sq Df  F value    Pr(>F)
## (Intercept)                 250.563  1 171.6920 5.697e-10 ***
```

```
## C(treatment, contrvnr, 1)   50.401  1  34.5361 2.340e-05 ***
## C(treatment, conthvreg, 1)   2.059  1   1.4111   0.25221
## C(treatment, contg1vg2, 1)  10.125  1   6.9379   0.01805 *
## Residuals                   23.350 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov.cons <- aov(vigor ~ treatment, data = PlantDat)
#Good Luck following this next funky command to get the proper ANOVA table
with the orthogonal contrasts included!
#Note the second list names your contrasts in the order you entered them into
Cons
summary(aov.cons, split=list(treatment=list("no removal vs removal"=1, "hand
vs regulator"=2,"g1 vs g2"=3)), type="III")
```

```
##                                    Df Sum Sq Mean Sq F value   Pr(>F)
## treatment                           3  65.33   21.78  14.921 6.75e-05 ***
##   treatment: no removal vs removal  1  53.14   53.14  36.415 1.74e-05 ***
##   treatment: hand vs regulator      1   2.06    2.06   1.411    0.252
##   treatment: g1 vs g2               1  10.13   10.13   6.938    0.018 *
## Residuals                          16  23.35    1.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
50.401+2.059+10.125
```

```
## [1] 62.585
```

The sum of squares of the the contrast do not add to the sum of squares of treatment as 50.401+2.059+10.125=62.585 which does not equal the sum of squares of treatment which canbe seen to be 65.33. This is not surprising as our treatment groups are of different size. The fact that the groups are not equal also means that they will not be orthogonal.

E) Remove the observations for plots 5, 10, 15, and 20. Re-compute the treatment and contrast sums of squares. Demonstrate that the three contrast sums of squares add up to the treatment sum of squares. Are you surprised by your results? Why or why not? Construct an ANOVA table which shows that with this balanced design, the sums of squares for treatments partitions into the sums of squares for the three contrasts. Are these contrasts now orthogonal? Why or why not?

```
test <- ChickWeight[-c(578),]

library(readxl)
PlantDat2 <- read_excel("plantvigordat.xlsx")
PlantDat2 <- PlantDat2[-c(20),]
PlantDat2 <- PlantDat2[-c(15),]
PlantDat2 <- PlantDat2[-c(10),]
PlantDat2 <- PlantDat2[-c(5),]
PlantDat2$treatment <- factor(PlantDat2$treatment)
con1=c(1,-1/3,-1/3, -1/3)
con2=c(0,1,-.5, -.5)
```

```r
con3=c(0,0,1, -1)
cons2=cbind(con1,con2, con3)

rownames(cons2)<= c("A","B", "C","D")

## logical(0)

t(cons2) %*% cons2

##          con1 con2 con3
## con1 1.333333  0.0    0
## con2 0.000000  1.5    0
## con3 0.000000  0.0    2

contrasts(PlantDat2$treatment) = cons2
#Now run your ANOVA
plant2.lm<-
lm(vigor~C(treatment,con1,1)+C(treatment,con2,1)+C(treatment,con3,1),
data=PlantDat2)
Anova(plant2.lm, type="III")

## Anova Table (Type III tests)
##
## Response: vigor
##                          Sum Sq Df  F value    Pr(>F)
## (Intercept)           208.803  1 120.1741 1.316e-07 ***
## C(treatment, con1, 1)  36.401  1  20.9501 0.0006356 ***
## C(treatment, con2, 1)   1.602  1   0.9218 0.3559437
## C(treatment, con3, 1)  10.125  1   5.8273 0.0326741 *
## Residuals              20.850 12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aov2.cons <- aov(vigor ~ treatment, data = PlantDat2)
#Good Luck following this next funky command to get the proper ANOVA table
with the orthogonal contrasts included!
#Note the second list names your contrasts in the order you entered them into
Cons
summary(aov2.cons, split=list(treatment=list("no removal vs removal"=1, "hand
vs regulator"=2,"g1 vs g2"=3)), type="III")

##                                    Df Sum Sq Mean Sq F value   Pr(>F)
## treatment                           3  48.13   16.04   9.233 0.001925 **
##   treatment: no removal vs removal  1  36.40   36.40  20.950 0.000636 ***
##   treatment: hand vs regulator      1   1.60    1.60   0.922 0.355944
##   treatment: g1 vs g2               1  10.13   10.13   5.827 0.032674 *
## Residuals                          12  20.85    1.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

50.401+2.059+10.125
```

```
## [1] 62.585
```

Now the Sum Sq for each treatment sum to be equal to the sum Sq of treatment as 36.40+1.60+10.13=48.13. I am not surprised by the results as not each treatment has the same number of observations. These are orthogonal as each treatmetn group has the same number of units thus, they are now orthogonal.