Lab 9 - ANOVA tables

In this lab, we'll work on creating more complex design matrices, interpreting coefficient outputs, and conducting ANOVA tests. The data used in this lab are provided in Quinn and Kough (2002) and were accessed from https://qkstats.com/data-files/.

Let's start with a data set from Hall et al. (2000), who examined the effects nitrogen and phosphorus on macroinvertebrate communities in artificial subtidal habitats. Microinvertebrate species richness was measured after either 2, 4, or 6 months, and habitats either were enriched with N and P or were in a control group where nutrients were not added. Species richness is included as a log transformed response variable in the data.

```
hall <- read.csv("P:/My Documents/BDA_Spring2018/hall.csv")

fit.hall <- lm(RICHNESS ~ TREAT * as.factor(TIME), data = hall)
```

Q1. How many parameters are we estimating?
We are estimating 6 parameters.

Q2. Create the design matrix by writing one row for each parameter included in the model.

```
treat <- factor(unique(hall$TREAT))
time <- factor(unique(hall$TIME))
model.df <- expand.grid(treat,time)
model.matrix(~Var1*Var2, data=model.df) #Var 1 is treatment and Var 2 is time
```

```
##   (Intercept) Var1nutrient Var24 Var26 Var1nutrient:Var24
## 1           1            0     0     0                   0
## 2           1            1     0     0                   0
## 3           1            0     1     0                   0
## 4           1            1     1     0                   1
## 5           1            0     0     1                   0
## 6           1            1     0     1                   0
##   Var1nutrient:Var26
## 1                  0
## 2                  0
## 3                  0
## 4                  0
## 5                  0
## 6                  1
## attr(,"assign")
## [1] 0 1 2 2 3 3
## attr(,"contrasts")
## attr(,"contrasts")$Var1
## [1] "contr.treatment"
##
## attr(,"contrasts")$Var2
## [1] "contr.treatment"
```

Q3. Write the equation for each parameter in terms of the βs.
1- Intercept, β_0: control+time 2 mean
2- β_1: + β_0 nutrient+time 2 mean 3- β_2: + β_0 control+time 4 mean 4- + β_0 + β_1 + β_4 nutrient+time 4 mean
5- β_3: + β_0 control+time 6 mean
6- + β_0 + β_1 + β_5 nutrient+time 6 mean

4- β_4: time effect offset of nutrient at time 4 from β_1
6- β_5: time effect offset between control and nutrient at time 6 from β_1

```
summary(fit.hall)
```

```
##
## Call:
## lm(formula = RICHNESS ~ TREAT * as.factor(TIME), data = hall)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -7.6   -0.8   -0.6    1.2    7.0
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         5.800      1.428   4.063 0.000482 ***
## TREATnutrient                       0.800      2.019   0.396 0.695589
## as.factor(TIME)4                   12.800      2.019   6.340 1.81e-06 ***
## as.factor(TIME)6                   20.200      2.142   9.433 2.27e-09 ***
## TREATnutrient:as.factor(TIME)4     10.200      2.855   3.572 0.001616 **
## TREATnutrient:as.factor(TIME)6      6.200      2.943   2.107 0.046273 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.192 on 23 degrees of freedom
## Multiple R-squared:  0.9348, Adjusted R-squared:  0.9207
## F-statistic: 66.01 on 5 and 23 DF,  p-value: 7.149e-13
```

```
anova(fit.hall)
```

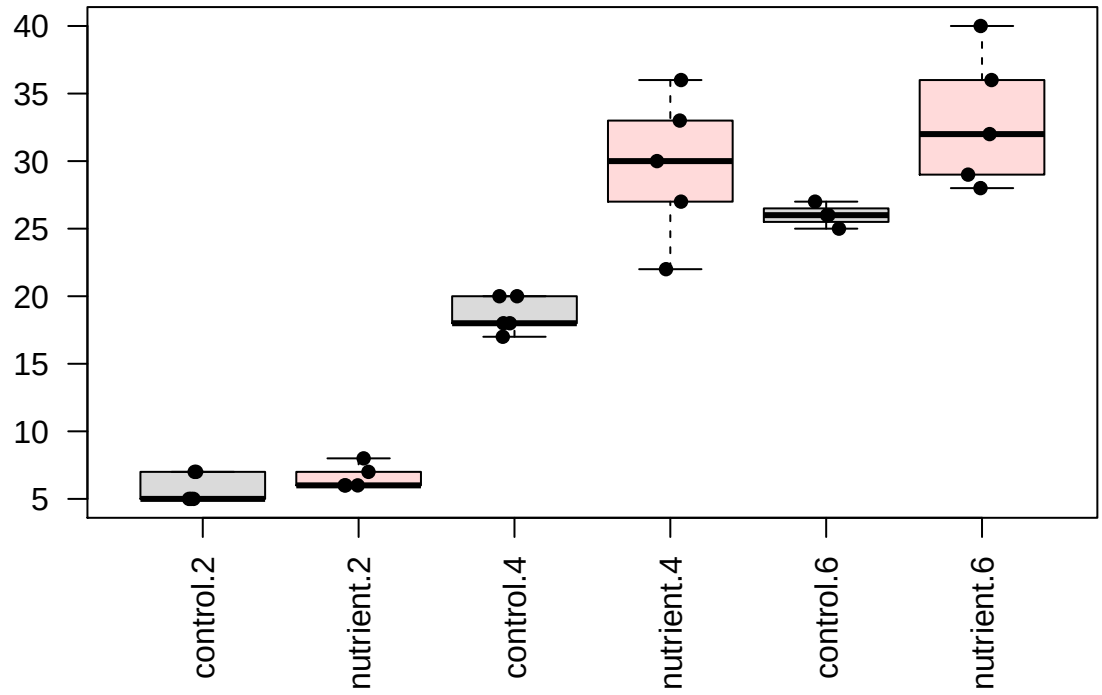
```
## Analysis of Variance Table
##
## Response: RICHNESS
##                        Df  Sum Sq Mean Sq  F value    Pr(>F)
## TREAT                   1  347.15  347.15  34.0629 6.013e-06 ***
## as.factor(TIME)         2 2884.34 1442.17 141.5097 1.185e-13 ***
## TREAT:as.factor(TIME)   2  131.91   65.96   6.4718  0.005892 **
## Residuals              23  234.40   10.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q4. Interpret the p-value. What does it tell us about our model?
The *p-value* for the linear model tells us that given the F distribution on 5 and 23 degrees of freedom, we have a probability of 7.1eˆ-13 of seeing that F ratio (a redution in residuals around the group means versus a grand mean). The p-value gives the chance of an F-statistic of 66.01 or larger if the experiment was repeated infinite times. The ANOVA breaks the p-values into which grouping reduces the residual sum of squares. The interaction factor, treatment at each time step, is signfiant. There is a difference between control and nutrient and that effect changes at each time step. It does not make sense to then interpret the difference between all control groups vs. nutrient groups nor does it make sense to interpret the difference of means at each time step when there is evidence that the interaction term best describes the data. **if they didnt' matter, would be surprising that the p-value is so low, the reduction of residuals and effect size is by chance. or doesn't matter... departure from what we'd expect given the data.** if really no signal from a grouping, then should be surprised that got such a low p-value. p-value is comparing to the null being that the grand mean is best way to explain/have lowest residuals. If main effects are also with interaction effect, then telling about a main effect is sloppy, there's more information when don't just talk about the main effects, there's something else that might be driving why different among groups (the

interaction effect)

Q5. You have a supervisor that wants to know about treatment effects. What do you tell them? What is the effect of increased nutrients on microinvertebrate communities?

```
boxplot(hall$RICHNESS~hall$TREAT+hall$TIME,col=c(rgb(0,0,0,.15),rgb(1,0,0,.15)),las=2, outline=F)
stripchart(RICHNESS~TREAT+TIME, data=hall, method="jitter", pch=16,
           vertical=TRUE, add=TRUE)
```




```
cont4 <- fit.hall$coefficients[1]+fit.hall$coefficients[3]
nut4 <- fit.hall$coefficients[1]+fit.hall$coefficients[3]+fit.hall$coefficients[2]+fit.hall$coefficients
cont6 <- fit.hall$coefficients[1]+fit.hall$coefficients[4]
nut6 <- fit.hall$coefficients[1]+fit.hall$coefficients[4]+fit.hall$coefficients[2]+fit.hall$coefficients

nut4-cont4
```

```
## (Intercept)
##          11
```

```
nut6-cont6
```

```
## (Intercept)
##           7
```

The effect of increased nutrients on microinvertebrate communities is an increasing in richness of 0.8 at time 2, 11 at time 4, and 7 at time 6.

Loyn (1987) was interested in the relationship between habitat characteristics and avian abundance and diversity. In the model below, we estimate abundance as a function of both patch area and stock grazing history, where 1 represents light grazing and 5 represents heavy grazing.

```
birds<-read.csv("P:/My Documents/BDA_Spring2018/loyn.csv")
str(birds)
```

```
## 'data.frame':    56 obs. of  7 variables:
##  $ ABUND  : num  5.3 2 1.5 17.1 13.8 14.1 3.8 2.2 3.3 3 ...
##  $ AREA   : num  0.1 0.5 0.5 1 1 1 1 1 1 1 ...
##  $ YR.ISOL: int  1968 1920 1900 1966 1918 1965 1955 1920 1965 1900 ...
##  $ DIST   : int  39 234 104 66 246 234 467 284 156 311 ...
##  $ LDIST  : int  39 234 311 66 246 285 467 1829 156 571 ...
##  $ GRAZE  : int  2 5 5 3 5 3 5 5 4 5 ...
##  $ ALT    : int  160 60 140 160 140 130 90 60 130 130 ...
```

```
fit.loyn <- lm(ABUND ~ AREA * as.factor(GRAZE), data = birds)
```

Q6. How many parameters are we estimating?
We are estimating 10 parameters

Q7. Create the design matrix by writing one row for each parameter included in the model.

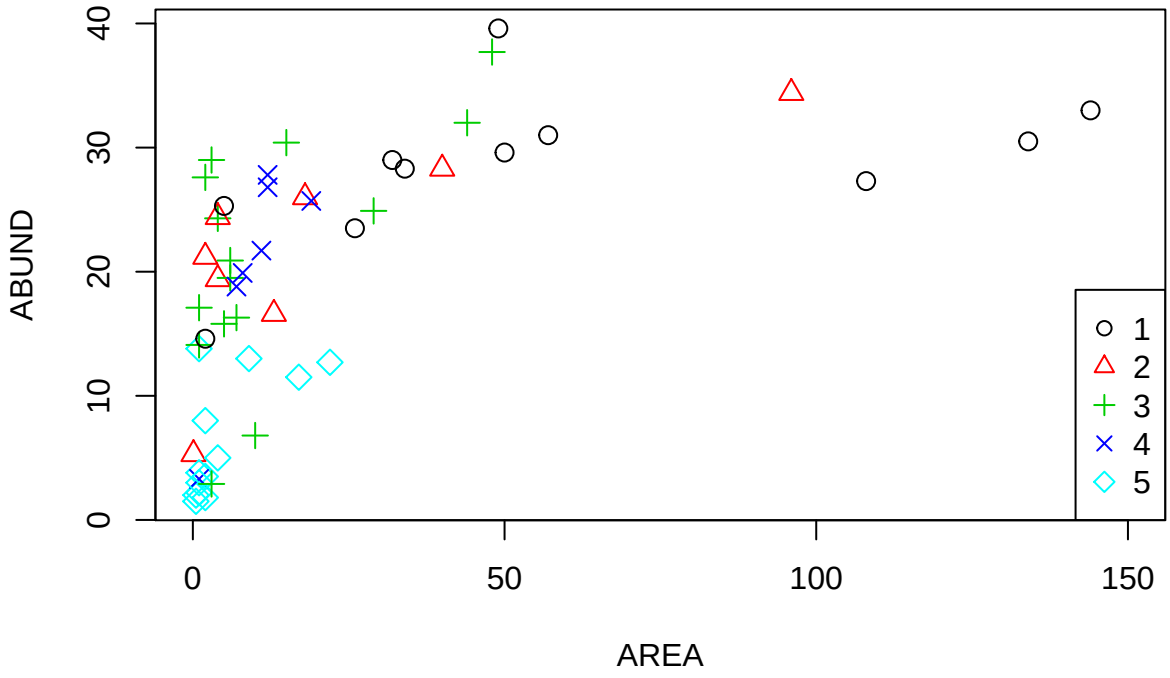
```
modelout <- tapply(birds$AREA,birds$GRAZE,mean)
model.matrix(~modelout*factor(unique(birds$GRAZE)))
```

```
##   (Intercept)   modelout factor(unique(birds$GRAZE))2
## 1           1 260.384615                            1
## 2           1  22.137500                            0
## 3           1  12.266667                            0
## 4           1  10.000000                            0
## 5           1   4.846154                            0
##   factor(unique(birds$GRAZE))3 factor(unique(birds$GRAZE))4
## 1                            0                            0
## 2                            0                            0
## 3                            1                            0
## 4                            0                            1
## 5                            0                            0
##   factor(unique(birds$GRAZE))5 modelout:factor(unique(birds$GRAZE))2
## 1                            0                              260.3846
## 2                            1                                0.0000
## 3                            0                                0.0000
## 4                            0                                0.0000
## 5                            0                                0.0000
##   modelout:factor(unique(birds$GRAZE))3
## 1                               0.00000
## 2                               0.00000
## 3                              12.26667
## 4                               0.00000
## 5                               0.00000
##   modelout:factor(unique(birds$GRAZE))4
## 1                                     0
## 2                                     0
## 3                                     0
## 4                                    10
## 5                                     0
##   modelout:factor(unique(birds$GRAZE))5
## 1                                0.0000
## 2                               22.1375
## 3                                0.0000
```

```
## 4                                   0.0000
## 5                                   0.0000
## attr(,"assign")
##  [1] 0 1 2 2 2 2 3 3 3 3
## attr(,"contrasts")
## attr(,"contrasts")$`factor(unique(birds$GRAZE))`
## [1] "contr.treatment"
```

```
grazefactor <- as.factor(birds$GRAZE)


plot(ABUND~AREA, col=grazefactor,data=birds,pch=as.numeric(grazefactor),cex=1.25,
     xlim=c(0,150))
legend("bottomright",col=1:5,legend=1:5,pch=1:5)
```



Q8. Write the equation for each parameter in terms of the βs.

```
summary(fit.loyn)
```

```
##
## Call:
## lm(formula = ABUND ~ AREA * as.factor(GRAZE), data = birds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.8807  -2.7226  -0.2619   2.9237  11.3766
##
## Coefficients:
```

```
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              28.130741   1.892685  14.863  < 2e-16 ***
## AREA                      0.001891   0.003352   0.564 0.575445
## as.factor(GRAZE)2       -10.501178   3.253563  -3.228 0.002303 **
## as.factor(GRAZE)3       -11.485099   2.771458  -4.144 0.000145 ***
## as.factor(GRAZE)4       -20.401704   5.361642  -3.805 0.000417 ***
## as.factor(GRAZE)5       -24.061712   2.807900  -8.569 4.34e-11 ***
## AREA:as.factor(GRAZE)2    0.193273   0.070354   2.747 0.008555 **
## AREA:as.factor(GRAZE)3    0.376454   0.104908   3.588 0.000804 ***
## AREA:as.factor(GRAZE)4    1.282348   0.446409   2.873 0.006140 **
## AREA:as.factor(GRAZE)5    0.456881   0.251194   1.819 0.075449 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.055 on 46 degrees of freedom
## Multiple R-squared:  0.7339, Adjusted R-squared:  0.6818
## F-statistic:  14.1 on 9 and 46 DF,  p-value: 1.481e-10
```

```
anova(fit.loyn)
```

```
## Analysis of Variance Table
##
## Response: ABUND
##                        Df  Sum Sq Mean Sq F value    Pr(>F)
## AREA                    1  415.27  415.27 11.3257  0.001551 **
## as.factor(GRAZE)        4 3065.31  766.33 20.9004 7.129e-10 ***
## AREA:as.factor(GRAZE)   4 1170.73  292.68  7.9825 5.653e-05 ***
## Residuals              46 1686.62   36.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q9. Interpret the p-value.

Q10. Interpret the coefficients.

Q11. Describe how the influence of grazing changes with patch size.

Q12. Use the function interaction.plot to show how the influence of area changes depending on grazing level. Hint: Abundance will go on the y-axis, area on x-axis, and you'll have a trace for the different levels of grazing. Use the following data set:

```
modified.data <- data.frame(GRAZE = birds$GRAZE,
AREA = seq(min(birds$AREA), 100, length.out = 2))
predicted.values <- predict(fit.loyn, newdata = modified.data)
```