ENTMLGY 6707 Entomological Techniques and Data Analysis

Supplemental activity (do not submit): Split Plots

1 Introduction

Stratified designs are used to increase precision of estimates by experimentally controlling or reducing variation. This tutorial covers how to analyze split-plot and split-split-plot designs. The examples include multiple ways of analyzing the same exact data.

You might recall that F-values in ANOVAs are calculated from F-ratios (e.g., mean squares for treatment divided by means squares for errors). Historically, for more complicated designs like split-plots, folks had to compute the F-ratios "by-hand": these designs have whole plot and sub plot factors and errors, and so F-ratios for whole plot factors use the the whole plot error as the denominator whereas sub plot factors use the sub plot error. When examining the R output below for split plots and split-split plots, note which mean square values are used in the F-ratios.

Nowadays, some folks use mixed-effects models to analyze stratified designs. Be aware that a fixed-effect only approach might be preferred, given that we assume random effects are normally distriubuted and it is hard to test that assumption when there are only a few levels of a random effect (e.g., 3 blocks or 4 whole plots).

Either way, you still have to be careful about specifying the random effects correctly. And you will notice that if you do, the sums of squares and F-statistics for each treatment are typically equivalent. There are sometimes differences, but rarely do they influence the overall conclusions.

The following packages are necessary to complete this tutorial.

```
library(car)
library(lme4)
library(lmerTest)
library(tidyverse)
```

```r
library(agricolae)
library(emmeans)
```

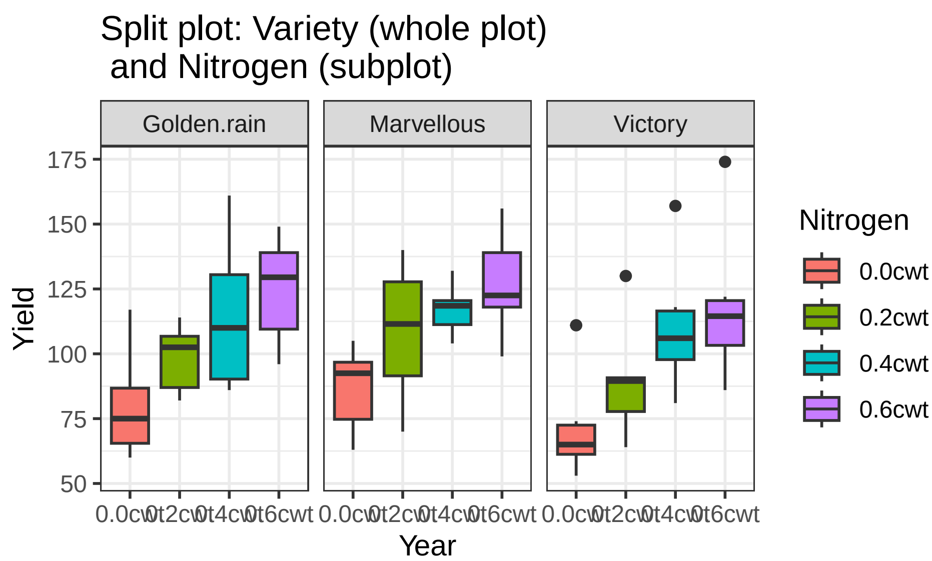
2 Split plot

These data report the yield of oats from a split-plot field trial. The treatment structure used in the experiment was a 3 x 4 full factorial, with three varieties of oats and four concentrations of nitrogen. The experimental units were arranged into six blocks, each with three whole-plots subdivided into four subplots. The varieties of oats were assigned randomly to the whole-plots and the concentrations of nitrogen to the subplots. All four concentrations of nitrogen were used on each whole-plot.

```
oats_df <- read.table("oats.txt", header=T, sep="\t",
                     colClasses = c("factor",   "factor","factor",
                                   "factor", "numeric"))
summary(oats_df)
```

```
 Block          Variety     Nitrogen      Yield      
 I  :12   Golden.rain:24   0.0cwt:18   Min.   : 53.0  
 II :12   Marvellous :24   0.2cwt:18   1st Qu.: 86.0  
 III:12   Victory    :24   0.4cwt:18   Median :102.5  
 IV :12                    0.6cwt:18   Mean   :104.0  
 V  :12                                3rd Qu.:121.2  
 VI :12                                Max.   :174.0  
```

```
ggplot(oats_df, aes(x=Nitrogen, y=Yield, fill=Nitrogen)) +
  geom_boxplot()+ theme_bw()+
  xlab("Year")+
  ylab("Yield")+
  facet_wrap(~Variety)+
  ggtitle("Split plot: Variety (whole plot) \n and Nitrogen (subplot)")
```



2.1 `Agricolae` package

```
library(agricolae)
sp.plot(oats_df$Block, oats_df$Variety, oats_df$Nitrogen, oats_df$Yield)
```

```
ANALYSIS SPLIT PLOT:  oats_df$Yield
Class level information

oats_df$Variety     :  Victory Golden.rain Marvellous
oats_df$Nitrogen    :  0.0cwt 0.2cwt 0.4cwt 0.6cwt
oats_df$Block   :  I II III IV V VI

Number of observations:  72

Analysis of Variance Table

Response: oats_df$Yield
                                 Df  Sum Sq Mean Sq F value    Pr(>F)
oats_df$Block                     5 15875.3  3175.1 17.9297 9.525e-10 ***
oats_df$Variety                   2  1786.4   893.2  1.4853    0.2724
Ea                               10  6013.3   601.3
oats_df$Nitrogen                  3 20020.5  6673.5 37.6856 2.458e-12 ***
```

```
oats_df$Variety:oats_df$Nitrogen  6   321.8    53.6  0.3028    0.9322
Eb                               45  7968.7   177.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cv(a) = 23.6 %, cv(b) = 12.8 %, Mean = 103.9722
```

2.2 aov()

```r
fit_aov_SP <- aov(Yield ~ Variety*Nitrogen + Error(Block/Variety), data=oats_df)
summary(fit_aov_SP)
```

```

Error: Block
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  15875    3175

Error: Block:Variety
          Df Sum Sq Mean Sq F value Pr(>F)
Variety    2   1786   893.2   1.485  0.272
Residuals 10   6013   601.3

Error: Within
                 Df Sum Sq Mean Sq F value   Pr(>F)
Nitrogen          3  20020    6673  37.686 2.46e-12 ***
Variety:Nitrogen  6    322      54   0.303    0.932
Residuals        45   7969     177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.3 mixed-effects model

```r
fit_lmer_SP <- lmer(Yield ~ Variety*Nitrogen + (1|Block) + (1|Block:Variety), data=oats_df)
anova(fit_lmer_SP, type=3)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
                  Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
Variety            526.1   263.0     2    10  1.4853    0.2724
```

```
Nitrogen         20020.5  6673.5     3    45 37.6857 2.458e-12 ***
Variety:Nitrogen   321.8    53.6     6    45  0.3028    0.9322
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.4 pairwise comparisons

```r
emmeans(fit_lmer_SP, pairwise~"Nitrogen")
```

```
NOTE: Results may be misleading due to involvement in interactions

$emmeans
 Nitrogen emmean   SE   df lower.CL upper.CL
 0.0cwt     79.4 7.17 6.79     62.3     96.5
 0.2cwt     98.9 7.17 6.79     81.8    116.0
 0.4cwt    114.2 7.17 6.79     97.2    131.3
 0.6cwt    123.4 7.17 6.79    106.3    140.5

Results are averaged over the levels of: Variety
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast        estimate   SE df t.ratio p.value
 0.0cwt - 0.2cwt   -19.50 4.44 45  -4.396  0.0004
 0.0cwt - 0.4cwt   -34.83 4.44 45  -7.853  <.0001
 0.0cwt - 0.6cwt   -44.00 4.44 45  -9.919  <.0001
 0.2cwt - 0.4cwt   -15.33 4.44 45  -3.457  0.0064
 0.2cwt - 0.6cwt   -24.50 4.44 45  -5.523  <.0001
 0.4cwt - 0.6cwt    -9.17 4.44 45  -2.067  0.1797

Results are averaged over the levels of: Variety
Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 4 estimates
```

3 Split-split splot

Prairie junegrass is a native species that has potential as a low input turf grass. Seed production strategies need to be investigated to determine best management practices. An experiment was conducted at the Sand Plain Research Center near Becker, MN to investigate varieties, fertility, and the use of growth regulators. Growth regulators in grasses tend to reduce plant height by influencing stem elongation. The experiment was a split split plot treatment design in a randomized complete block with 4 replications (blocks).

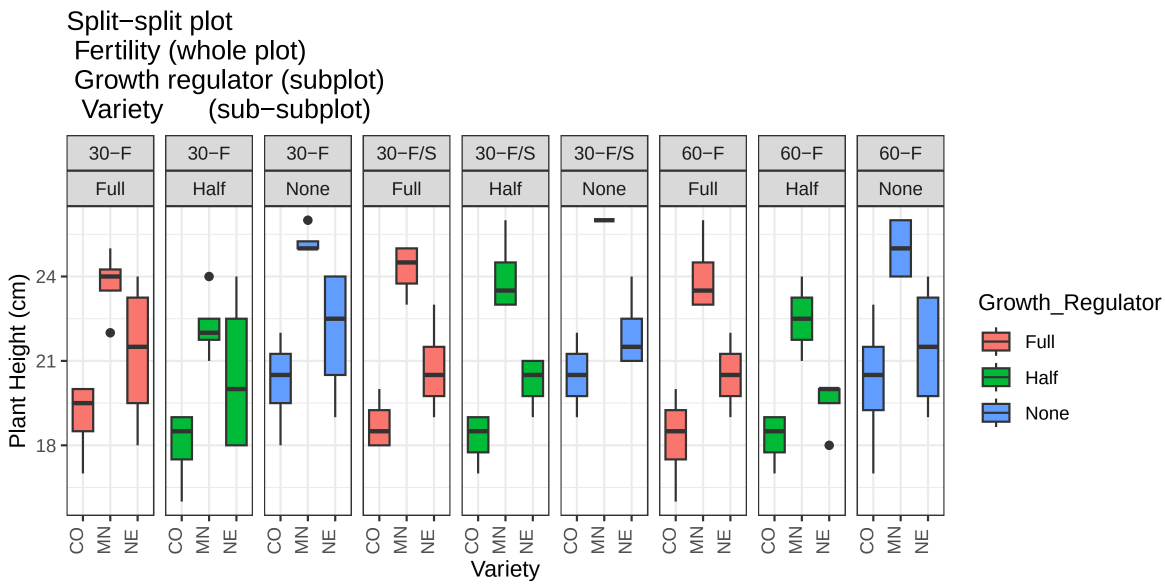
- Whole plots: Fertility
  - 30-F = 30 lbs N applied in Fall
  - 30-F/S = 30 lbs N applied in Fall + 30 lbs N applied in Spring
  - 60-F = 60 lbs N applied in Fall
- Subplots: Growth regulator applications
  - None = no Apogee growth regulator
  - Full = full rate (8 oz.)
  - Half = half rate (4 oz.)
- Sub-Subplots: 3 prairie junegrass populations
  - CO = Koeleria Colorada
  - NE = Koeleria Nebraska
  - MN = Koeleria Weaver Dunes, MN

```r
june_df <- read.table('EPP_junegrass.txt', header=T, sep="\t",
                      colClasses = c("factor","factor","factor","factor",
                                     "numeric", "numeric", "numeric"))

summary(june_df)
```

```
 Variety Growth_Regulator  Fertility   Rep    Seed_Yield_samples
 CO:36   Full:36          30-F  :36   1:27   Min.   : 2.00
 MN:36   Half:36          30-F/S:36   2:27   1st Qu.:11.00
 NE:36   None:36          60-F  :36   3:27   Median :15.00
                                      4:27   Mean   :15.81
                                             3rd Qu.:20.25
                                             Max.   :40.00
 Seed_Yield_lbs_per_acre  Plant_Height
 Min.   : 18.0           Min.   :16.00
 1st Qu.: 98.0           1st Qu.:19.00
 Median :134.0           Median :21.00
```

```
 Mean   :140.7          Mean   :21.36
 3rd Qu.:180.2          3rd Qu.:24.00
 Max.   :356.0          Max.   :26.00
```



3.1 `Agricolae` package

```
ssp.plot(june_df$Rep, june_df$Fertility, june_df$Growth_Regulator,
         june_df$Variety, june_df$Plant_Height)
```

```
ANALYSIS SPLIT-SPLIT PLOT:  june_df$Plant_Height
Class level information

june_df$Fertility   :  30-F 30-F/S 60-F
june_df$Growth_Regulator    :  None Full Half
june_df$Variety     :  CO NE MN
june_df$Rep     :  1 2 3 4

Number of observations:  108

Analysis of Variance Table
```

```
Response: june_df$Plant_Height
                                                         Df Sum Sq Mean Sq
june_df$Rep                                               3 109.88  36.627
june_df$Fertility                                         2   5.56   2.778
Ea                                                        6  13.26   2.210
june_df$Growth_Regulator                                  2  84.39  42.194
june_df$Fertility:june_df$Growth_Regulator                4   1.39   0.347
Eb                                                       18  18.44   1.025
june_df$Variety                                           2 470.72 235.361
june_df$Variety:june_df$Fertility                         4   5.72   1.431
june_df$Variety:june_df$Growth_Regulator                  4   2.72   0.681
june_df$Variety:june_df$Fertility:june_df$Growth_Regulator  8   2.67   0.333
Ec                                                       54  64.17   1.188
                                                          F value    Pr(>F)
june_df$Rep                                               30.8234 9.508e-12
june_df$Fertility                                          1.2570    0.3500
Ea
june_df$Growth_Regulator                                  41.1777 1.921e-07
june_df$Fertility:june_df$Growth_Regulator                 0.3389    0.8482
Eb
june_df$Variety                                          198.0701 < 2.2e-16
june_df$Variety:june_df$Fertility                          1.2039    0.3198
june_df$Variety:june_df$Growth_Regulator                   0.5727    0.6836
june_df$Variety:june_df$Fertility:june_df$Growth_Regulator   0.2805    0.9696
Ec

june_df$Rep                                              ***
june_df$Fertility
Ea
june_df$Growth_Regulator                                 ***
june_df$Fertility:june_df$Growth_Regulator
Eb
june_df$Variety                                          ***
june_df$Variety:june_df$Fertility
june_df$Variety:june_df$Growth_Regulator
june_df$Variety:june_df$Fertility:june_df$Growth_Regulator
Ec
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cv(a) = 7 %, cv(b) = 4.7 %, cv(c) = 5.1 %, Mean = 21.36111
```

3.2 aov()

```
fit_aov_SSP <-aov(Plant_Height   ~  Rep + Fertility*Growth_Regulator*Variety +
                   Error(Rep/Fertility/Growth_Regulator),data=june_df)
summary(fit_aov_SSP)
```

```
Error: Rep
    Df Sum Sq Mean Sq
Rep  3  109.9   36.63

Error: Rep:Fertility
          Df Sum Sq Mean Sq F value Pr(>F)
Fertility  2  5.556   2.778   1.257   0.35
Residuals  6 13.259   2.210

Error: Rep:Fertility:Growth_Regulator
                           Df Sum Sq Mean Sq F value   Pr(>F)
Growth_Regulator            2  84.39   42.19  41.178 1.92e-07 ***
Fertility:Growth_Regulator  4   1.39    0.35   0.339    0.848
Residuals                  18  18.44    1.02
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
                                   Df Sum Sq Mean Sq F value Pr(>F)
Variety                             2  470.7  235.36 198.070 <2e-16 ***
Fertility:Variety                   4    5.7    1.43   1.204  0.320
Growth_Regulator:Variety            4    2.7    0.68   0.573  0.684
Fertility:Growth_Regulator:Variety  8    2.7    0.33   0.281  0.970
Residuals                          54   64.2    1.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.3 mixed-effects model

```
fit_lmer_SSP <- lmer(Plant_Height ~ Fertility*Growth_Regulator*Variety +
                   (1|Rep) + (1:Rep:Fertility) + (1:Rep:Fertility:Growth_Regulator),
                   data=june_df)
anova(fit_lmer_SSP, type=3)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
                                   Sum Sq Mean Sq NumDF DenDF  F value
Fertility                            5.56   2.778     2    78   2.2600
Growth_Regulator                    84.39  42.194     2    78  34.3293
Variety                            470.72 235.361     2    78 191.4895
Fertility:Growth_Regulator           1.39   0.347     4    78   0.2825
Fertility:Variety                    5.72   1.431     4    78   1.1639
Growth_Regulator:Variety             2.72   0.681     4    78   0.5537
Fertility:Growth_Regulator:Variety   2.67   0.333     8    78   0.2712
                                      Pr(>F)
Fertility                             0.1111
Growth_Regulator                   2.022e-11 ***
Variety                            < 2.2e-16 ***
Fertility:Growth_Regulator            0.8885
Fertility:Variety                     0.3333
Growth_Regulator:Variety              0.6969
Fertility:Growth_Regulator:Variety    0.9735
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.4 pairwise comparisons

```r
emmeans(fit_lmer_SSP, pairwise~Variety)
```

```
NOTE: Results may be misleading due to involvement in interactions

$emmeans
 Variety emmean    SE   df lower.CL upper.CL
 CO        19.1 0.602 3.42     17.3     20.8
 MN        24.1 0.602 3.42     22.3     25.9
 NE        20.9 0.602 3.42     19.1     22.7

Results are averaged over the levels of: Fertility, Growth_Regulator, Rep
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast estimate    SE df t.ratio p.value
 CO - MN     -5.06 0.261 78 -19.347  <.0001
 CO - NE     -1.86 0.261 78  -7.122  <.0001
```

```
 MN - NE      3.19 0.261 78  12.225  <.0001

Results are averaged over the levels of: Fertility, Growth_Regulator, Rep
Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 3 estimates
```

4 R Activity

You will need the following packages to complete this problem set:

```r
library(tidyverse)
library(car)
library(lme4)
library(lmerTest)
library(emmeans)
```

An experiment was designed to assess the effect of nitrogen fertilizer on yield (grams) from three varieties of wheat. The experimental design was a split plot with replicates (=blocks). Five rates of nitrogen fertilizer were applied to whole plots at rates of 0, 40, 80, 120 and 160 kg/ha, and the three varieties were planted in sub plots. The data are in the "`EPP_yield.txt`" data file. Even though split plots are often designed to evaluate interactions, please ignore interactions for this activity.

1. Load in the data. Note that column names starting with `Rep` represent a block.

```
yield_df <- read.table('EPP_yield.txt', header=T, sep="\t",
                    colClasses = c("factor","factor",NA,NA,NA))
summary(yield_df)
```

```
 Nitrogen Variety      Rep1            Rep2             Rep3      
 0  :3    1:5     Min.   :3.60   Min.   :4.30   Min.   :4.200  
 120:3    2:5     1st Qu.:5.35   1st Qu.:5.30   1st Qu.:5.300  
 160:3    3:5     Median :6.40   Median :6.50   Median :5.900  
 40 :3            Mean   :6.58   Mean   :6.58   Mean   :6.313  
 80 :3            3rd Qu.:8.00   3rd Qu.:7.75   3rd Qu.:7.300  
                  Max.   :9.20   Max.   :9.70   Max.   :9.300  
```

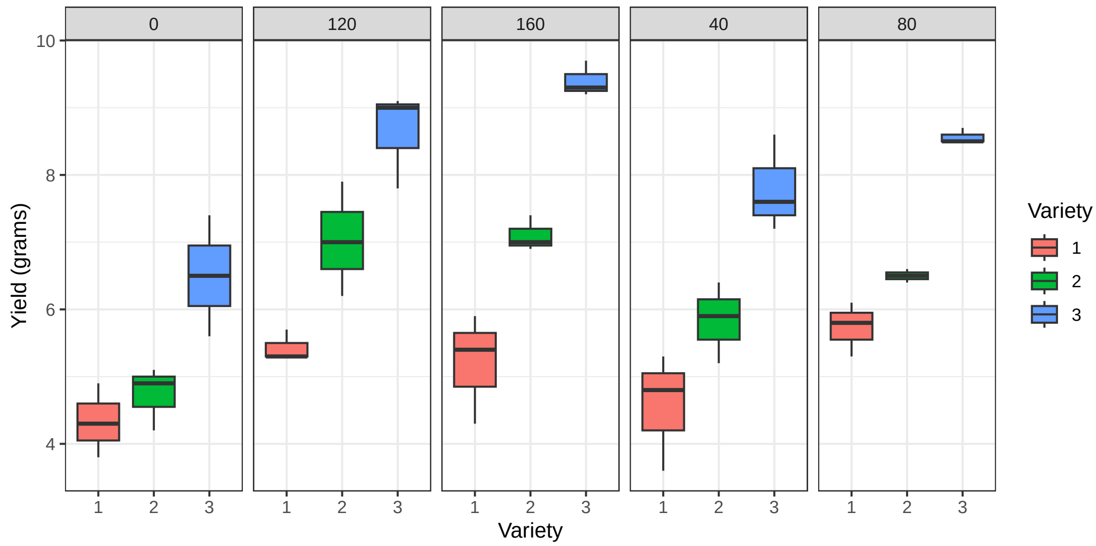
2. You might notice the data do not have one observation per row (which R expects when fitting linear models). Convert the data from "wide" to "long" format using R. Check out last week's R activity for some example code.

```
yield_df_long <- yield_df %>%
   pivot_longer(
   cols = starts_with("Rep"),
   names_to = "Rep",
   values_to = "Yield")
summary(yield_df_long)
```

```
 Nitrogen Variety     Rep                Yield      
 0  :9    1:15    Length:45          Min.   :3.600  
 120:9    2:15    Class :character   1st Qu.:5.300  
 160:9    3:15    Mode  :character   Median :6.400  
 40 :9                               Mean   :6.491  
 80 :9                               3rd Qu.:7.600  
                                     Max.   :9.700  
```

3. Graph the data using a boxplot. In the plot, group the data by the whole plot factor (one panel for each level of `Nitrogen`) and display the sub plot factor on the x-axis.

```r
ggplot(yield_df_long, aes(x=Variety, y=Yield, fill=Variety)) +
  geom_boxplot()+ theme_bw()+
  xlab("Variety")+
  ylab("Yield (grams)")+
  facet_wrap(~Nitrogen, ncol=5)
```



4. Conduct an analysis of variance (ANOVA) using the `aov()` command (i.e., assess if `Nitrogen` and `Variety` explain variation in `yield`). Treat `Nitrogen` as a factor for all analyses.

```r
aov_yield_1 <- aov(Yield ~ Rep + Nitrogen + Variety + Error(Rep/Nitrogen),
data=yield_df_long)
summary(aov_yield_1)
```

```
Error: Rep
    Df Sum Sq Mean Sq
Rep  2 0.7111  0.3556

Error: Rep:Nitrogen
          Df Sum Sq Mean Sq F value   Pr(>F)
Nitrogen   4 26.248   6.562   23.25 0.000184 ***
Residuals  8  2.258   0.282
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
```

```
          Df Sum Sq Mean Sq F value   Pr(>F)    
Variety    2  74.74   37.37   90.66 5.87e-13 ***
Residuals 28  11.54    0.41                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Conduct the same ANOVA as you did in the previous step but use the `lmer()` command from the `lme4` package and the `anova()` command from the `lmerTest` package. Note: if you get an error that says `boundary (singular) fit: see ?isSingular`, ignore it.

```
lmer_yield_1 <- lmer(Yield ~ Nitrogen + Variety +  (1|Rep) + (1|Rep:Nitrogen),
  data=yield_df_long)
```

```
boundary (singular) fit: see help('isSingular')
```

```
anova(lmer_yield_1, type=3)
```

```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)    
Nitrogen 26.248   6.562     4    38  17.185 3.976e-08 ***
Variety  74.739  37.370     2    38  97.868 1.024e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6. Pick one of the variables to conduct pairwise comparisons of treatment levels. Explain the reasoning for your choice and conduct the comparisons using the `lmer()` model.

```
emmeans(lmer_yield_1, pairwise~Nitrogen)
```

```
$emmeans
 Nitrogen emmean    SE df lower.CL upper.CL
 0          5.19 0.206 10     4.73     5.65
 120        7.03 0.206 10     6.57     7.49
 160        7.23 0.206 10     6.77     7.69
 40         6.07 0.206 10     5.61     6.53
 80         6.93 0.206 10     6.47     7.39

Results are averaged over the levels of: Variety 
Degrees-of-freedom method: kenward-roger 
Confidence level used: 0.95 

$contrasts
 contrast                    estimate    SE df t.ratio p.value
 Nitrogen0 - Nitrogen120       -1.844 0.291  8  -6.332  0.0015
 Nitrogen0 - Nitrogen160       -2.044 0.291  8  -7.018  0.0008
```

```
 Nitrogen0 - Nitrogen40      -0.878 0.291  8  -3.013  0.0914
 Nitrogen0 - Nitrogen80      -1.744 0.291  8  -5.989  0.0022
 Nitrogen120 - Nitrogen160   -0.200 0.291  8  -0.687  0.9539
 Nitrogen120 - Nitrogen40     0.967 0.291  8   3.319  0.0602
 Nitrogen120 - Nitrogen80     0.100 0.291  8   0.343  0.9964
 Nitrogen160 - Nitrogen40     1.167 0.291  8   4.005  0.0239
 Nitrogen160 - Nitrogen80     0.300 0.291  8   1.030  0.8352
 Nitrogen40 - Nitrogen80     -0.867 0.291  8  -2.975  0.0963

Results are averaged over the levels of: Variety
Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 5 estimates
```

```
emmeans(lmer_yield_1, pairwise~Variety)
```

```
$emmeans
 Variety emmean   SE df lower.CL upper.CL
 1         5.05 0.16 14     4.71     5.40
 2         6.24 0.16 14     5.90     6.58
 3         8.18 0.16 14     7.84     8.52

Results are averaged over the levels of: Nitrogen
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast            estimate    SE df t.ratio p.value
 Variety1 - Variety2    -1.19 0.226 28  -5.259  <.0001
 Variety1 - Variety3    -3.13 0.226 28 -13.857  <.0001
 Variety2 - Variety3    -1.94 0.226 28  -8.598  <.0001

Results are averaged over the levels of: Nitrogen
Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 3 estimates
```

7. Write 4-5 sentences comparing the conclusions from the two approaches (i.e., using `aov()` vs. `lmer()`) including conclusions drawn from any pairwise comparisons you conducted. At least 1-2 of your sentences should include a conclusion written in "biologically meaningful" terms.

   Answer: The analyses using `aov()` and `lmer()` were equivalent in terms of sums of squares and mean squares, but the F-values differed (FYI: it's due to the differences in residual degrees of freedom). Yield varied across nitrogen levels ($F_{4,38} = 17.19, p <$

0.0001) and between varieties ($F_{2,38} = 97.87, p < 0.0001$). Pairwise comparisons indicated that variety 3 produced a higher yield than variety 1 (Tukey's range test (TRT): $t_{28} = 13.86, p < 0.0001$) or variety 2 (TRT: $t_{28} = 8.60, p < 0.0001$). Even though nitrogen applications of 160 kg/ha were associated with the highest yield, yield at that rate did not differ from yield at applications 80 kg/ha (TRT: $t_8 = 1.03, p = 0.84$); thus, to maximize yield, plant variety 3 should be planted and returns on yield from nitrogen applications beyond 80-100 kg/ha might be considered negligible.