3) Anova

2023-04-10

Pre-processing the data-set

```
data <- read.csv("Anova_Dataset.csv", header = TRUE)
processed_data <- na.omit(data)
head(processed_data)
```

```
##   density block fertilizer    yield
## 1       1     1          1 177.2287
## 2       2     2          1 177.5500
## 3       1     3          1 176.4085
## 4       2     4          1 177.7036
## 5       1     1          1 177.1255
## 6       2     2          1 176.7783
```

```
summary(processed_data)
```

```
##     density         block        fertilizer     yield
##  Min.   :1.0   Min.   :1.00   Min.   :1    Min.   :175.4
##  1st Qu.:1.0   1st Qu.:1.75   1st Qu.:1    1st Qu.:176.5
##  Median :1.5   Median :2.50   Median :2    Median :177.1
##  Mean   :1.5   Mean   :2.50   Mean   :2    Mean   :177.0
##  3rd Qu.:2.0   3rd Qu.:3.25   3rd Qu.:3    3rd Qu.:177.4
##  Max.   :2.0   Max.   :4.00   Max.   :3    Max.   :179.1
```

```
str(processed_data)
```

```
## 'data.frame':    96 obs. of  4 variables:
##  $ density   : int  1 2 1 2 1 2 1 2 1 2 ...
##  $ block     : int  1 2 3 4 1 2 3 4 1 2 ...
##  $ fertilizer: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ yield     : num  177 178 176 178 177 ...
```

```
nrow(processed_data)
```

```
## [1] 96
```

Splitting the model

```
library(caret)
```

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

```
## Loading required package: lattice
```

```
indexs = createDataPartition(processed_data$yield, times = 1, p = 0.8, list = F)
#times = no. of times to be split
#p = percentage of data to be used for training, here 80% is used of training and 20%
for testing

train = processed_data[indexs, ]
nrow(train)
```

```
## [1] 80
```

```
test = processed_data[-indexs, ]
nrow(test)
```

```
## [1] 16
```

Creating the model - One way Anova

```
## ONE-WAY ANOVA
av1 <- aov(train$yield ~ train$density, data = train)
av1
```

```
## Call:
##    aov(formula = train$yield ~ train$density, data = train)
##
## Terms:
##                 train$density Residuals
## Sum of Squares        4.07647  32.65356
## Deg. of Freedom             1        78
##
## Residual standard error: 0.6470204
## Estimated effects may be unbalanced
```

```
summary(av1)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## train$density  1   4.08   4.076   9.738 0.00253 **
## Residuals     78  32.65   0.419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
av2 <- aov(train$yield ~ train$block, data = train)
av2
```

```
## Call:
##    aov(formula = train$yield ~ train$block, data = train)
##
## Terms:
##                 train$block Residuals
## Sum of Squares      0.13710  36.59294
## Deg. of Freedom           1        78
##
## Residual standard error: 0.6849381
## Estimated effects may be unbalanced
```

```
summary(av2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## train$block  1   0.14  0.1371   0.292   0.59
## Residuals   78  36.59  0.4691
```

```
av3 <- aov(train$yield ~ train$fertilizer, data = train)
av3
```

```
## Call:
##    aov(formula = train$yield ~ train$fertilizer, data = train)
##
## Terms:
##                 train$fertilizer Residuals
## Sum of Squares          6.977567 29.752471
## Deg. of Freedom                1        78
##
## Residual standard error: 0.6176099
## Estimated effects may be unbalanced
```

```
summary(av3)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## train$fertilizer  1  6.978   6.978   18.29 5.32e-05 ***
## Residuals        78 29.752   0.381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Creating the model - Two way Anova

```
## ONE-WAY ANOVA
av12 <- aov(train$yield ~ train$density + train$block + train$fertilizer, data = trai
n)
av12
```

```
## Call:
##    aov(formula = train$yield ~ train$density + train$block + train$fertilizer,
##     data = train)
##
## Terms:
##                 train$density train$block train$fertilizer Residuals
## Sum of Squares       4.076475    0.171089         6.638443 25.844032
## Deg. of Freedom             1           1                1        76
##
## Residual standard error: 0.5831407
## Estimated effects may be unbalanced
```

```
summary(av1)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## train$density  1   4.08   4.076   9.738 0.00253 **
## Residuals     78  32.65   0.419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

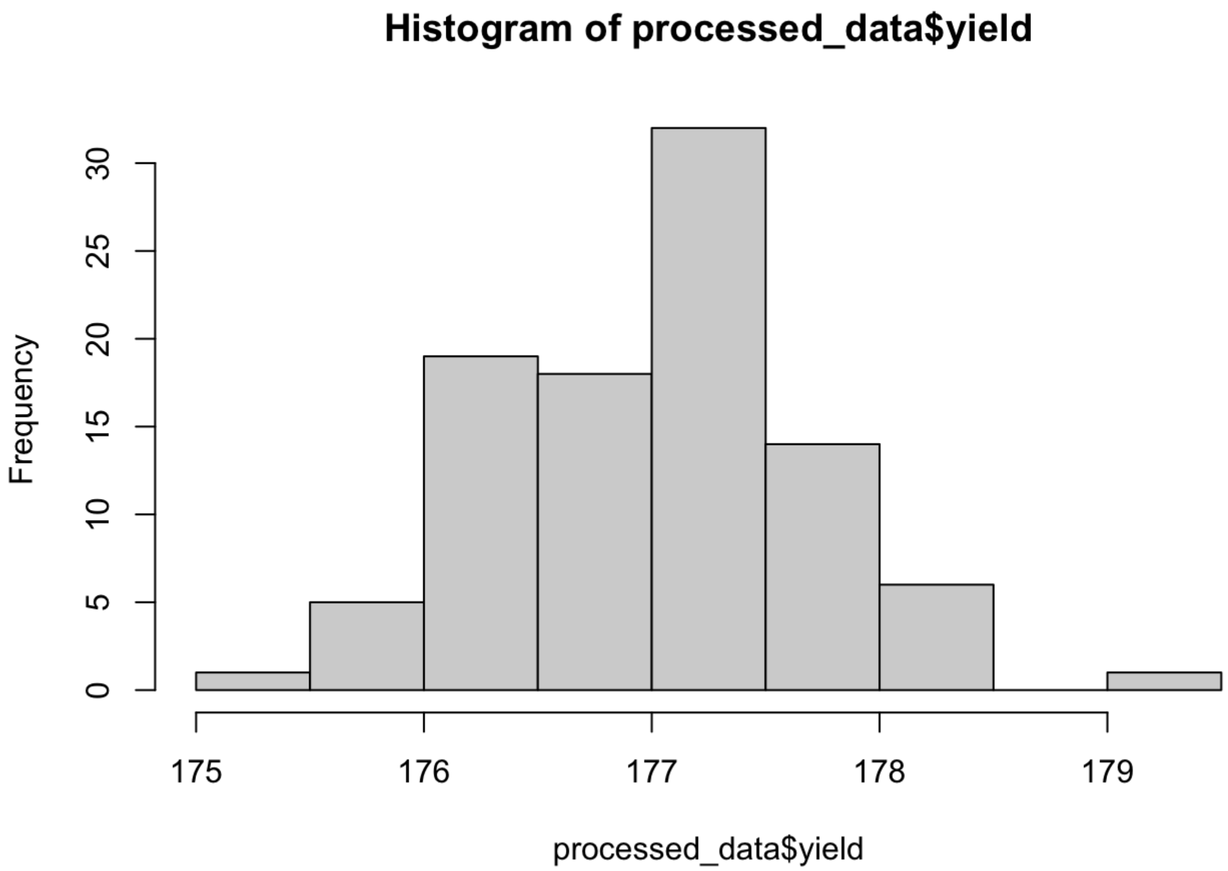
Finding the best fit

```
library(AICcmodavg)

one.way <- av3
two.way <- av12
intr <- aov(train$yield ~ train$density*train$fertilizer, data = train)

model.set <- list(one.way, two.way, intr)
model.names <- c('one.way', 'two.way', 'intr')

aictab(model.set, modnames = model.names)
```

```
##
## Model selection based on AICc:
##
##         K   AICc Delta_AICc AICcWt Cum.Wt     LL
## two.way 5 147.45       0.00   0.55   0.55 -68.32
## intr    5 147.89       0.45   0.44   0.98 -68.54
## one.way 3 154.22       6.77   0.02   1.00 -73.95
```

Creating histogram

```
hist(processed_data$yield)
```



Conclusion: We found a statistically-significant difference in average crop yield by both fertilizer type (F(2)=9.018, p < 0.001) and by planting density (F(1)=15.316, p < 0.001).