Two Way ANOVA with Interaction

Pranav Gopalkrishna, 1940223

2021-08-25

# **SETTING WORK DIRECTORY**

setwd("~/Documents/Study/computerScience/programming/r/data/")

# DATA SET

This data set contains measurements of the crop yields of various plots with different levels of soil density and different types of fertilizers applied (with the same intensity). It also contains data about the irrigation status (control implies unirrigated, which helps us determine the effect of irrigation when we want to study it, by comparing the irrigated plots to the unirrigated or control plots).

I will be focusing on three variables...

1. Crop yield (the response)
2. Fertilizer type (factor 1)
3. Soil density (factor 2)

I aim to determine, with 95% confidence, whether the mean yield can be said to differ significantly in general due to the effect of

* Fertilizer type
* Soil density
* Interaction between fertilizer type and soil density

myData = read.csv("cropYield.csv")  
head(myData)

## X yield block irrigation density fertilizer  
## 1 1 90 A control low N  
## 2 2 95 A control low P  
## 3 3 107 A control low NP  
## 4 4 92 A control medium N  
## 5 5 89 A control medium P  
## 6 6 92 A control medium NP

summary(myData)

## X yield block irrigation density fertilizer  
## Min. : 1.00 Min. : 60.00 A:18 control :36 high :24 N :24   
## 1st Qu.:18.75 1st Qu.: 86.00 B:18 irrigated:36 low :24 NP:24   
## Median :36.50 Median : 95.00 C:18 medium:24 P :24   
## Mean :36.50 Mean : 99.72 D:18   
## 3rd Qu.:54.25 3rd Qu.:114.00   
## Max. :72.00 Max. :136.00

x1 = myData$fertilizer  
x2 = myData$density  
y = myData$yield  
myData = data.frame(x1, x2, y)  
head(myData)

## x1 x2 y  
## 1 N low 90  
## 2 P low 95  
## 3 NP low 107  
## 4 N medium 92  
## 5 P medium 89  
## 6 NP medium 92

# HYPOTHESES

## For the effect of fertilizers alone

H\_0: There is no difference in the means of fertilizers

H\_1: Means are not equal with respect to fertilizer

## For the effect of soil density alone

H\_0: There is no difference in the means of densities

H\_1: Means are not equal with respect to density

## For the effect of interaction between soil density and fertilizer type

H\_0: There is no difference in the means of the interactions between density and fertilizer

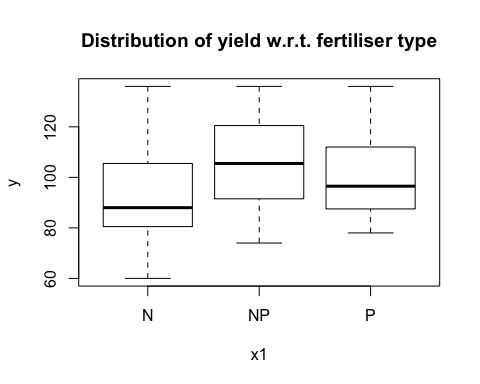
H\_1: There is difference in the means of the interactions between density and fertilizer

**Significance level = 0.05**

# VISUALIZING DISTRIBUTION OF YIELD W.R.T. FACTORS

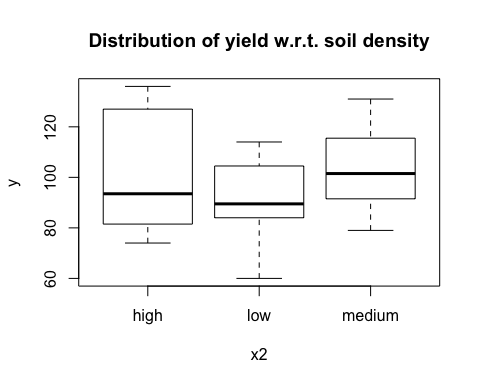
## With respect to fertilizer alone

boxplot(y~x1, main = "Distribution of yield w.r.t. fertiliser type")



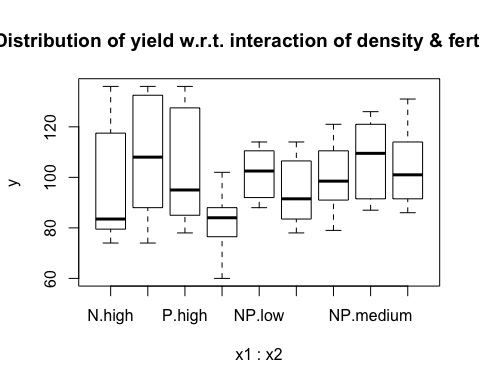
## With respect to soil density alone

boxplot(y~x2, main = "Distribution of yield w.r.t. soil density")



## With respect to the interaction of soil density and fertilizer type

boxplot(y~x1:x2, main = "Distribution of yield w.r.t. interaction of density & fertiliser")



# ANOVA

Now performing ANOVA test for both factors and their interaction...

model = aov(y~x1+x2+x1:x2)  
# x1:x2 denotes all the interactions of the levels of x1 and x2.  
summary(model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## x1 2 1977 988.7 3.159 0.0493 \*  
## x2 2 1758 879.2 2.809 0.0678 .  
## x1:x2 4 305 76.2 0.244 0.9125   
## Residuals 63 19716 312.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

As we can see, with 0.05 being the significance level, since Pr(>F) for x i.e. the p-value of x1 i.e. fertilizer is lower than 0.05, the effect of fertilizer type on the mean yield is statistically significant for the sample, and it may be concluded that there is difference in mean yield due to fertilizer type in general.

**Hence, we may reject H0 for the effect fertilizer type.**

However, for soil density and the interaction between fertilizer type and soil density, we see that p-values are greater than 0.05, meaning that their effects are not statistically significant for the sample, and it may be concluded that there is no difference in mean yield due to soil density or the interaction between soil density and fertilizer type in general.

**Hence, we may accept H0 for the effect of soil density, and for the effect of the interaction between soil density and fertilizer type.**

# POST-HOC ANALYSIS

Since only fertilizer type's effect is statistically significant, the only meaningful post-hoc analysis is on fertilizer type alone i.e. x1 alone. However, I will run the post-hoc analysis for all factors and their interaction, just to emphasize the previous conclusions. To perform post-hoc analysis, we use the Tukey HSD test, available in R through the package "multcompView".

library(multcompView)

TukeyHSD(model)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = y ~ x1 + x2 + x1:x2)  
##   
## $x1  
## diff lwr upr p adj  
## NP-N 12.750000 0.4921043 25.007896 0.0396887  
## P-N 7.666667 -4.5912290 19.924562 0.2972354  
## P-NP -5.083333 -17.3412290 7.174562 0.5824727  
##   
## $x2  
## diff lwr upr p adj  
## low-high -10.0416667 -22.299562 2.216229 0.1290335  
## medium-high 0.8333333 -11.424562 13.091229 0.9854304  
## medium-low 10.8750000 -1.382896 23.132896 0.0920046  
##   
## $`x1:x2`  
## diff lwr upr p adj  
## NP:high-N:high 12.000 -16.410887 40.410887 0.9094299  
## P:high-N:high 7.250 -21.160887 35.660887 0.9958478  
## N:low-N:high -14.000 -42.410887 14.410887 0.8103533  
## NP:low-N:high 5.125 -23.285887 33.535887 0.9996528  
## P:low-N:high -2.000 -30.410887 26.410887 0.9999998  
## N:medium-N:high 3.625 -24.785887 32.035887 0.9999747  
## NP:medium-N:high 10.750 -17.660887 39.160887 0.9501099  
## P:medium-N:high 7.375 -21.035887 35.785887 0.9953338  
## P:high-NP:high -4.750 -33.160887 23.660887 0.9998028  
## N:low-NP:high -26.000 -54.410887 2.410887 0.0987353  
## NP:low-NP:high -6.875 -35.285887 21.535887 0.9971220  
## P:low-NP:high -14.000 -42.410887 14.410887 0.8103533  
## N:medium-NP:high -8.375 -36.785887 20.035887 0.9891213  
## NP:medium-NP:high -1.250 -29.660887 27.160887 1.0000000  
## P:medium-NP:high -4.625 -33.035887 23.785887 0.9998386  
## N:low-P:high -21.250 -49.660887 7.160887 0.3008216  
## NP:low-P:high -2.125 -30.535887 26.285887 0.9999996  
## P:low-P:high -9.250 -37.660887 19.160887 0.9795542  
## N:medium-P:high -3.625 -32.035887 24.785887 0.9999747  
## NP:medium-P:high 3.500 -24.910887 31.910887 0.9999807  
## P:medium-P:high 0.125 -28.285887 28.535887 1.0000000  
## NP:low-N:low 19.125 -9.285887 47.535887 0.4422993  
## P:low-N:low 12.000 -16.410887 40.410887 0.9094299  
## N:medium-N:low 17.625 -10.785887 46.035887 0.5541399  
## NP:medium-N:low 24.750 -3.660887 53.160887 0.1365391  
## P:medium-N:low 21.375 -7.035887 49.785887 0.2933961  
## P:low-NP:low -7.125 -35.535887 21.285887 0.9963151  
## N:medium-NP:low -1.500 -29.910887 26.910887 1.0000000  
## NP:medium-NP:low 5.625 -22.785887 34.035887 0.9993120  
## P:medium-NP:low 2.250 -26.160887 30.660887 0.9999994  
## N:medium-P:low 5.625 -22.785887 34.035887 0.9993120  
## NP:medium-P:low 12.750 -15.660887 41.160887 0.8771267  
## P:medium-P:low 9.375 -19.035887 37.785887 0.9777871  
## NP:medium-N:medium 7.125 -21.285887 35.535887 0.9963151  
## P:medium-N:medium 3.750 -24.660887 32.160887 0.9999672  
## P:medium-NP:medium -3.375 -31.785887 25.035887 0.9999854

As can be seen, only the difference of means between the fertilizer types NP and N is statistically significant, since only p-value associated with this pair i.e. 0.0396887 is below the significance level 0.05.

As concluded, the differences of means between the pairs of the levels of the other factor and the interaction between factors are all statistically insignificant, as seen by their p-values always exceeding the significance level 0.05.