TWO FACTORIAL DESIGN TO STUDY THE EFFECTS OF FERTILIZER TYPE & SOIL DENSITY ON CROP YIELD

Pranav Gopalkrishna, 1940223

2021-11-16

# INTRODUCTION

Fertilizers can contain phosphates, nitrates or both. We will be studying the effects of these three fertilizer types on crop yield. At the same time, we will be studying the effects of soil density on crop yield, and study the effect of the interactions of these two factors.

Each fertilizer type and soil density level are applied randomly to crops within each block, and the crop yield is measured for each interaction of fertilizer type, soil density and block (block will not be considered as a factor, as it is a logistical aspect of the experiment, and is not supposed to affect the end result).

# OBJECTIVE

The objective of this study is to estimate the significance of the above factors, both separately and together, in terms of their effects on crop yield.

## Two factorial design

A two-factor factorial design is an experimental design in which data is collected for all possible combinations of the levels of the two factors of interest. In our case, we also have two factors of interest, namely fertilizer type and soil density.

Data for crop yield is collected for all combinations of these factors’ levels, hence having a two-factorial experimental design. 2^3 factorial design in particular specifies that there are two factors, and each factor has three levels, which is the case with our data set here.

# 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)  
   **Varieties:**
   1. Nitrates (N)
   2. Phosphates (P)
   3. Nitrates & phosphates (NP)
3. Soil density (factor 2)  
   **Levels:**
   1. Low
   2. Medium
   3. High

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

setwd("~/Documents/Study/computerScience/programming/r/data/")  
myData = read.csv("cropYield.csv")  
head(myData)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TABLE 1: DATA TABLE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **X** | **Yield** | **Block** | **Irrigation** | **Density** | **Fertilizer** |
| 1 | 90 | A | control | low | N |
| 2 | 95 | A | control | low | P |
| 3 | 107 | A | control | low | NP |
| 4 | 92 | A | control | medium | N |
| 5 | 89 | A | control | medium | P |
| 6 | 92 | A | control | medium | NP |

## **Summary of data set**

summary(myData)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TABLE 2: DATA SUMMARY**

|  |  |  |  |
| --- | --- | --- | --- |
| **Yield** | **Irrigation** | **Density** | **Fertilizer** |
| Min. : 60.00 | control : 36 | high :24 | N :24 |
| 1st Qu.: 86.00 | irrigated: 36 | low :24 | NP:24 |
| Median : 95.00 |  | medium:24 | P :24 |
| Mean : 99.72 |  |  |  |
| 3rd Qu.:114.00 |  |  |  |
| Max. :136.00 |  |  |  |

## Variables

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

# ANOVA

## HYPOTHESES

### For fertilizer

H0: There is no difference in the means of fertilizers

H1: Means are not equal with respect to fertilizer

### For soil density

H0: There is no difference in the means of densities

H1: Means are not equal with respect to density

### For interaction

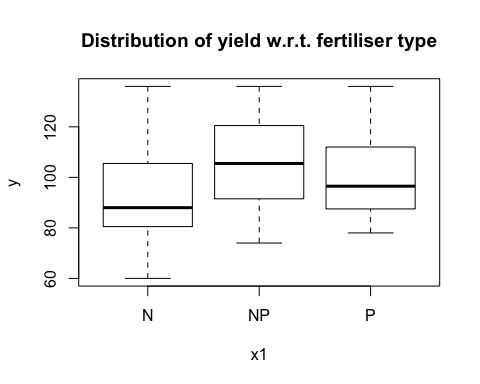
H0: There is no difference in the means of the interactions between density and fertilizer

H1: There is difference in the means of the interactions between density and fertilizer

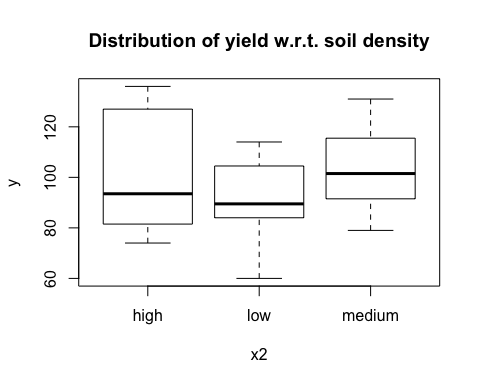
***Significance level = 0.05***

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

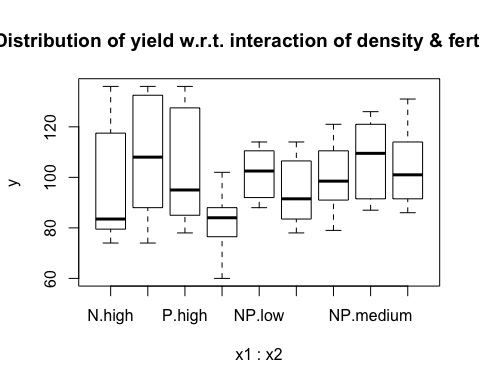
boxplot(y~x1, main = "Distribution of yield w.r.t. fertiliser type")



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



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



Hence, we see that both factors' levels have visibly different mean responses, and even their interactions have significantly distinct means on the plots. This gives an indication that the factors and their levels have notable effects on the crop yield.

## ANOVA

To create the ANOVA test model to test the significance of our 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)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TABLE 3: ANOVA TABLE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum squared** | **Mean squared** | **Calc. F value** | **P** |
| Fertilizer | 2 | 1977 | 988.7 | 3.159 | 0.0493 |
| Density | 2 | 1758 | 879.2 | 2.809 | 0.0678 |
| Interaction | 4 | 305 | 76.2 | 0.244 | 0.9125 |
| Error | 63 | 19716 | 312.9 |  |  |

The F-value is derived from the mean squared of the factor and the mean squared error of the model. It is used to test the significance of the differences in the mean responses between the levels of the factor. The p-value is the probability of getting the given F-value (probability is derived using the standard F-distribution, and the significance defines the proportion of lowest probability values that must be considered as outliers or "significantly different" for our test).

# INITIAL CONCLUSIONS

Judging by the p-value, we can say that we may only reject the null hypothesis for fertilizer type. In other words, only fertilizer types have mean crop yields that are significantly different from each other, indicating a significant impact of fertilizer on crop-yield.

# POST-HOC ANALYSIS

Since null hypothesis is rejected for fertilizer type, now we will aim to find out which fertilizer types in particular have significantly different mean crop yields, using the Tukey Honest Significance Difference test.

library(multcompView)

model = aov(y~x1)  
TukeyHSD(model)

Tukey multiple comparisons of means  
(95% family-wise confidence level)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TABLE 4: TUKEY HSD POST HOC ANALYSIS TABLE**

|  |  |  |
| --- | --- | --- |
| **Levels** | **Difference in means** | **P-adjusted value** |
| NP-N | 12.750000 | 0.0401813 |
| P-N | 7.666667 | 0.2996603 |
| P-NP | -5.083333 | 0.5849013 |

# FINAL CONCLUSIONS

Here, the p-value indicates whether the difference is significant or not, by comparing it to a critical difference (if the actual difference is above this, then it is significant). We can see that only NP and P have significantly different mean crop yields, for a 0.05 significance level.

## Inspecting values

**Selecting subsets of the table with fertilizer types NP, N and P...**  
np = subset(myData, x1 == 'NP')  
n = subset(myData, x1 == 'N')  
p = subset(myData, x1 == 'P')

**Comparing mean crop yields...**  
mean(np$y)

## [1] 105.6667

mean(n$y)

## [1] 92.91667

mean(p$y)

## [1] 100.5833

As we can see, the mean crop yields for NP and N are much further apart, while the mean crop yield for P lies between both, and is hence not significantly different from either one. In any case, it is likely that NP fertilizer type would produce more crop yield than N fertilizer type, since their difference is significant. However, N may be a substitute for either one, and hence, we can conclude that NP and P fertilizer types are the most desirable for increasing crop yield.