Report 1. Advanced Data Analysis

Lucas W Hahn De Castro. UNI: lwh2119

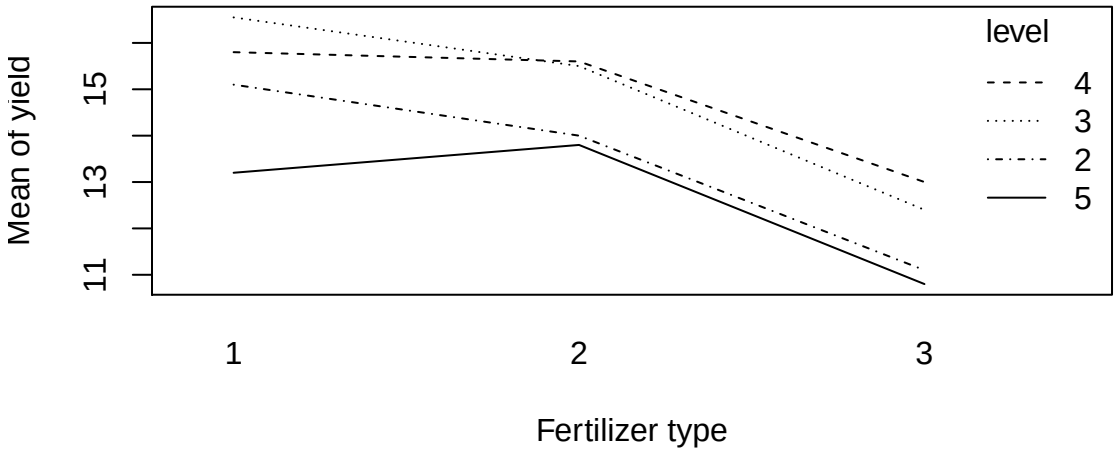
Due March 15, 2019

Summary:

The goal of this study is to determine the effect on the yield of a crop produced by two factors: the fertilizer type and the level of the fertilizer application. The results of the analysis show that there is no interaction between the factors, and bot the fertilizer and its level have significant effects in the yield of the crop. The fertilizer and level of application that produce the highest yield are fertilizer 1 and Level 3 (200 pounds per acre).

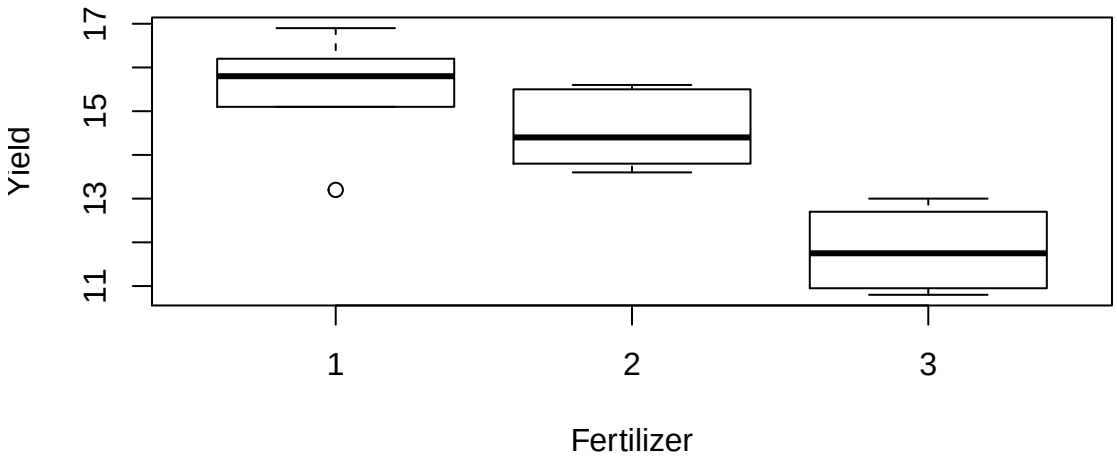
Data description and exploratory data analysis:

• The data used in this analysis was obtained from a randomized experiment. In this experiment, three different brands of fertilizer where applied with 4 different levels of intensity (levels of 2,3,4 and 5). Each level corresponds to 100, 200, 300 and 400 pounds per acre. The experiment was run in 14 different plots, which were randomly assigned to a fertilizer type and a level of application.
• The results of this experiment are important because they will contribute to a better understanding of the best agricultural practices that will help to improve the yield of the crop under study. In other words, by analyzing the results of the experiment we will be able to determine a type of fertilizer and a level of intensity that can optimize the production of the crop.
• An exploratory analysis of the data is conducted by visually examining the relatinship between the yield and the two different factors under consideration. The following plot is an interaction plot, which shows the mean of the yield for each type of the brands of fertilizer and the levels of application. The plot suggests that the brand of the fertilizer might have an interaction with the level of the intensity, given the change of the trend between the types of the fertilizer. Nonetheless, a statistical assessment of the presence of an interaction will be further conducted in the following section.



• Moreover, we can also see that there are differences in the means of the yield for each combination of fertilizers and levels. The next table shows the mean of the yield for each of the combinations in the data, and the following boxplot suggests that there might be significant differences according to the fertilizer used:

		Levels			
		2	3	4	5
Fertilizers	1	15.1	16.55	15.8	13.2
	2	14.0	15.5	15.6	13.8
	3	11.1	12.4	13.0	10.8



Analysis of the results:

• The final estimated model is:

$$Yield = 13.92 + 1.32F_1 + 0.77F_2 - 0.56L_2 + 1L_3 + 0.88L_4$$

Where F_1, F_2 represent dummy variables for Fertilizers 1 and 2, and L_1, L_2, L_3 represent dummy variables for Levels 2, 3 and 4. The base variables selected for the analysis were Fertilizer 3 and Level 5.
• This model was obtained by following the procedure described next. Given that the design of the experiment is unbalanced, initially we used a two-way analysis of variance (ANOVA) procedure with Type 3 sum of squares. The results of the procedure are shown in the following table:

Source	df	SS	F	P-value
Intercept	1	2530.8	8958.6	0.0001
Fertilizer	2	27.5	48.7	0.02
Level	3	11.42	13.5	0.07
Interaction	6	1.63	0.96	0.59
Residuals	2	0.56		

• The results of the Anova analysis show that there is not enough statistical evidence to conclude the existence of an interaction between the fertilizer and the intensity of its application. This because the p-value of the interaction terms is not statistically significant at any usual level of confidence (0.1, 0.05 or 0.001). Therefore, we proceed to estimate the model using the same procedure but excluding the interaction terms. The results of the analysis are the following:

Source	df	SS	F	P-value
Intercept	1	2534.9	9616.6	0.000
Fertilizer	2	28.67	52.3	0.000
Level	3	12.56	15.28	0.001
Residuals	8	2.19		

• The results of the analysis show that both the fertilizer and the level of application have a highly statistically significant effect on the yield, using a significance level of 0.01. This means that there exists a combination of fertilizer and level of application that produce a higher yield than the rest of combinations. Therefore, we proceed to revise the coefficients of the regression to determine the combination that produces the highest yield.
• The results of the regression analysis are the following. The selected base variables to compare to in the analysis were fertilizer 3 and level 5, the highest one. We can see that all the coefficients are statistically significant, which means that al the other treatments are different from the base treatment. An examination of the magnitude and sign of the coefficients shows that the fertilizer with the highest yield was type 1, and the level of application with the highest yield was level 3. The coefficients of the model and their p-values are shown next.

Variable	Intercept	F_1	F_2	L_2	L_3	L_4
Coefficient	13.919	1.320	0.773	-0.562	1.000	0.880
p-value	0.000	0.000	0.005	0.045	0.003	0.009

Conclusion

• This analysis concluded with a high statistical significance that the fertilizer type and its level of application have different effects on the yield of a particular crop. It was found that both factors do not show evidence of an existence of interacting effects between them. Moreover, the treatment that was found to produce the highest yield was the one with fertilizer 1 with a level of application of 200 pounds per acre.