

Danmarks
Tekniske
Universitet



Analysis of Correlated Data Assignment 3

AUTHOR

Marios-Dimitrios Lianos
s233558

November 8, 2023

Summary

For the requirements of Assignment 3 in the Analysis of Correlated Data: Mixed Linear Models class, a statistical examination was executed. The focus of the study was to identify the different effects of nitrogen fertilizer levels on rice yield when nitrogen is considered as a categorical variable, to when considered as a numerical. A field was systematically divided into blocks, within which varied levels of nitrogen fertilizer were applied, and different rice varieties were sown. The analysis is structured into three sections. The initial segment involves conducting model diagnostics to identify a mixed linear model that accurately reflects our evaluations. The subsequent part encompasses the core statistical analysis, while the final section is dedicated to drawing conclusions based on the findings.

Contents

1	Introduction	1
2	Exploratory Data Analysis	2
2.1	Variables in the Dataset	2
2.2	Explorative Plots and Diagrams	2
2.3	Split-Plot Model	5
2.3.1	The Model when N is considered as Categorical	6
2.3.2	The Model when N is considered as Numerical	6
3	Statistical Analysis	7
3.1	Statistical Analysis: N Treated as a Categorical Variable	7
3.1.1	Model Diagnostics	7
3.1.2	ANOVA	11
3.1.3	Post-hoc Analysis	11
3.2	Statistical Analysis: N Treated as a Numerical Variable	15
3.2.1	Model Diagnostics	15
3.2.2	ANOVA	18
3.2.3	Post-hoc Analysis	19
4	Conclusion	21
5	Appendix A	22

1 Introduction

Our study aims to explore the impacts of various nitrogen fertilizer levels and rice varieties on rice yield. Conducted meticulously across multiple blocks, each experiment was subjected to six distinct nitrogen fertilizer levels and involved the cultivation of four different rice varieties. The objective is to acquire reliable data to refine and enhance fertilizer application strategies and rice variety selections in agriculture. The analysis unfolds in two pivotal segments. Initially, the nitrogen fertilizer level (N) is analyzed as a categorical variable, enabling a comparison between distinct, non-ordered groups. Subsequently, we transition into interpreting N as a numerical variable, allowing for an evaluation based on a continuous, ordered scale. This dual approach aims to unveil the respective benefits and drawbacks of each model, providing an understanding of their influences on rice yield.

2 Exploratory Data Analysis

Initially, the nitrogen fertilizer levels will be treated as categorical variables ($N = 1, 2, 3, 4, 5, 6$). This approach allows us to categorize the different levels of fertilizer into discrete, non-ordered groups and analyze their impact on rice yield without assuming a specific order or quantity. Subsequently, we will treat the nitrogen fertilizer levels as numerical variables, associating each level with specific quantitative values representing the actual amount of fertilizer applied ($1 = 0, 2 = 60, 3 = 90, 4 = 120, 5 = 150, 6 = 180$).

2.1 Variables in the Dataset

- **Block:** Categorical variable representing the specific sections of the field where the experiment was conducted.
- **N (Nitrogen Fertilizer Level):** In the first part we treat the variable as a categorical one and in the second part as numerical. It indicates the level of nitrogen fertilizer applied.
- **Var (Rice Variety):** This categorical variable represents the specific rice variety used in each part of the experiment.
- **Yield:** This is a continuous variable that represents the measured yield of rice.

It is crucial to highlight that the dataset is balanced. Each combination of nitrogen fertilizer levels and rice varieties is well-represented across the blocks, ensuring a comprehensive and nuanced understanding of their impacts on rice yield. In addition, Block is crossed with both N levels and Var, as each Block contains all levels of the other two factors. N is crossed with Var, as each nitrogen level is paired with every rice variety.

- **Fixed Effects:** Nitrogen Fertilizer Level (N) and Rice Variety (Var), as these are the primary variables of interest.
- **Random Effects:** Block, to account for spatial or environmental variability.

2.2 Explorative Plots and Diagrams

Fig: 1 reveals that the Yield is not that much affected by the factor Block while in Fig: 2 and 3 we see that Yield increases as N increases except for $N=6=180$ where a small decrease is observed and as Var increases Yield decreases.

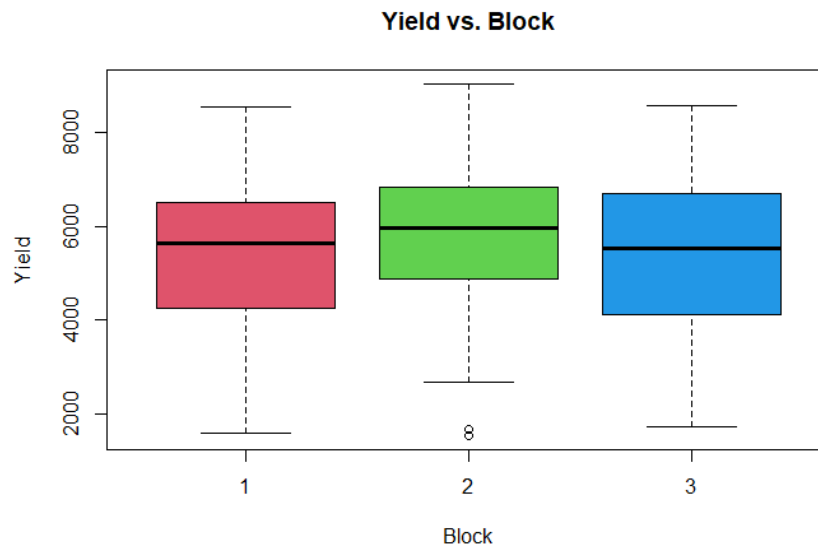


Figure 1: Boxplot Yield - Block

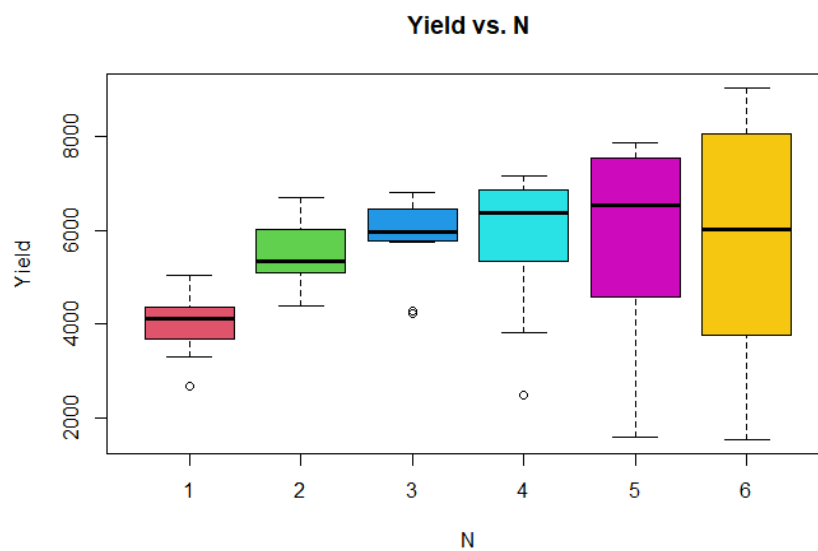


Figure 2: Boxplot Yield - N

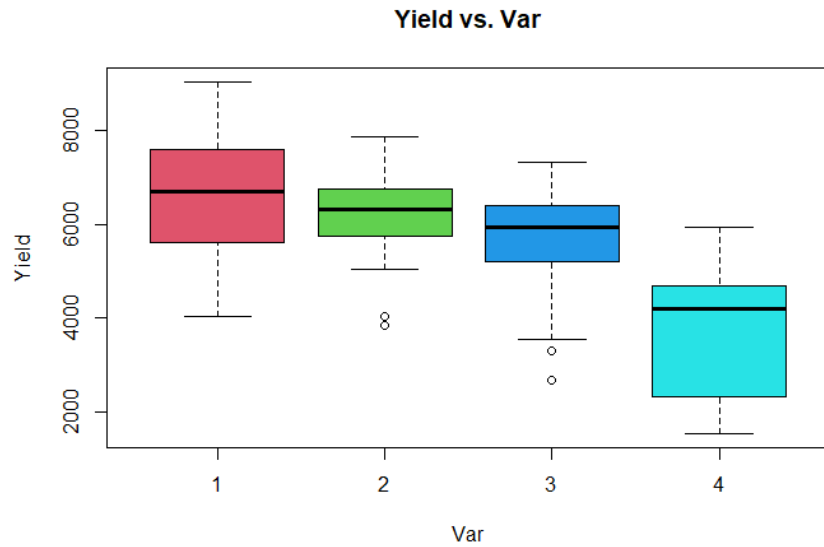


Figure 3: Boxplot Yield - Var

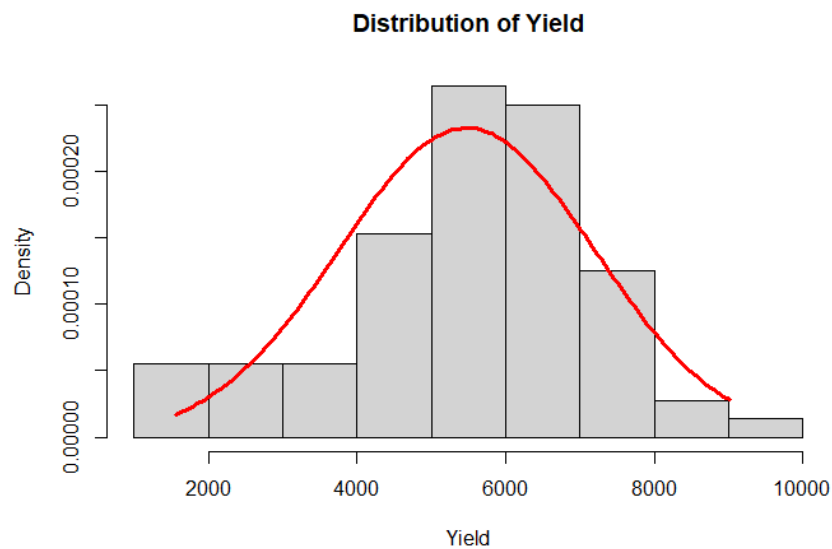


Figure 4: Histogram

Fig: 4 shows that the distribution of Yield is not exactly normally distributed and this may be caused by the influence of N and Var on the Yield.

2.3 Split-Plot Model

Representation of the model:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \epsilon_{ijkl} \quad (1)$$

- N as a categorical variable :

Y_{ijkl} is the observed yield, μ is the overall mean yield, α_i is the effect of the i -th level of nitrogen fertilizer (N), where $i = 1, 2, 3, 4, 5, 6$. β_j is the effect of the j -th level of rice variety (Var), where $j = 1, 2, 3, 4$. γ_k is the effect of the k -th block, where $k = 1, 2, 3$. $(\alpha\beta)_{ij}$ is the interaction effect between the nitrogen fertilizer levels and the rice varieties, and $(\alpha\gamma)_{ik}$ is the nitrogen fertilizer levels and the blocks. ϵ_{ijkl} is the random error term. The interaction $(\beta\gamma)_{ij}$ doesn't make any realistic sense, so we will exclude it from the model.

- N as a numerical variable :

α_i represents the effect of a one-unit increase in the numerical value of the nitrogen fertilizer level (Nc), where the values of i correspond to specific quantities of nitrogen fertilizer applied, such as 0, 60, 90, 120, 150, and 180.

In this split-plot design (1), the Block and nitrogen fertilizer levels N form the whole plot and whole plot factor, respectively, while the combinations of these with different rice varieties Var form the subplots, with the rice varieties acting as the subplot factors.

The factor diagram of the model (1) is given in Fig: 5. The upper numbers indicate the levels of the factor while the sub-numbers show the degrees of freedom.

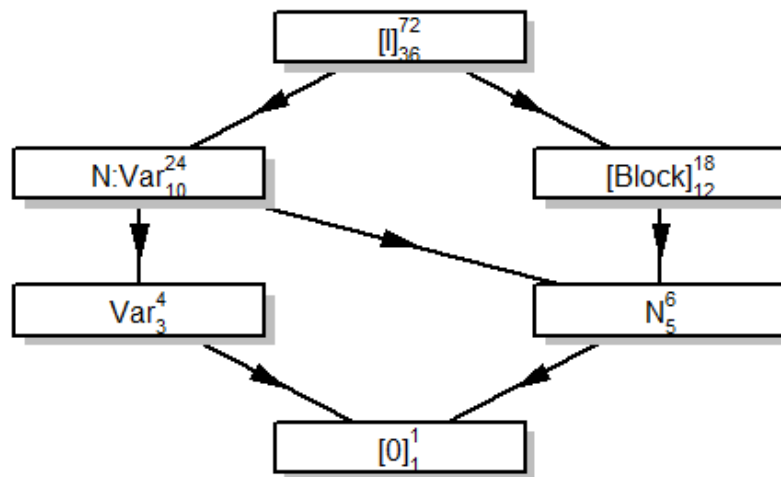


Figure 5: Factor Diagram

2.3.1 The Model when N is considered as Categorical

We'll conduct likelihood tests to see if any factor or an interaction effect should not be considered in the model (1).

Firstly we will compare model (1) with the model without the interaction of Block and N.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkl} \quad (1)$$

'log Lik.' 0.1702522. As we can see the interaction effect of Block and N is not statistically significant.

In addition, we compare the model without the interaction of Block and N with a model that does not contain the factor Block.

'log Lik.' 0. So it wouldn't help our analysis to consider a model that does not have the factor Block.

The final model when N is a categorical variable:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkl} \quad (2.1)$$

2.3.2 The Model when N is considered as Numerical

Here, we will first consider the model in which Nc has the values (0, 60, 90, 120, 150, 180). Then we will consider in the model the quadratic term for nitrogen concentration $I(Nc)^2$.

Finally, we will compare those two models to decide which one we will keep.

From the above procedure, we get the table:

Model	npar	AIC	BIC
interactionNc1.lmer	11	1169.4	1194.4
interactionNc2.lmer	12	1155.9	1183.3

Table 1: Comparison of the two models

Lower values of AIC and BIC suggest a better model fit to the data. As a result, the model with the quadratic term is more promising.

By conducting the same procedure as the one that we've done to retrieve model (2.1) we get the expression for the final model when N is considered to be a numerical variable:

$$Y_{ijkl} = \mu + \alpha_i + I(Nc)_i^2 + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkl} \quad (2.2)$$

3 Statistical Analysis

Our analysis is divided into two comprehensive parts, each focusing on a unique treatment of the factor N. In the first segment of our study, factor N is analyzed as a categorical variable. Following this, in the second segment, we pivot our approach, treating N as a numerical variable. Each segment of the analysis is structured into distinct phases: initial model diagnostics, result interpretation, and a conclusive post-hoc analysis.

3.1 Statistical Analysis: N Treated as a Categorical Variable

3.1.1 Model Diagnostics

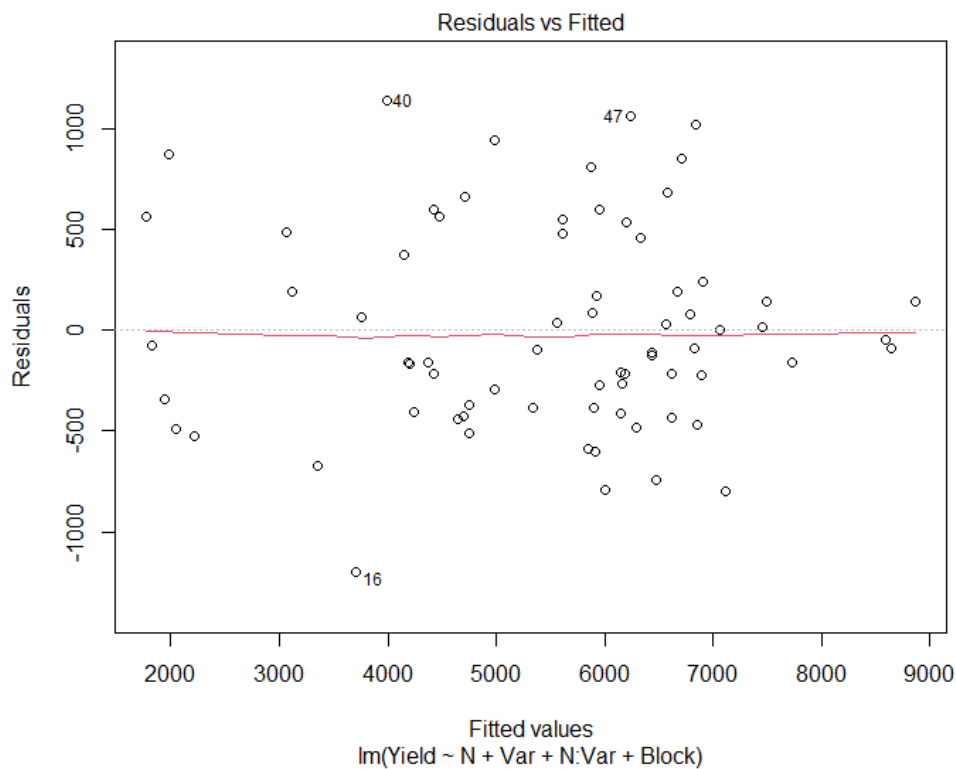


Figure 6: Residuals vs Fitted values

Fig: 6 and Fig: 7 displays residuals that are largely symmetrical in their distribution, closely aligning with a normal distribution, except for a few notable outliers at the extremes. Fig: 10 and 11 show that N and Var do not have big differences in variability.

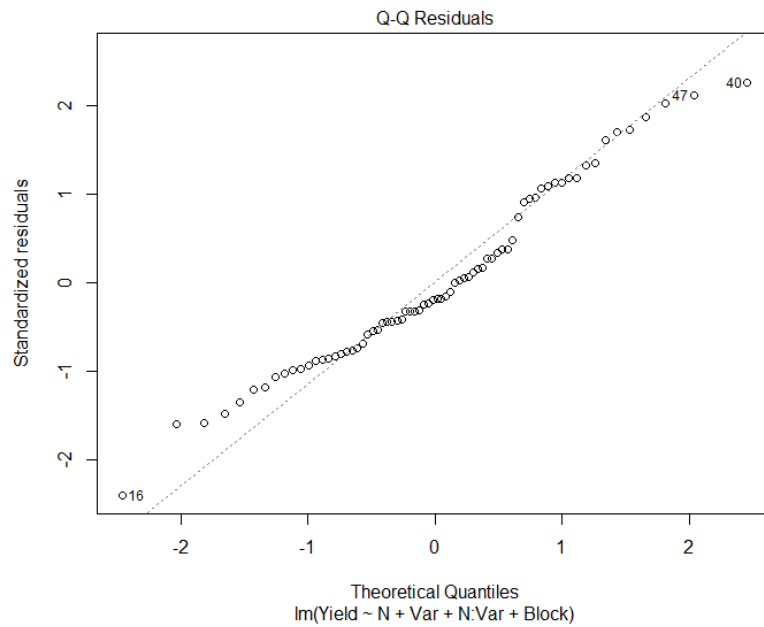


Figure 7: Q-Q Residuals

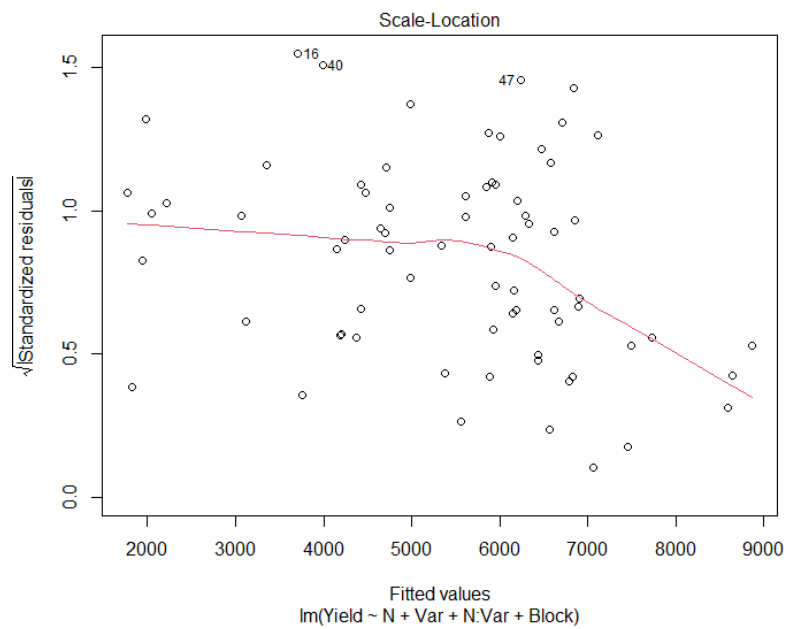


Figure 8: Scale-Location

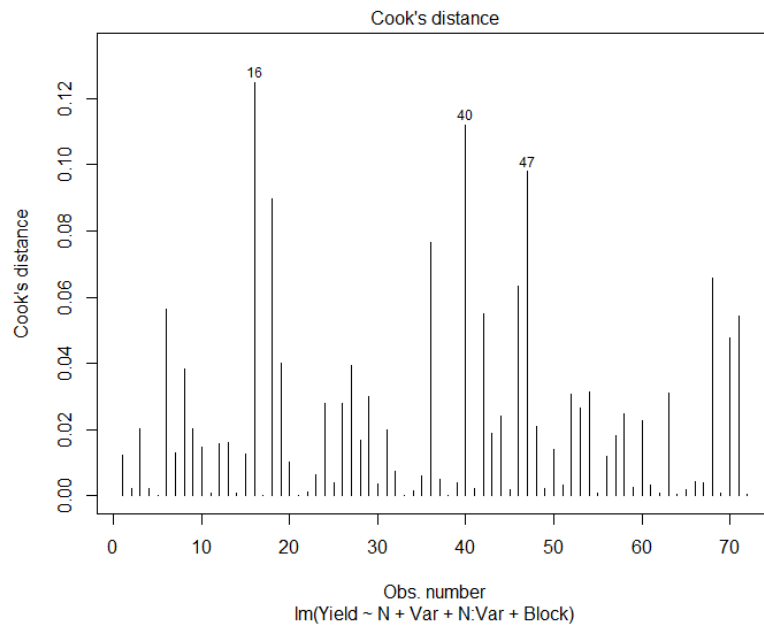


Figure 9: Cook's distance

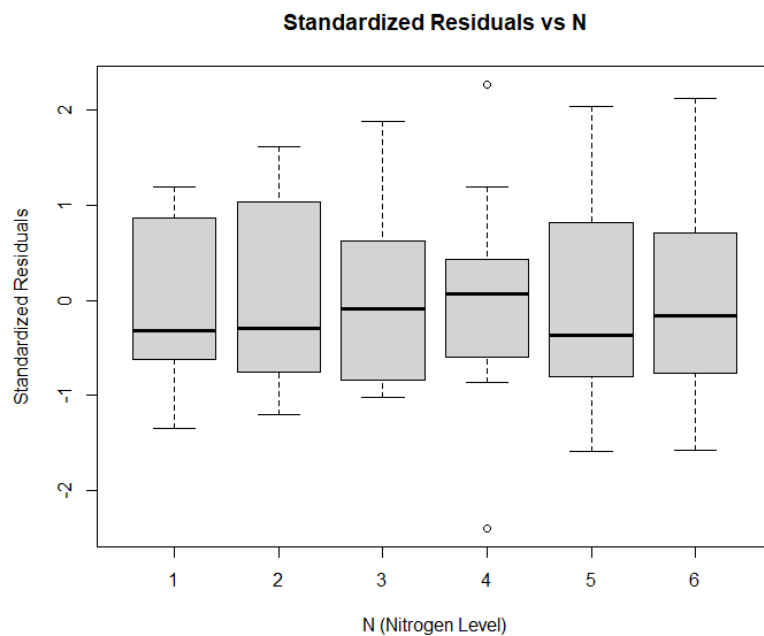


Figure 10: Studentized Residuals vs N

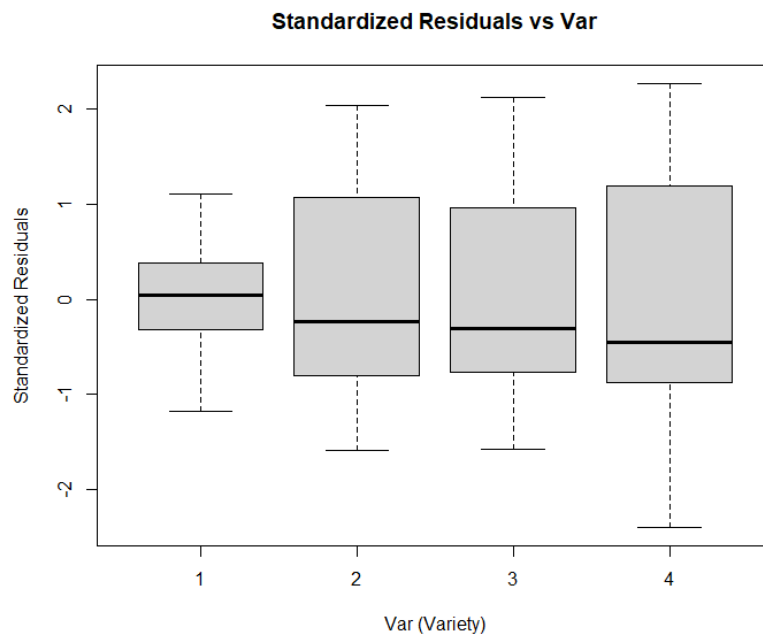


Figure 11: Studentized Residuals vs Var

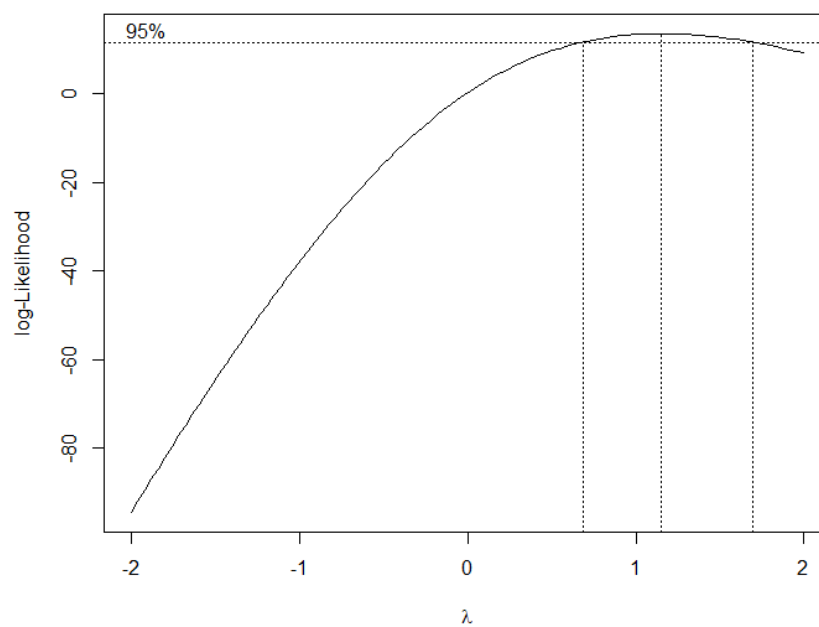


Figure 12: BoxCox

After conducting model diagnostics it didn't seem necessary to modify our model.

3.1.2 ANOVA

Performing the ANOVA model analysis as outlined in Equation (2.1) we have the results in the following tables:

Model	logLik	AIC	LRT	Pr(>Chisq)
<none>	-390.84	833.68		
(1 Block)	-390.89	831.78	0.10794	0.7425

Table 2: Anova for Random Effects

Source	NumDF	DenDF	F value	Pr(>F)
N	5	46	15.468	6.509e-09 ***
Var	3	46	76.022	< 2.2e-16 ***
N:Var	15	46	11.736	4.472e-11 ***

Table 3: Anova for Fixed Effects.

The analysis reveals that both N and Var, along with their interaction, are statistically significant,

3.1.3 Post-hoc Analysis

At a 95% profile likelihood confidence limits, the uncertainties of the estimated values on the standard deviation scale

N	Estimate	SE	Lower	Upper
1	4054	187	3671	4437
2	5478	187	5095	5861
3	5866	187	5483	6249
4	5864	187	5481	6248
5	5812	187	5429	6195
6	5797	187	5414	6180

Table 4: N Estimations

Var	Estimate	SE	Lower	Upper
1	6554	155	6227	6880
2	6156	155	5829	6482
3	5563	155	5237	5890
4	3642	155	3315	3969

Table 5: Var Estimations

Estimation of the Fixed Parameters

Tukey-adjusted pairwise comparisons were used to assess differences between levels of the N and Var providing a detailed evaluation of the distinct categories.

contrast	Estimate	SE	p.value
Var1 - Var2	398	209	0.2413
Var1 - Var3	990	209	0.0001
Var1 - Var4	2911	209	<.0001
Var2 - Var3	592	209	0.0336
Var2 - Var4	2513	209	<.0001
Var3 - Var4	1921	209	<.0001

Table 6: Var Contrast

contrast	Estimate	SE	p.value
N1 - N2	-1423.83	256	<.0001
N1 - N3	-1811.92	256	<.0001
N1 - N4	-1810.08	256	<.0001
N1 - N5	-1757.67	256	<.0001
N1 - N6	-1742.42	256	<.0001
N2 - N3	-388.08	256	0.6571
N2 - N4	-386.25	256	0.6616
N2 - N5	-333.83	256	0.7821
N2 - N6	-318.58	256	0.8134
N3 - N4	1.83	256	1.0000
N3 - N5	54.25	256	0.9999
N3 - N6	69.50	256	0.9998
N4 - N5	52.42	256	0.9999
N4 - N6	67.67	256	0.9998
N5 - N6	15.25	256	1.0000

Table 7: N Contrast

From Tables 6 and 7 and with Figures 13 and 14, N1 appears to have a significantly negative impact on yield compared to the other nitrogen levels tested. The pattern of results suggests that Var1 and Var2 have notably higher effects than Var3 and Var4, with Var1 having the highest mean effect of all. Var4 has the lowest mean effect according to these results. The statistical significance tells us that these are not likely to be random findings, but instead reflect true differences in the effects of these variable levels.

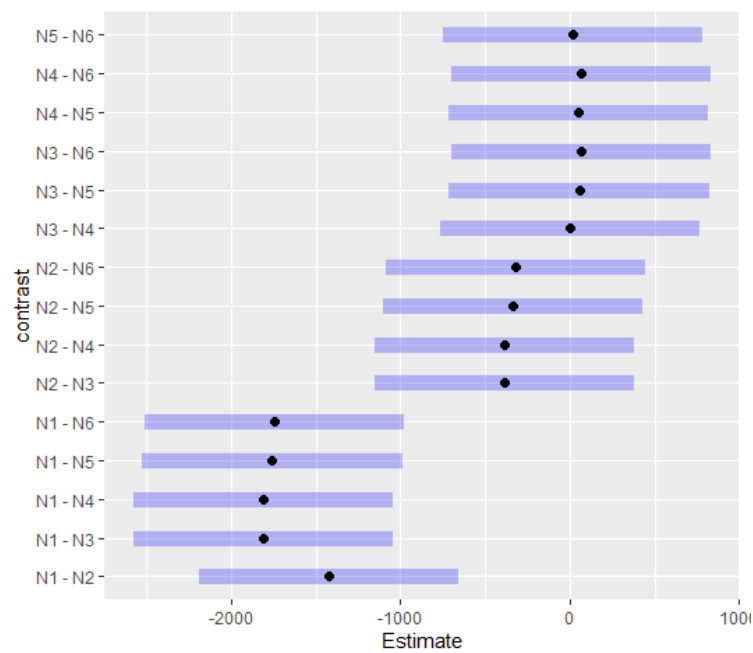


Figure 13: Contrast Plot for N

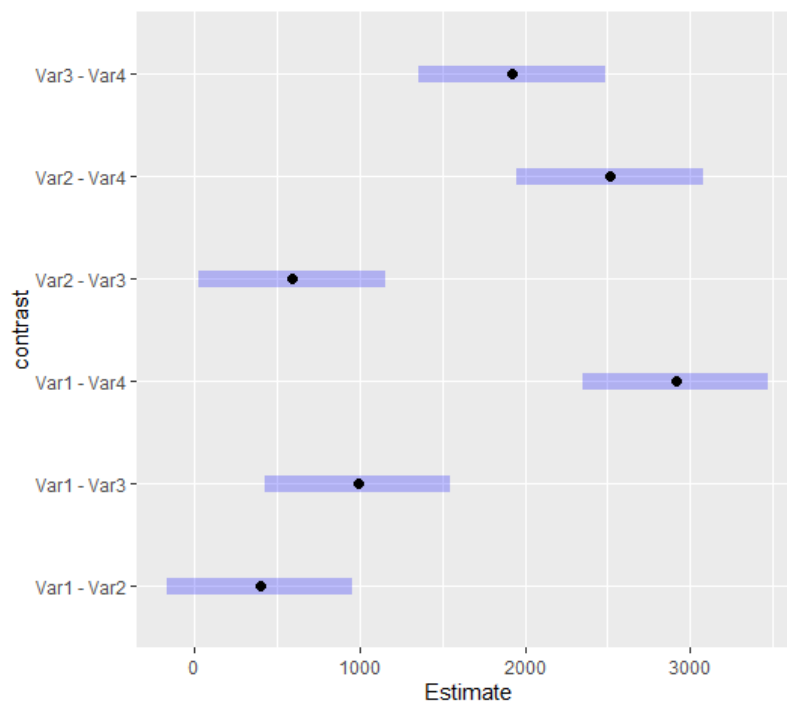


Figure 14: Contrast Plot for Var

N1 is significantly different from all other groups.

N	Estimate
1	4054a
2	5478b
6	5797b
5	5812b
4	5864b
3	5866b

Table 8: Pair-wise N comparison summary

Var 4 has the lowest mean, Var 3 is significantly different from Var 4 but not from Var 2 and Var 1. Var 2 is not significantly different from Var 1 or Var 3. Var 1, having the highest mean, is not significantly different from Var 2 but is from Var 3 and Var 4.

Var	Estimate
4	3642a
3	5563b
2	6156bc
1	6554c

Table 9: Pair-wise Var comparison summary

3.2 Statistical Analysis: N Treated as a Numerical Variable

3.2.1 Model Diagnostics

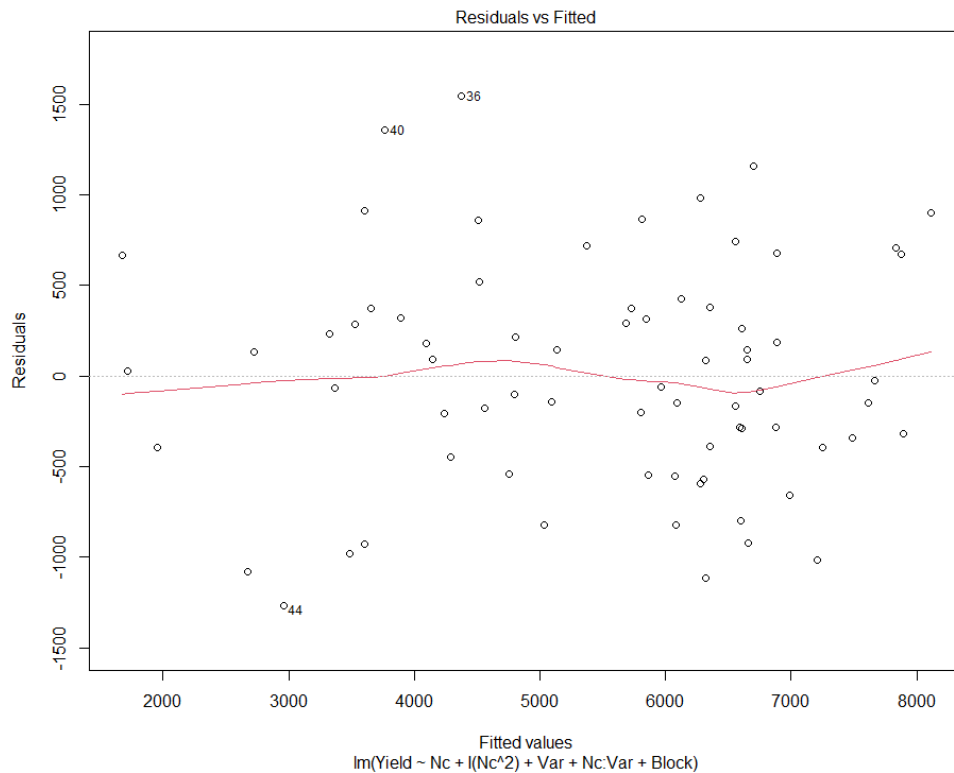


Figure 15: Residuals vs Fitted values

Fig: 15 and Fig: 16 illustrate residuals that are largely symmetrical in their distribution, closely aligning with a normal distribution, except for a few notable outliers at the extremes. Fig: 19 shows that Var does not have big differences in variability.

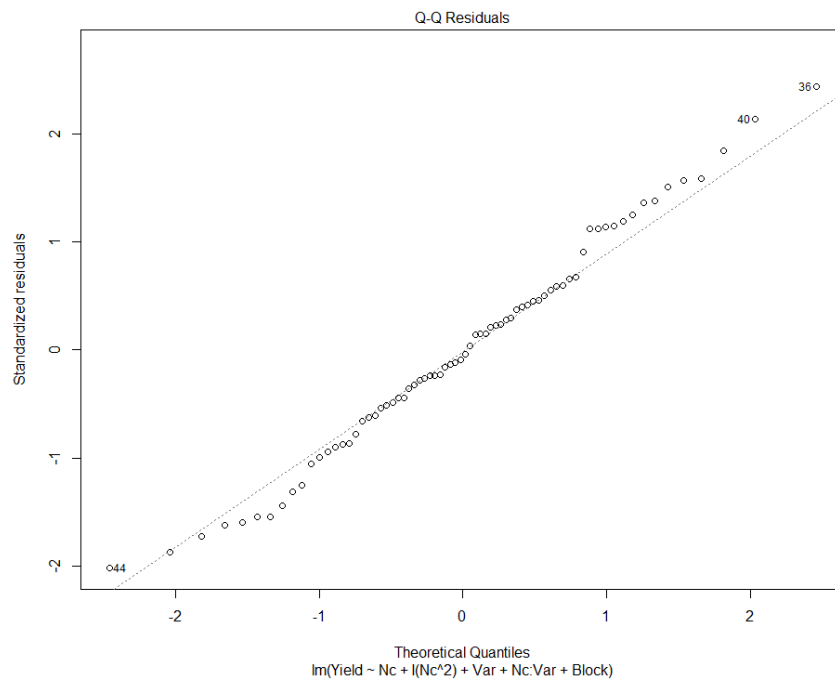


Figure 16: Q-Q Residuals

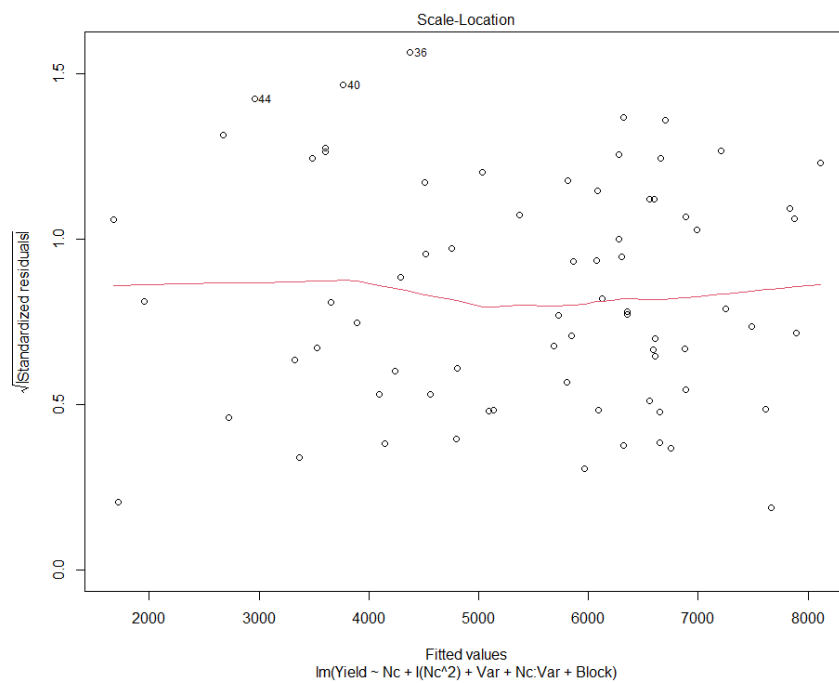


Figure 17: Scale-Location

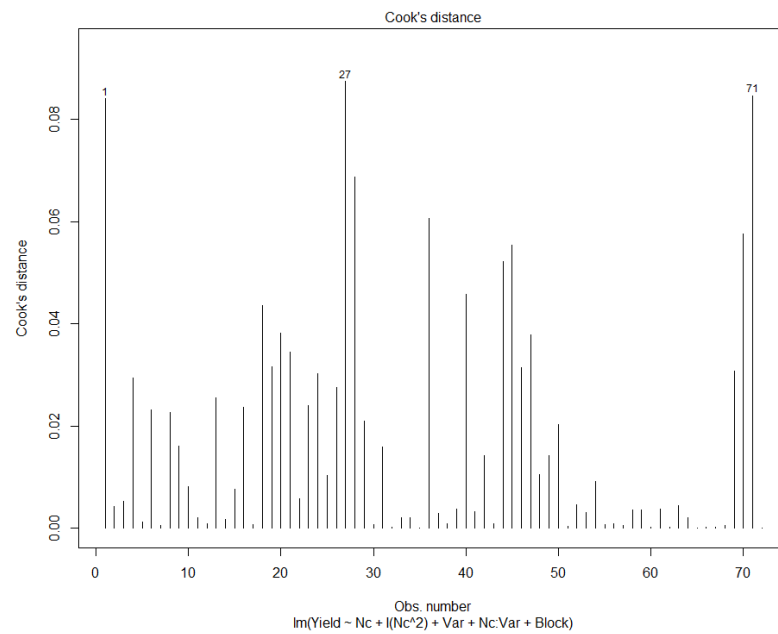


Figure 18: Cooks distance

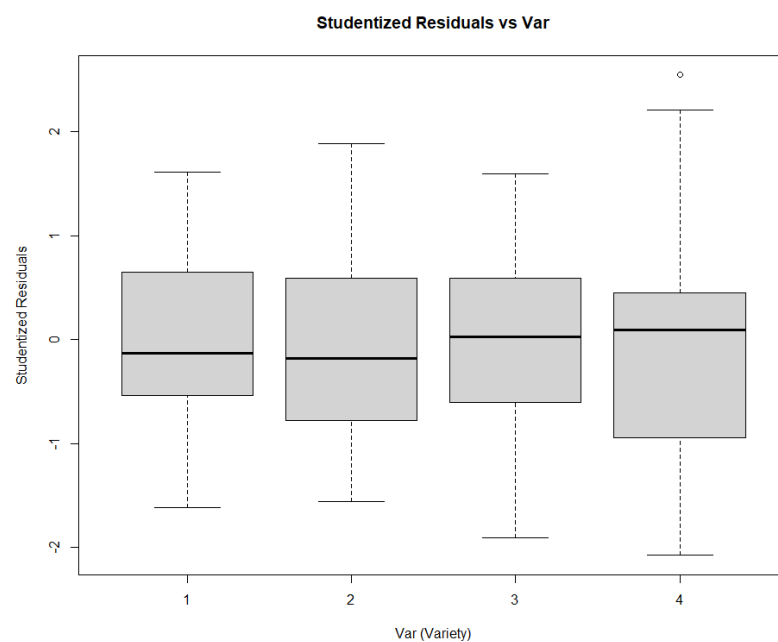


Figure 19: Studentized Residuals vs Var

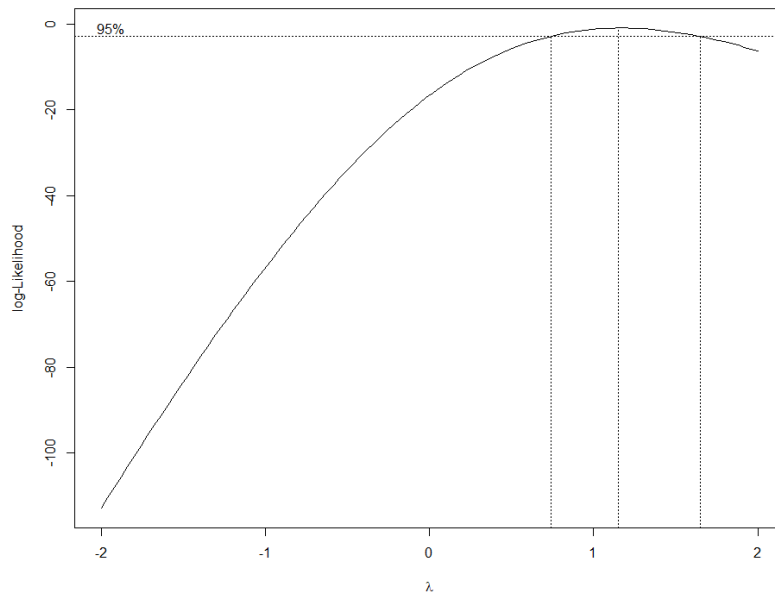


Figure 20: BoxCox

3.2.2 ANOVA

In Table 11 we see that Nc, $I(Nc)^2$, Var, and their interaction are all statistically significant.

Model	logLik	AIC	LRT	Pr(>Chisq)
<none>	-537.56	1097.1		
(1 Block)	-537.58	1095.2	0.038501	0.8444

Table 10: Anova for Random Effects.

Term	NumDF	DenDF	F value	Pr(>F)
Nc	1	61	41.8731	1.864e-08 ***
I(Nc ²)	1	61	21.6842	1.785e-05 ***
Var	3	61	4.3135	0.007991 **
Nc:Var	3	61	45.2938	1.591e-15 ***

Table 11: Anova for Fixed Effects.

3.2.3 Post-hoc Analysis

Estimation of the Fixed Parameters

Tukey-adjusted pairwise comparisons were used to assess differences between levels of Var.

Var	Estimate	SE	Lower	Upper
1	6935	178	6579	7291
2	6537	178	6181	6893
3	5945	178	5589	6301
4	4023	178	3667	4380

Table 12: Var Estimations

Contrast	Estimate	SE	p.value
Var1 - Var2	398	224	0.2924
Var1 - Var3	990	224	0.0002
Var1 - Var4	2911	224	<.0001
Var2 - Var3	592	224	0.0490
Var2 - Var4	2513	224	<.0001
Var3 - Var4	1921	224	<.0001

Table 13: Var Contrast

The contrast estimates and their corresponding p-values provide evidence of significant differences among the varieties: Var 1 vs. Var 3 and 4: The large and statistically significant positive estimates indicate that Var 1 has a significantly higher predicted yield compared to Var 3 and 4. The p-values are less than 0.001, which suggests that these differences are highly statistically significant. Var 2 vs. Var 4: Similarly, Var 2 has a significantly higher yield than Var 4, as shown by a significant positive estimate and a very low p-value. Var 2 vs. Var 3: The difference between Var 2 and Var 3 is also statistically significant but less pronounced than the differences involving Var 4.

Figure 22 shows how the predicted Yield changes with Nc for each level of Var.

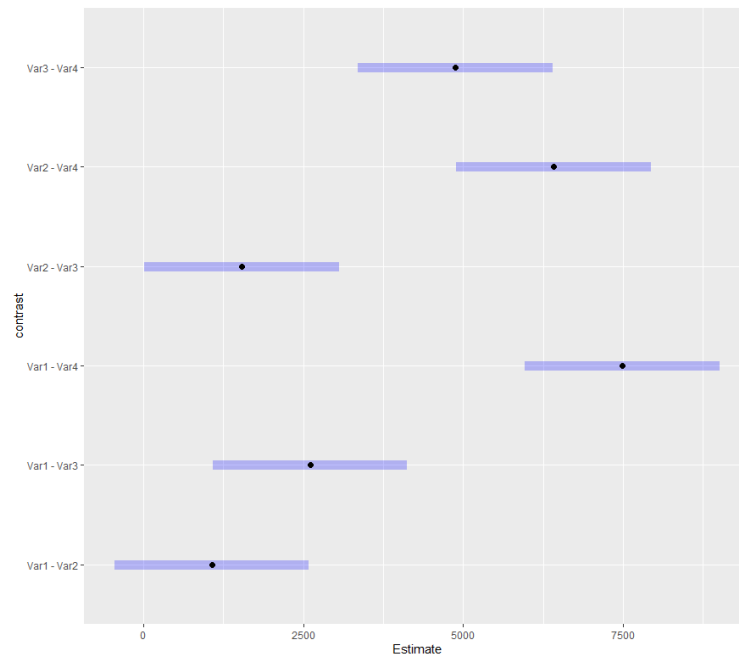


Figure 21

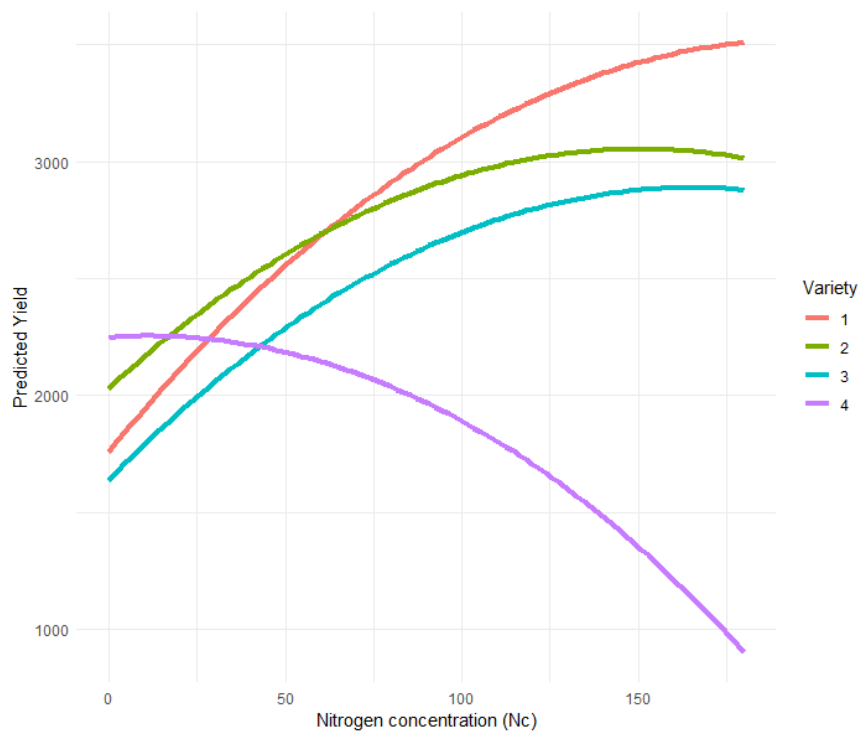


Figure 22

4 Conclusion

Each approach carries its unique strengths, with the categorical analysis offering clarity in group comparisons and the numerical analysis providing depth in trend exploration.

According to the results obtained in the Statistical Analysis when N is a categorical variable the pattern of results suggests that Var1 and Var2 have notably higher effects than Var3 and Var4, with Var1 having the highest mean effect of all. Var4 has the lowest mean effect according to these results. N1 appears to have a significantly negative impact on yield compared to the other nitrogen levels tested. So a good combination could be Var1 with N2. When N is a numerical variable, Yield predictions for all varieties initially increase with nitrogen concentration but at different rates. Var 1 and 2 appear to reach a plateau at higher yields. At the same time, Var 3 shows a slower increase, and Var 4 demonstrates a sharp decrease after a certain point, indicating a potential negative response to higher nitrogen levels. Here a good combination would be Var1 with N1 or N2.

5 Appendix A

```
1 #Libraries
2 library(ggplot2)
3 library(diagram)
4 library(emmeans)
5 library(lmerTest)
6 library(emmeans)
7 library(dplyr)
8 library(MASS)
9 library(car)
10 library(multcomp)
11
12 # Reading in data:
13 assignment3 <- read.table("assignment3.txt", header = TRUE, sep = "\t"
14   ")
15 str(assignment3)
16 head(assignment3)
17 #-----Convert N to Nc-----#
18 assignment3$Nc <- rep(rep(c(0,60,90,120,150,180), each = 4), 3)
19 #-----#
20
21 # Converting necessary variables to factors
22 assignment3$Block <- as.factor(assignment3$Block)
23 assignment3$N <- as.factor(assignment3$N)
24 assignment3$Var <- as.factor(assignment3$Var)
25
26 #-----Explorative-Data-And-Plots-----#
27
28 # Check if Factors are Balanced
29 assignment3 %>%
30   group_by(Block, N, Var) %>%
31   summarise(count = n()) %>%
32   ungroup() %>%
33   arrange(count)
34
35 #-----Checking the effect of the factors and the interactions
36
37 #For N:Categorical
38 interactionN1.lmer <- lmer(Yield ~ N + Var + N:Var + (1|Block) + (1|
39   Block:N), data = assignment3)
40 interactionN2.lmer <- lmer(Yield ~ N + Var + N:Var + (1|Block), data
41   = assignment3)
42 0.5*(1-pchisq(2*(logLik(interactionN1.lmer)-logLik(interactionN2.lmer
43   )),1))
```

```
41 interactionN.lm <- lm(Yield ~ N + Var + N:Var, data = assignment3)
42 0.5*(1-pchisq(2*(logLik(interactionN2.lmer)-logLik(interactionN.lm)),
43   1))
44 #For N:Numerical
45 interactionNc1.lmer <- lmer(Yield ~ Nc + Var + Nc:Var + (1|Block) + (
46   1|Block:Nc), data = assignment3)
47 interactionNc2.lmer <- lmer(Yield ~ Nc + I(Nc^2) + Var + Nc:Var + (1|
48   Block) + (1|Block:Nc), data = assignment3)
49 anova(interactionNc1.lmer, interactionNc2.lmer)
50 interactionNc1.lmer <- lmer(Yield ~ Nc + I(Nc^2) + Var + Nc:Var + (1|
51   Block) + (1|Block:Nc), data = assignment3)
52 interactionNc2.lmer <- lmer(Yield ~ Nc + I(Nc^2) + Var + Nc:Var + (1|
53   Block), data = assignment3)
54 0.5*(1-pchisq(2*(logLik(interactionNc1.lmer)-logLik(interactionNc2.
55   lmer)),1))
56 interactionNc.lm <- lm(Yield ~ Nc + I(Nc^2) + Var + Nc:Var, data =
57   assignment3)
58 0.5*(1-pchisq(2*(logLik(interactionNc2.lmer)-logLik(interactionNc.lm)
59   ),1))
60 #-----Diagramms-and-Plots-----#
61 par(mfrow=c(1,1))
62 # Boxplots for each categorical variable
63 for (var_name in c('Block', 'N', 'Var')) {
64   k <- length(levels(as.factor(assignment3[,var_name]))) + 1
65   boxplot(as.formula(paste("Yield ~", var_name)),
66     col = 2:k, main = paste("Yield vs.", var_name),
67     data = assignment3)
68 }
69 # Histogram for Yield
70 f <- function(x) {
71   dnorm(x, mean = mean(assignment3$Yield), sd = sd(assignment3$Yield)
72   )
73 }
74 hist(assignment3$Yield, xlab='Yield', probability=T, main = "
75   Distribution of Yield")
76 curve(f, from = min(assignment3$Yield), to = max(assignment3$Yield),
77   lwd=3, col="red", add=T)
78 par(mfrow=c(1,1))
```

```
76 # Interaction plots for different pairs of variables
77 with(assignment3, {
78     interaction.plot(N, Var, Yield, legend=FALSE,
79                     bty="n", col=2:6, xtick = TRUE,
80                     main = "Interaction Between N and Var")
81     interaction.plot(N, Block, Yield, legend=FALSE,
82                     bty="n", col=2:4, xtick = TRUE,
83                     main = "Interaction Between N and Block")
84     interaction.plot(Var, Block, Yield, legend=FALSE,
85                     bty="n", col=2:5, xtick = TRUE,
86                     main = "Interaction Between Var and Block")
87 })
88
89 par(mfrow=c(1,1))
90
91 #Factor Diagram
92 y.names <- c(expression("[I]" [36]^{72}),
93               expression("N:Var" [10]^{24}),
94               expression("[Block]" [12]^{18}),
95               expression("Var" [3]^{4}),
96               expression("N" [5]^{6}),
97               expression("[0]" [1]^{1}))
98 M <- matrix(nrow = 6, ncol = 6, byrow = TRUE, data = 0)
99 M[2,1] <- M[3,1] <- M[5,3] <- M[5,2] <- M[4,2] <- M[6,5] <- M[6,4] <-
100 " "
101 plotmat(M, pos = c(1,2,2,1), name = y.names, lwd = 2,
102         box.lwd = 2, cex.txt = 1, box.size = 0.12,
103         box.type = "square", box.prop = 0.2, arr.type = "triangle",
104         curve = 0)
105
106 #-----Models For N:categorical-----#
107 analysis.categ.lm <- lm(Yield ~ N + Var + N:Var, data = assignment3)
108 anova(analysis.categ.lm)
109
110 #-----Models For N:numerical-----#
111 #Fixed Model#
112 analysis.num.lm <- lm(Yield ~ Nc + Var + Nc:Var, data = assignment3)
113 anova(analysis.num.lm)
114
115 #-----Statistical analysis-----#
116 #-----N:Categorical-----#
```

```

119
120 model1 <- lm(Yield ~ N + Var + N:Var + Block, data = assignment3)
121
122 #-----Model Diagnostic-----#
123
124 par(mfrow=c(2,2))
125 plot(model1, which=1:4)
126 par(mfrow=c(1,1))
127 par(mfrow=c(1,1))
128 stdresid = rstandard(model1)
129 par(mfrow=c(1,1))
130 plot(stdresid ~ fitted(model1),
131       xlab = "Fitted Values",
132       ylab = "Standardized Residuals",
133       main = "Standardized Residuals vs Fitted Values")
134 with(assignment3, plot(stdresid ~ N,
135                         xlab = "N (Nitrogen Level)",
136                         ylab = "Standardized Residuals",
137                         main = "Standardized Residuals vs N"))
138 with(assignment3, plot(stdresid ~ Var,
139                         xlab = "Var (Variety)",
140                         ylab = "Standardized Residuals",
141                         main = "Standardized Residuals vs Var"))
142 par(mfrow=c(1,1))
143
144
145 # Box Cox
146 boxcox(model1)
147 par(mfrow = c(1, 1))
148
149 #-----Results-----#
150 model1 <- lmer(Yield ~ N + Var + N:Var + (1|Block), data = assignment
151               3)
152 summary(model1)
153 anova(model1)
154
155 #-----Post-hoc-----#
156
157 # Estimated mean levels and their differences for N
158 emm_N <- emmeans(model1, pairwise ~ N)
159 print(emm_N)
160 tukey_N <- cld(emm_N$emmeans, adjust = "tukey")
161 print(tukey_N)
162 plot(emm_N[[2]], xlab = 'Estimate', main = "Pairwise N Comparison")
163

```

```
164 # Estimated mean levels and their differences for Var
165 emm_Var <- emmeans(model1, pairwise ~ Var)
166 print(emm_Var)
167 tukey_Var <- cld(emm_Var$emmeans, adjust = "tukey")
168 print(tukey_Var)
169 plot(emm_Var[[2]], xlab = 'Estimate', main = "Pairwise Var Comparison
    ")
170
171
172 #-----N:Numerical
    -----#
173
174 model2 <- lm(Yield ~ Nc + I(Nc^2) + Var + Nc:Var + Block, data =
    assignment3)
175
176 #-----Model Diagnostic-----#
177 #1st Model Diagnostics
178 par(mfrow=c(1,1))
179 plot(model2, which=1:4)
180 par(mfrow=c(1,1))
181 plot(as.numeric(assignment3$Nc), rstandard(model2),
182       xlab = 'Nc (Nitrogen)', ylab = 'Standardized residuals',
183       main = 'Residuals vs. Nitrogen Level')
184 plot(predict(model2), rstandard(model2),
185       xlab = 'Predicted values', ylab = 'Standardized residuals',
186       main = 'Residuals vs. Predicted Values')
187 plot(as.numeric(assignment3$Var), rstandard(model2),
188       xlab = 'Var (Variety)', ylab = 'Standardized residuals',
189       main = 'Residuals vs. Variety')
190 par(mfrow=c(1,1))
191 studresid = studres(model2)
192 par(mfrow=c(1,1))
193 plot(studresid ~ fitted(model2),
194       xlab = "Fitted Values",
195       ylab = "Studentized Residuals",
196       main = "Studentized Residuals vs Fitted Values")
197 with(assignment3, plot(studresid ~ Nc,
198                       xlab = "Nc (Nitrogen)",
199                       ylab = "Studentized Residuals",
200                       main = "Studentized Residuals vs Nc"))
201 with(assignment3, plot(studresid ~ Var,
202                       xlab = "Var (Variety)",
203                       ylab = "Studentized Residuals",
204                       main = "Studentized Residuals vs Var"))
205 par(mfrow=c(1,1))
206
```

```

207 # Box Cox
208 boxcox(model2)
209 par(mfrow = c(1, 1))
210
211 #-----Results-----#
212
213 model2 <- lmer(Yield ~ Nc + I(Nc^2) + Var + Nc:Var + (1|Block), data
    = assignment3)
214 summary(model2)
215 ranova(model2)
216 anova(model2)
217
218 #-----Post-hoc-----#
219
220 emm_Var <- emmeans(model2, pairwise ~ Var)
221 print(emm_Var)
222 plot(emm_Var[[2]], xlab = 'Estimate')
223
224 # Prediction across a range of 'Nc' values for each 'Var'
225 vars <- levels(assignment3$Var)
226 nc_vals <- seq(from = min(assignment3$Nc), to = max(assignment3$Nc),
    length.out = 100)
227 new_data <- expand.grid(Nc = nc_vals, Var = vars, Block = factor(
    levels(assignment3$Block)))
228 # Predict Yield
229 new_data$Yield_pred <- predict(model2, newdata = new_data, re.form =
    NA)
230 new_data$Yield_pred <- new_data$Yield_pred
231 # Plot the predictions
232 ggplot(new_data, aes(x = Nc, y = Yield_pred, color = Var)) +
233   geom_line(size = 1.5) +
234   labs(x = "Nitrogen concentration (Nc)", y = "Predicted Yield",
    color = "Variety") +
235   theme_minimal()

```