University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Biometry

THE DETERMINATION OF VARIABILITY IN TYPE II AND III EXPERIMENTS OF IMPROVED YIELDS

Research Report in Biometry, Number . . . 2021

David Mugi Waigwa

July 2021

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Master Project in Biometry **University of Nairobi**

July 2021

THE DETERMINATION OF VARIABILITY IN TYPE II AND III EXPERIMENTS OF IMPROVED YIELDS

Research Report in Biometry, Number . . . 2021

David Mugi Waigwa

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

The goal is to determine the variability of yields in Type II and III agricultural experiments involving improved yields in Murangiri, Kenya, while determining an appropriate model that explains this variability. In this study, we investigated the effects of different treatments ranging from organic, inorganic and mixture of the two on the maize yields in Murangiri, Tharaka Nithi constituency in Kenya. We applied three statistical models to the data obtained from Type II and Type III experiments namely;FIXED EFFECTS MODEL, GENERALIZED LINEAR MODEL (GLM) and MIXED EFFECTS MODEL. We focused on interpretations and computation of model parameters and also investigated which model best fits the two datasets from the two experiments. Our study found that the treatments in general had the effects on the Maize yields in the two experiments as shown by all models fitted since the p-values of both mixed and fixed effect model are less than level of significance 0.05 while for GLM by using the deviance we showed that the fitted model with treatments were significant on both cases. On the best model, we used the model comparisons Akaike's Information Criterion (AIC) to determine model that best fit the two datasets from the Type II and Type III experiments respectively. The study found that the mixed model was the best among the three models considered under this study as it was having the smallest values of AIC 169.071 and 280.01 for Type II and Type III experiments as indicated in tables 5 and 9 respectively. Despite the mixed model showing the smallest AIC value among the three models, the differences among these values were not very significant, implying that all three models could be used to explain the variability of yields in Type II and III agricultural experiments.

Master Thesis in Biometry at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Biometry

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

19/07/2021

Signature Date

DAVID MUGI WAIGWA
Reg No. I56/7875/2002

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

20/07/2021

Signature Date

Dr. Vincent Onguso Oeba
Kenya Forestry Research Institute (KEFRI)
Box 20412, 00200.
E-mail: voeba@kefri.org

Dedication

This project is dedicated to my family and parents who have always believed that I would one day complete my studies.

Contents

Abstract ii

Declaration and Approval iv

Dedication vi

List of Figures ix

List of Tables x

Acknowledgments xi

1 INTRODUCTION 1

- 1.1 **RATIONALE** 1
- 1.2 **BACKGROUND** 2
- 1.3 **PROBLEM STATEMENT** 3
- 1.4 **RESEARCH OBJECTIVES** 4
  - 1.4.1 OVERALL OBJECTIVES 4
  - 1.4.2 SPECIFIC OBJECTIVES 4
- 1.5 **SIGNIFICANCE OF THE STUDY** 5

2 LITERATURE REVIEW 6

3 RESEARCH METHODOLOGY 9

- 3.1 RESEARCH DESIGN 9
  - 3.1.1 ***Data source and its description*** 9
  - 3.1.2 ***Study variables*** 9
  - 3.1.3 ***Exploratory data analysis*** 9
- 3.2 STATISTICAL MODELS 10
  - 3.2.1 ***Fixed effects model*** 10
  - 3.2.2 ***Generalized Linear Model*** 10
  - 3.2.3 ***Mixed Effect Model*** 11
- 3.3 MODEL COMPARISON TECHNIQUES 13
- 3.4 MODEL CHECKING TECHNIQUES FOR MIXED EFFECT MODEL 13
- 3.5 DIAGNOSTIC TESTS 14
  - 3.5.1 Shapiro-Wilk Test 14
  - 3.5.2 Normality plots. 14
  - 3.5.3 Ethical Considerations. 14

4 DATA ANALYSIS AND RESULTS 15

- 4.1 Introduction 15
- 4.2 Exploratory Data Analysis 15
  - 4.2.1 Descriptive statistics of field data 15
- 4.3 Diagnostics test of Normality 16

4.3.1 SHAPIRO –WILK NORMALITY TEST 17
4.4 Inferential Statistics 17
4.4.1 Fixed effect model 17
Plot of residuals & fitted values and QQ norm for the fixed effect model 18
4.4.2 Generalized Linear model (GLM) 19
4.4.3 Mixed Model 20
4.4.4 Model comparisons 21
4.5 Statistical models fitted Type III experiment data 21
4.5.1 Fixed effect model 21
Plot of residuals & fitted values and QQ norm for the fixed effect model 22
4.5.2 Generalized Linear model (GLM) 23
4.5.3 Mixed Model 24
4.5.4 Model comparisons 25
5 CONCLUSIONS AND RECOMMENDATION 26
5.1 Introduction 26
5.1.1 Interpretations of results 26
5.1.2 Recommendations 27
References 28
APPENDIX A 30

List of Figures

Figure 1. Strip chart for Type II and Type III data. 15
Figure 2. Box Plot for Type II and Type III data.......... 16
Figure 3. Normality plot.......... 16
Figure 4. Plot of residuals & fitted values and QQ norm for the fixed effect model 18
Figure 5. Plot of residuals & fitted values and QQ norm for the fixed effect model 22

List of Tables

Table 1. Shapiro-wilk normality test 17
Table 2. Fixed effect model 17
Table 3. Results for GLM Type II experiment 19
Table 4. Results for Mixed model 20
Table 5. Model Comparison Results 21
Table 6. Fixed effect model – Type III 21
Table 7. Results from generalized linear model (GLM) assuming the Gaussian family 23
Table 8. Results for Mixed model 24
Table 9. Model Comparison Results 25

Acknowledgments

I would like to thank God for granting me the grace to finish my project in a timely manner. I pass my sincere gratitude towards my supervisor Dr. Vincent Oeba for his guidance and support towards completion of the project. I am also thankful to my family Julia, Jonah, Margaret, and Kellos for their unwavering support, even during the times I would be away so that I could study and do the project instead of spending time with them.

My appreciation to the University of Nairobi, and especially to Prof. Stephen Luketero, Head of Mathematics Department, and Prof. Lydia Njenga, Director, School of Graduate Studies for not only believing in me, but offering me a second chance to complete my studies after being absent for such a long time.

Of notable mention is Dr. Daniel Njai a senior lecturer from the University of Nairobi and a family friend who was key in my success by always making a follow up and asking about my progression in studies. He not only offered words of encouragement, but also advice filled with lots of wisdom.

Lastly, I wish to acknowledge the moral and spiritual support of my parents Mr. John Waigwa and Mrs. Margaret Waigwa who were instrumental in my starting the Masters course in 2002, and always encouraging me to push on and finish the program.

David Mugi Waigwa

Nairobi, 2021.

1 INTRODUCTION

1.1 RATIONALE

In agricultural research, there are three types of experiments Type I, II, and III that can be conducted to improve the yield production while minimizing costs. A Type I experiment is one that is designed and managed by the researcher. A Type II experiment is designed by the researcher and managed by the farmer. The researcher looks for the farmer contribution in the designed experiments. The farmer provides the land, and the researcher monitors the treatments that were successful at the on-station level on the farmer's field. The farmer is provided instructions and researcher monitors farmer contribution. This provides some aspect of variation that is not explained by the treatment effect. A Type III experiment is designed and managed by the farmer.

The farmers go to the on-station and learn various methodologies of interest. They then utilize their knowledge to replicate some of the experiments that they found useful. It is expected that the Type III experiment will have the most variability when a comparison is done with the other experimental units of Type I and II. This is because there is no design structure on the farms under Type III when compared to those on Type I and II that are designed by the researcher. The on-station designs would likely have less variability due to the environment, treatments used and effective monitoring systems. The purpose of this study is to determine the variability in Type II and III experiments of improved yields.

1.2 BACKGROUND

The study was conducted in Tharaka Nithi county, Tharaka constituency. The primary economic activities of this county are subsistence dairy farming, rearing of goats and sheep, tea farming, and coffee planting. Farmers from this study used different treatments on their maize crops. The treatments were organic, inorganic, or a mixture of both. Organic farming entails the cultivating plants or the rearing of animals using natural methods. It involves using cover crops, manure (plant and animal), crop rotation, and other techniques to control weeds, pests, and diseases. Inorganic farming entails the use of synthetic fertilizers and pesticides to improve yield production (Kakar et al., 2020 [13]).

The organic treatments used on the farms were Mucuna pruriens, Tithonia diversifolia, Calliandra haematocephala, Crotalaria retusa, Leucaena leucocephala and Manure. Mucuna pruriens is used as a green manure because it introduces nitrogen to the soil making it more fertile. Tithonia diversifolia decomposes rapidly as releases nitrogen, phosphorous and potassium into the soil improving its fertility and increasing yield production (Ajao & Moteetee, 2017 [3]). Calliandra haematocephala is an important cover crop as it provides nitrogen to the soil. Crotalaria retusa also introduces nitrogen into the soil as has been shown to improve soil fertility.

Leucaena leucocephala is a good cover crop and also adds nitrogen into the soil. The manure used by most farmers came from cows, goats and sheep. According to Moyin-Jesu, 2012 [16], Liquid Cattle Manure (LCM) is very useful as a treatment as it contains nitrogen that can be used to enhance plant growth and improve yield production. LCM can also improve the soil salinity as well as increasing micro nutrients in it. The inorganic treatment used was NPK (nitrogen, phosphorous, and potassium) based fertilizer. Nitrogen improves the leaf structure enabling the plants to produce better yield. Phosphorous assists in seed germination and root development, while potassium is vital for maintaining growth while enabling the plant become disease resistant.

1.3 PROBLEM STATEMENT

A lot has been done on organic and inorganic fertilizers and their effect in improving yields. More has been done on different agricultural experimental designs, and discussions generated about the integrated soil fertility management (Dafallah, 2017 [9] & Bailey-Serres et.al., 2019 [4]). Some of these studies focus on treatments used in the on-farm and on-station experiments and their effect on yield production, but do not discuss about the variability of these experiments (Acer et.al, 2004; [1]; Moyin-Jesu, 2012 [16]). Not much has been done on determining the variability between the Type II and Type III experiments.

In controlled experiments such as the on-station researcher designed and managed, it is possible to have experimental plots receiving the same treatment but exhibiting different outcomes showing the existence of variability. It is also expected that experiments that are farmer designed and managed at the on-farm level will show more variability than those at the on-station level managed by the researcher. The variability may be attributed to environmental factors such as weather, climate, pests, soils and topography. Gupta et.al (2015) [11] posited that variability in agricultural designs may be wanted and desirable, or unwanted and undesirable implying the diversity of these experiments and the need to conduct more research. Proper experimentation designs ensure the identification and isolation of natural variation so that true effects due to treatments can be measured with some degree of accuracy.

To minimize field variability, blocking is done on the fields so that the slope and soil characteristics of the land are taken into consideration. There are several methods that are used to measure the variability in agricultural designs. In this study we will discuss the variability using Analysis of Variance (ANOVA), Generalized Linear Models (GLM) and Mixed Modelling methods and determine which best model to use for Type II and III experiments.

1.4 RESEARCH OBJECTIVES

1.4.1 OVERALL OBJECTIVES

To determine the variability in type II and III experiments involving improved yields in Murangiri, Kenya.

1.4.2 SPECIFIC OBJECTIVES

1. To determine the best model that explains the variability of yields in type II and III experiments involving improved yields in Murangiri, Kenya.
2. To determine the variability of yields in type II and III experiments involving improved yields in Murangiri, Kenya.

1.5 SIGNIFICANCE OF THE STUDY

Our study reveals the relevance of statistics to answer specific questions in the can be controlled as the researcher can design the on-station farms, observe the environment and control the treatments used According to **Bailey-Serres et.al (2019) [4]**, future food security will heavily rely on trans-formative methods used to improve yield production, due to environmental factors caused by climate change. As the world population increases, it is imperative that we utilize our resources accordingly, while ensuring that there will not be any shortage of food at any time. Heat waves, the continual increase in droughts, torrential rainfall and other extreme weather patterns that are seen across the world have a dismal effect on agricultural production, and are predicted to continue doing so in future **Bailey-Serres et.al (2019) [4]**. This sensitizes our need to look at ways of improving food production and yields for a sustainable future.

One of the sustainable development goals created through the United Nations Development Program (UNDP) focuses to end hunger, achieve food security and improved nutrition, and promote sustainable agriculture (The 2030 Agenda and the Sustainable Development Goals: An Opportunity for Latin America and the Caribbean, 2018). While not all variability such as the experimental error can be explained, the study will explain some of the variability found in the Type II and III experimental methods and possibly look for patterns on the improved yields to be used for recommendations to future researchers and farmers.

Poverty is the leading cause of death in many developing countries due to fluctuations in the availability of food **(Bationo et al., 2011 [19])**. To slow down this food insecurity problems in the developing countries while empowering their communities to be self-sufficient, agricultural researchers need to step up their continual efforts to improve food security by developing sustainable programs that can work for those communities. Poverty can be decreased by ensuring agricultural production is high so that majority of these countries facing food security issues can be self-sustained.

2 LITERATURE REVIEW

(V.K. Gupta et al., 2015 [11]) performed a study on the significance of experimental designs in agricultural research. The researchers discussed the importance of factoring in variability when designing experiments in agricultural research. The purpose of the study was to explain some modern efficient and useful designs used by agricultural researchers when conducting their experiments. The designs used were the Incomplete, Resolvable, and Augmented block designs, and block design with factorial structure. The incomplete block design is useful for single factor experiments while the augmented designs are used for making comparisons between the control and the treatments. (Horsley, n.d. [12]). The resolvable block designs are structured such that the experiments can be performed one replication at a time. This allows to control and monitor the variation caused by location and time periods. The block designs with factorial structure are useful when the experiments have structured treatments. The researchers used ANOVA technique to monitor the variability and tested whether all the treatments were the same or whether at least one treatment mean was different. Despite the extensive study done on agricultural experimental designs, the researchers did not mention the variation on Type II and III experiments implying that more work was needed on this area.

Farjana et al., (2019) [10] conducted a study on cabbage *Brassica oleracea* to monitor its growth and yield when both organic and inorganic treatments were used. The experiment consisted of four varied types of fertilizers with the control being considered as one of the fertilizers. The other fertilizers were organic, inorganic and a mixture of both. Three different types of mulching were used in this experiment namely water hyacinth, rice straw, and black polythene. The researchers used a Randomized Complete Block Design with three replications. They found out that there were significant differences among the treatments used. The study stated that using a combination of both fertilizers with black polythene had the highest yield production of cabbages. A limitation of the study was lack of an agricultural experimentation between the organic and inorganic fertilizers to determine the best performance.

Aina et al., (2019) [2] conducted a study that took a comparison between the levels of bioactive compounds and antioxidant characteristics of produce that were planted using varied soil compositions. The soils were treated with cow dung, chicken droppings, and NPK based fertilizer. Tomato seedlings were planted and watered for five months. The soils treated with NPK showed the highest yield performance, followed by chicken droppings, and then cow dung. Despite the inorganic fertilizer showing the highest yield, the researchers reported that more use of organic manures that are found on farms should be used to enhance the quality of soil nutrition and that it will reduce environmental degradation. They also recommended that extensive and detailed study such as agricultural experimental procedures of research designed and farmer managed (Type II), or designed and managed by the farmer (Type III) need to be considered for future research.

Sharada & Sujathamma, (2018) [18] conducted a study on how both organic and inorganic fertilizers can affect rice production. The researchers posited that using inorganic fertilizers caused water to be contaminated and reduced soil productivity levels which in turn affect rice production. They also stated that using organic manures assisted the soil to regain its health but was insufficient to provide the necessary nutrients for optimal growth. They recommended usage of a mixture of organic and inorganic fertilizers to produce rice. The experiment was conducted on-farm using the randomized complete block design with three replications. The treatment had two controls with ten different organic and inorganic fertilizers. There were higher yields for the inorganic treatments compared to the organic ones. The researchers advised that there was need to conduct further research to compare the effect of using both organic and inorganic fertilizers on crop produce.

Waqas et al. (2020) [21] did a study on the nature of managing nutrients on crop production, and yield stability under different climates. The study was long-term (more than 10 years) and was conducted using a meta-analysis approach. The area under investigation was partitioned into four groups with the basis on the local climatic conditions which were warm dry and moist, and cool dry and moist in China. The results showed that the effect of managing soil nutrient on soil carbon storage and yield stability varies under the different zones depending on the climate. The study also investigated how using unbalanced and balanced mineral fertilizer led to a depletion of the soil carbon storage by 6% and 11% respectively. It also showed that using balanced mineral fertilizer was the most appropriate nutrient management strategy.

Voltr et al. (2021) [20] conducted a study on the organic matter found in soil and its properties. The researchers posited that despite the current studies on the soil's condition compounded with its management methodologies, not much had been done on the determination of a common model for both. The study focused on the use of fodder, inorganic and organic fertilizers, harvest residues as soil management practices. The study used one and multidimensional linear regression for analysis and was conducted in the Czech Republic between 2008 and 2018. The results showed an increase in the soil organic carbon storage while the hot water ex-tractable carbon was decreased. These two fundamental factors influence the soil productivity and health. The multivariate linear regression showed that the hot water ex-tractable carbon content was significantly affected by the type of soil, phosphorous content, and the nitrogen content in the soil. The results also showed that soil organic matter plays an integral part in maintaining the overall biochemical and biological properties that enhances proper soil health (agroecosystems) and ensures food security.

Ismail (2018) [22] carried out a study on the application of Generalized Linear Models (GLM) in agriculture. In GLM, the Akaike Information Criterion (AIC) is used for choosing an adequate model. The study fitted both classical as well as GLM's on a dataset that determined yield and yield attributes of tulip (Tulipa Liliodeae). It was observed that the parameter estimates in these models can be greatly influenced by the error assumptions, and that GLM's can be adequately used to identify appropriate error structure in the modeling production of tulips.

(Bell et al., 2019) [6] did a study on making informed choices when using the fixed and random effects model. The paper aim was to assist researchers analyzing longitudinal and multilevel data make good and sound decisions when choosing the right models. They discussed the capabilities and limitations of the fixed and random effects models. Their simulations found out that failing to add random slopes can generate standard errors. They also discussed that assuming random intercepts are normally distributed when they are not introduces bias in the analysis.

(Clarke et al., 2018) [8] performed a study on the choice between using fixed or random effects model. The did the study with consideration to educational research and set out some of the assumptions of the two methods. After performing an analysis on the determinants of student achievement in primary schools, they concluded that the fixed effect approach was preferable when looking at individual characteristics within the school. They also posited that the random effect model was suitable when researching from a wider population.

(Lord et al., 2008) [14] conducted a study on the Conway-Mazwell-Poisson Generalized Linear Model to determine the application for analyzing car accidents. Apart from using the GLM model for the analysis, the researchers also compared the results with the Negative Binomial model. To test the objectives, several models were created using the two approaches. The results showed that Generalized Linear Models performed as well as the Negative Binomial model when they compared the goodness of fit statistics and the predictive strength of the models. They concluded that the GLM offered a better model than the Negative Binomial for modeling the car accidents due to the limitations of the Negative Binomial model.

3 RESEARCH METHODOLOGY

3.1 RESEARCH DESIGN

3.1.1 *Data source and its description*

The study used routine data collected from the farmers in Tharaka Nithi County. The data categorized into type II and Type III depending on whether it was managed and designed by researcher or farmer. The data was collected and recorded with maize yields being measured in tonnes per individual plot assigned different treatments. Data in excel was cleaned and analysis was done in R software.

3.1.2 *Study variables*

The variables were classified as; dependent variables which consisted maize yields and independent variable which consisted the treatments applied to each plot including organic and inorganic.

3.1.3 *Exploratory data analysis*

This is the technique used to visualize the patterns of data relative to research interests. This implies that EDA can help us to discover more information regarding raw data obtained from the plot by plotting individual curves and carrying out descriptive statistics to examine the data before doing any inferential statistics such as model fitting.

3.2 STATISTICAL MODELS

We apply three statistical models namely; fixed effects model, generalized linear model (GLM) and mixed models to fit the type II and Type III data and compare these three models to determine which best fit the data.

3.2.1 *Fixed effects model*

The fixed effects model is given by:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \left\{ \begin{matrix} i=1,2,\ldots m \\ j=1,2,\ldots n \end{matrix} \right\} \tag{1}$$

Where,
μ_i is parameter common to all treatments and is known as overall mean,
τ_i is the effect of the i^{th} treatments and ε_{ij} is random error component that incorporates all other sources of variability in the experiment including variability arising from uncontrolled factors, differences between the experimental units to which the treatments are applied, and the general background noise in the process.

3.2.1.1 Assumptions of the fixed effect model.

i. The error term ε_{ij} has conditional mean of zero.

ii. The variables are independently and identically distributed.

iii. Large outliers are unlikely.

iv. There is no perfect multi-collinearity.

3.2.2 *Generalized Linear Model*

McCullagh and Nelder (1989) [15] came up with the idea of Generalized Linear Models (GLM). Let us consider n observations and let $Y_i(x_i)$ denote a continuous response. The classic linear model is given by;

$$Y_i = \chi_i^T \beta + \varepsilon_i, \varepsilon_i \sim N\left(0, \sigma^2\right), 1 \leq i \leq n \tag{2}$$

Where,
$N\left(0, \sigma^2\right)$ denotes a normal distribution with mean μ and σ^2 variance. One major limitation is that it only applies to continuous response Y_i .The generalized linear models (GLM) extend the classic linear model to non-continuous response such as binary. To express the GLM, we first rewrite the

linear regression in (1) as:

$$Y_i|X_i \sim N\left(0,\sigma^2\right), \mu_i = E(Y_i|X_i) = \chi_i^T \beta, 1 \leq i \leq n \quad \textbf{(3)}$$

Where,
$Y_i|X_i$ denotes the conditional distribution of Y_i and X_i and $E(Y_i|X_i)$ denotes the conditional mean of Y_i given X_i . By replacing the normal in equation (3) with other distributions appropriate for the type of response, we obtain the class GLM.

$$Y_i|X_i \sim f\left(\mu_i\right), g\left(\mu_i\right), \mu_i = E(Y_i|X_i) = \chi_i^T \beta, 1 \leq i \leq n \quad \textbf{(4)}$$

Where $f\left(\mu_i\right)$ denotes some distribution with mean μ and $g\left(\mu\right)$ is a function of μ. Since $g\left(\mu\right)$ links the mean to the explanatory variables, $g\left(\mu\right)$ is called the link function. μ, the pdf of Y, is the random component of the GLM. $\chi_i^T \beta$ is the systematic component. The specification of $f\left(\mu_i\right)$ and $g\left(\mu\right)$ depends on the type of response Y_i .In our case we assume Y_i Gaussian.

Inference for GLM can be based on Maximum Likelihood (ML) or estimates equations (EE). The classic ML provides most efficient estimates, if the response Y_i follows the specified distribution such as normal in the linear regression (1). In our studies we limit estimating of parameters by GLM to ML.

3.2.2.1 Assumptions of the Generalized Linear Model

a. The data values $Y_1, Y_2, ..., Y_n$ are independently distributed.

b. The response variable Y_i does not have to be normally distributed. It however assumes a distribution that comes from an exponential family.

c. Homogeneity of the variance does not need to be fulfilled.

d. Error terms do not need to be normally distributed but have to be independent.

e. Since it used the ML estimation instead of the Ordinary Least Squares to estimate the parameters, it has to depend on large sample approximations.

3.2.3 *Mixed Effect Model*

A mixed model comprises of a fixed effect and a random effects component. It is commonly used for modeling continuous response variables where data are collected longitudinally and have some dependency structure between observations. The random component has blocking and error components. A fixed effect is an unknown constant that we try to estimate from given data. Since a random effect is a random variable it is estimated by using the parameters that describe the distribution of the random effect. If predictor variables are selected for analysis, they are considered fixed. However, if they are selected randomly from a large population of other predictor variables, they are considered random.

A fixed effects model with normal errors can be represented as:

$$Y = X\beta + \varepsilon, orY \sim N\left(X\beta, \sigma^2 I\right)$$

Where, X is an nxp matrix and β is a vector of length p.
This can be generalized to mixed effects model with a vector y of q random effects with corresponding matrix Z which is of dimension nxp.
The mixed model can be written as;

$$Y = X\beta + Z\gamma + \varepsilon, orY \sim N\left(X\beta + Z\gamma, \sigma^2 I\right)$$

Where,
$X\beta$ represents the fixed component, where β is a vector of parameters associated with the fixed factors, $Z\gamma$ the random part with γ the vector of parameters associated with the random effects, and ε the unknown error term.

The parameters are estimated using the Maximum Likelihood Estimator (MLE). Maximum likelihood estimation seeks to find parameter values that make the mode's predicted values closest in match to the observed values (Baayen et al., 2008) [5].

3.2.3.1 Assumptions of the Mixed Effects Model

a. The within-group errors are independent with mean zero and variance σ^2.

b. The within group errors are independent of the random effects.

c. The random effects are independent and normally distributed.

3.3 MODEL COMPARISON TECHNIQUES

Akaike's Information Criterion (AIC) is used which is a measure of goodness of fit of an estimated statistical model. It is a tool for model selection. The AIC penalizes the likelihood by the number of covariance parameters in the model, therefore;
$AIC = -2\log(L) + 2p$
A lower AIC value shows a better fitting model. The usefulness of the AIC is that it can assist researchers determine the model that explains the variation in their data values.

3.4 MODEL CHECKING TECHNIQUES FOR MIXED EFFECT MODEL

In Mixed effects model, it is assumed that random effects are normally distributed and uncorrelated with error term. Residual plots can be used visually to check normality of these effects and to identify any outlying effect categories. Examining the plot of the standardized residuals versus fitted values by any covariates of interest can give a better feeling Verbeke and Molenberghs, 2009) [17].

3.5 DIAGNOSTIC TESTS

We carryout Normality test to determine whether the data are normally distributed or not, the study will employ Shapiro-wilk test (Ghasemi & Zahedias, 2012) and also, the Normality plot.

3.5.1 Shapiro-Wilk Test

This test examines whether the null hypothesis that a sample $x_1, x_2, ..., x_n$ comes from a normal distribution. The test statistic can be written as;

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{[i]}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Where,
$x_{[i]}$ is the i^{th} - smallest number in the sample, and $\bar{x}$ is the sample mean.
The coefficients a_i can be written as;

$$(a_1, a_2, ..., a_n) = \frac{\left(m^T V^{-1}\right)}{C}$$

Where, C is a vector norm function defined as;

$$C = ||V^{-1}m|| = \left(m^T V^{-1} V^{-1} m\right)^{1/2}$$

while the vector $m = (m_1, m_2, ..., m_n)^T$ consists of expected values of the order statistics of identically distributed random variables sampled from the standard normal distribution.
V represents the covariance matrix of the normal order statistics.

3.5.2 Normality plots.

These are plots used to assess whether or not a dataset comes from a normal distribution. It assists in the identification of outliers, skewness, and kurtosis. They consist of raw data, residuals from model fits, and estimated parameters.

3.5.3 Ethical Considerations.

A permission to undertake the study has been obtained from ethical clearance at University of Nairobi Ethics Review Committee and The Tharaka Nithi Ethics Committee

4 DATA ANALYSIS AND RESULTS

4.1 Introduction

This chapter describes the findings and discussion of results. We employed statistical techniques both descriptive and inferential statistics to analyze the data from the Type II and Type III experiments. Exploratory data analysis was used to help us visualize the data through the strip chart and the boxplot. Quantile-Quantile plots were also employed to check normality of the data and the residuals. The inferential statistics were achieved through fitting the three models to each dataset and the testing the hypotheses.

4.2 Exploratory Data Analysis

4.2.1 Descriptive statistics of field data

The strip chart is used to present Maize yields produced by different treatments Type II and Type III experiments as shown in Figure 1 below.

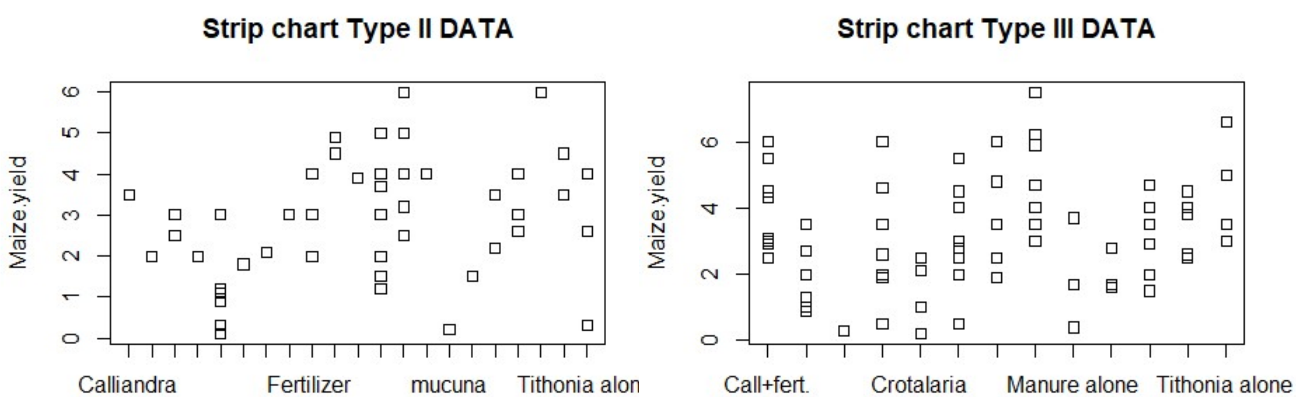

Figure 1. Strip chart for Type II and Type III data.

Exploratory Data Analysis (EDA) used the boxplot to check the distribution of the maize yields from different treatments as shown in figure 2 below.

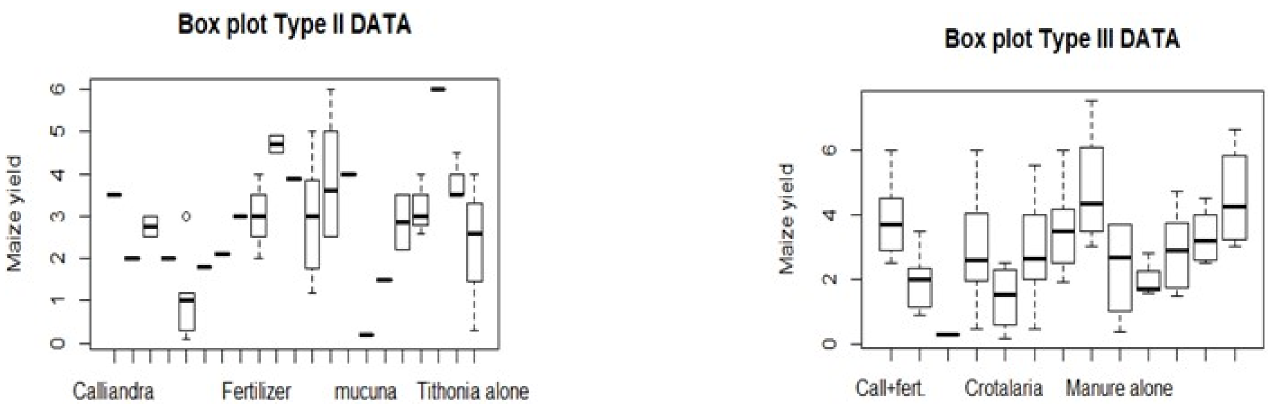

Figure 2. Box Plot for Type II and Type III data.

4.3 Diagnostics test of Normality

We carry out normality tests to determine whether the data are normally distributed or not, the study will employ Shapiro-wilk test (Ghasemi & Zahedias, 2012) and the Normality plot as shown in figure 3 below.

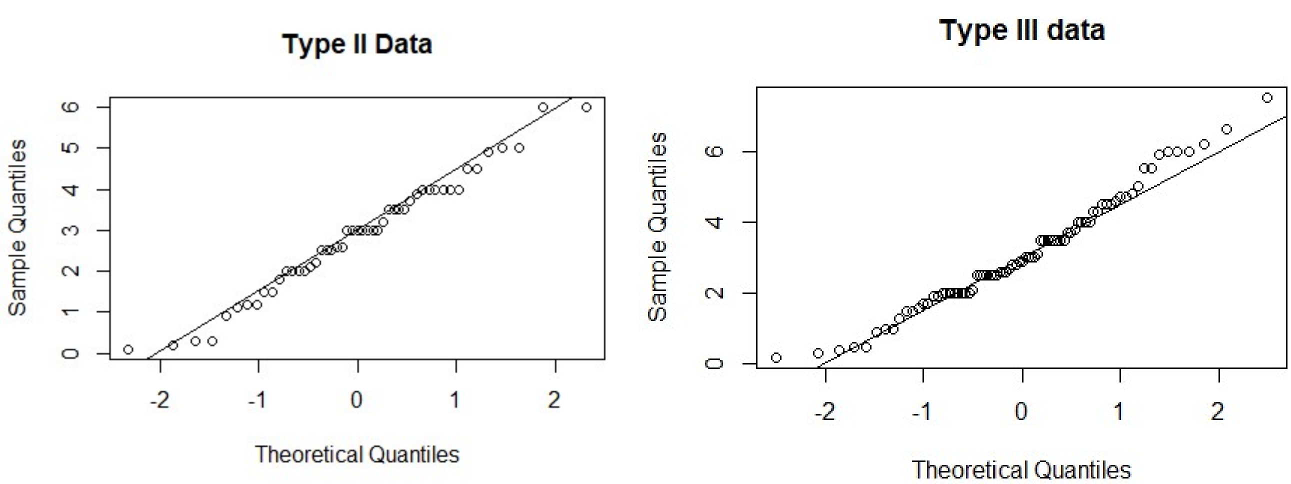

Figure 3. Normality plot.

From Figure 3 above, we can conclude the data is approximately normally distributed

4.3.1 SHAPIRO –WILK NORMALITY TEST

Data type	w-value	p-value
Type II experiment	0.98051	0.5868
Type III experiment	0.97608	0.1438

Table 1. Shapiro-wilk normality test

At level of significance 5%, we can conclude that the data is normal since *p*-values are greater than 5% hence confirmatory from figure 3.

4.4 Inferential Statistics

In this section we apply the statistical models discussed in chapter three to data from Type II and Type III experiments and plot the residuals for the model fitted.

4.4.1 Fixed effect model

We begin by fitting the fixed effect model to data from type II experiment so that to investigate the hypothesis that the fixed effect model significantly fit the data from the Type II experiment as shown table 2 below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	20	60.43	3.021	2.168	0.0294 *
Residuals	28	39.03	1.394		
Signif. Codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

Table 2. Fixed effect model

From the above table, we can conclude the model is significant at 0.05 since the *p-value* is $0.0294 < 0.05$

Plot of residuals & fitted values and QQ norm for the fixed effect model

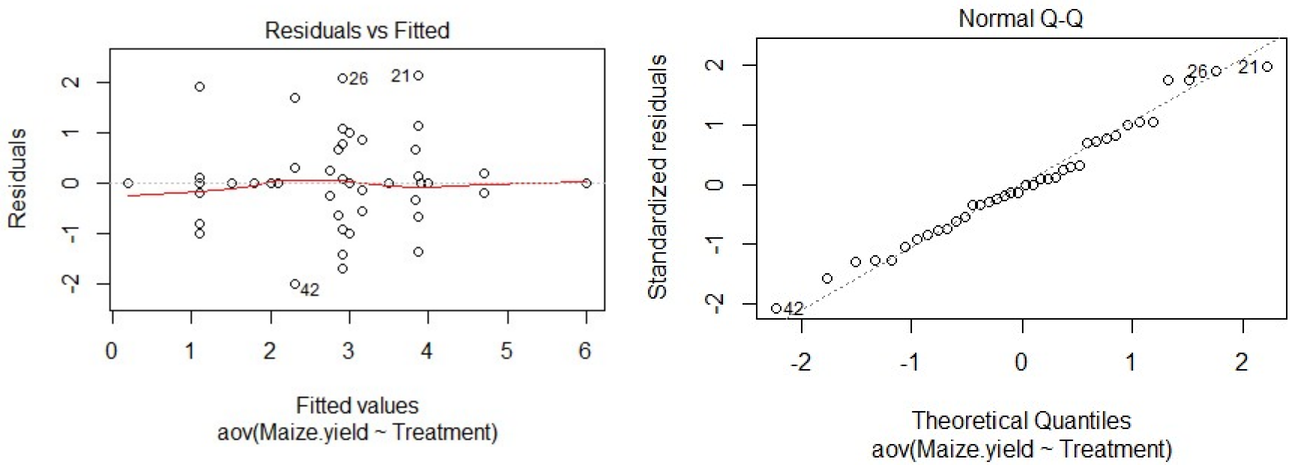

Figure 4. Plot of residuals & fitted values and QQ norm for the fixed effect model

The residuals vs Fitted values show that the error terms are scattered with no obvious pattern. This implies that the error terms are normally distributed.
The Quantile-quantile plot shows the values clustered around the line, further confirming that the residuals are normally distributed.

4.4.2 Generalized Linear model (GLM)

We fit the GLM to data from Type II experiment and the results obtained are presented in the table below.

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	3.5000	1.1806	2.965	0.00613 **
TreatmentCalliandra + fertilizer	-1.5000	1.6697	-0.898	0.37664
TreatmentCalliandra + manure	-0.7500	1.4460	-0.519	0.60806
TreatmentCalliandra + tithonia + fart -1.5000	1.6697	-0.898	0.37664	
Treatmentcontrol	-2.4000	1.2752	-1.882	0.07026 .
TreatmentCrotalaria _+ leucaena	-1.7000	1.6697	-1.018	0.31731
TreatmentCrotalaria + fertilizer	-1.4000	1.6697	-0.838	0.40885
TreatmentCrotalaria + manure + fertil -0.5000	1.6697	-0.299	0.76680	
TreatmentFertilizer	-0.5000	1.3633	-0.367	0.71655
TreatmentLeucaena+ manure	1.2000	1.4460	0.830	0.41362
TreatmentManure + calliandra	0.4000	1.6697	0.240	0.81241
Treatmentmanure + fertilizer	-0.5857	1.2621	-0.464	0.64619
TreatmentManure alone	0.3667	1.2752	0.288	0.77582
TreatmentMuc + Man + fert.	0.5000	1.6697	0.299	0.76680
Treatmentmucuna	-3.3000	1.6697	-1.976	0.05803 .
TreatmentMucuna + fertilizer	-2.0000	1.6697	-1.198	0.24102
TreatmentMucuna + fertilizer + manure	-0.6500	1.4460	-0.450	0.65651
TreatmentTith + fert	-0.3500	1.3200	-0.265	0.79283
TreatmentTithonia + manure	2.5000	1.6697	1.497	0.14550
TreatmentTithonia + manure + fertilizer 0.3333	1.3633	0.245	0.80862	
TreatmentTithonia alone	-1.2000	1.3633	-0.880	0.38622
—				
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
(Dispersion parameter for gaussian family taken to be 1.393878)				
Null deviance: 99.456 on 48 degrees of freedom				
Residual deviance: 39.029 on 28 degrees of freedom				
AIC: 171.91				
Number of Fisher Scoring iterations: 2				

Table 3. Results for GLM Type II experiment

From the table 3 above, represents results from generalized linear model (GLM) assuming the Gaussian family. We notice that some treatment have negative effects while others have positive effects. Among those are significantly explaining the response variable at 0.05 level. These are shown by '**' at the end while the rest are non-significant.
Also on the deviance, we have null deviance at 99.456 on 48 degrees of freedom for the null model with intercept only while the residual deviance is 39.029 on 28 degrees of freedom hence we conclude the model fitted is the good fit.
The fisher scoring algorithm we only needed two iterations to perform the fit.

4.4.3 Mixed Model

The mixed model is fitted to data from Type II experiment and results represented in the table 4 below.

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	3.5000	1.1806	2.965	0.00613 **
TreatmentCalliandra + fertilizer	-1.5000	1.6697	-2.198	0.03766 .
TreatmentCalliandra + manure	-0.7500	1.4460	-0.519	0.60806
TreatmentCalliandra + tithonia + fart	-1.5000	1.6697	-2.198	0.03766 .
Treatmentcontrol	-2.4000	1.2752	-1.882	0.07026 .
TreatmentCrotalaria _+ leucaena	-1.7000	1.6697	-1.018	0.31731
TreatmentCrotalaria + fertilizer	-1.4000	1.6697	-0.838	0.40885
TreatmentCrotalaria + manure + fert	-0.5000	1.6697	-0.299	0.76680
TreatmentFertilizer	-0.5000	1.3633	-0.367	0.71655
TreatmentLeucaena+ manure	1.2000	1.4460	0.830	0.41362
TreatmentManure + calliandra	0.4000	1.6697	2.240	0.01812 .
Treatmentmanure + fertilizer	-0.5857	1.2621	-0.464	0.64619
TreatmentManure alone	0.3667	1.2752	0.288	0.77582
TreatmentMuc + Man + fert.	0.5000	1.6697	2.299	0.01766 .
Treatmentmucuna	-3.3000	1.6697	-2.176	0.04803 .
TreatmentMucuna + fertilizer	-2.0000	1.6697	-2.198	0.02410 .
TreatmentMucuna + fertilizer + manure	-0.6500	1.4460	-0.450	0.65651
TreatmentTith + fert	-0.3500	1.3200	-0.265	0.79283
TreatmentTithonia + manure	2.5000	1.6697	2.597	0.01455 .
TreatmentTithonia + manure + fertilizer 0.3333	1.3633	0.245	0.80862	
TreatmentTithonia alone	-1.2000	1.3633	-0.880	0.38622
—				
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error: 1.181 on 28 degrees of freedom				
Multiple R-squared:	0.6076,	Adjusted R-squared:	0.3273	
F-statistic: 2.168 on 20 and 28 DF,	p-value: 0.02944			

Table 4. Results for Mixed model

From the table 4 above, we conclude the mixed model is significant at the 0.05 level, since $0.02944 < 0.05$. Also, some treatments are individually significant as shown and flagged by '.'

4.4.4 Model comparisons

In this subsection we compare the three models fitted above to data from Type II experiment by using Akaike's Information Criterion.

Models	AIC
Fixed effects model	170.131
Generalized Linear model	171.9072
Mixed model	169.071

Table 5. Model Comparison Results

From the table 5 above ,we can conclude that the mixed model is the best fit for the data from Type II experiment since it is AIC values is the smallest among the three.

4.5 Statistical models fitted Type III experiment data

4.5.1 Fixed effect model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	12	74.44	6.203	3.35	0.00077 ***
Residuals	66 122.23	1.852			
—					
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 6. Fixed effect model – Type III

Since $p-value = 0.00077 < 0.05$, we conclude that the model is significant fitting the data from Type III experiment.

Plot of residuals & fitted values and QQ norm for the fixed effect model

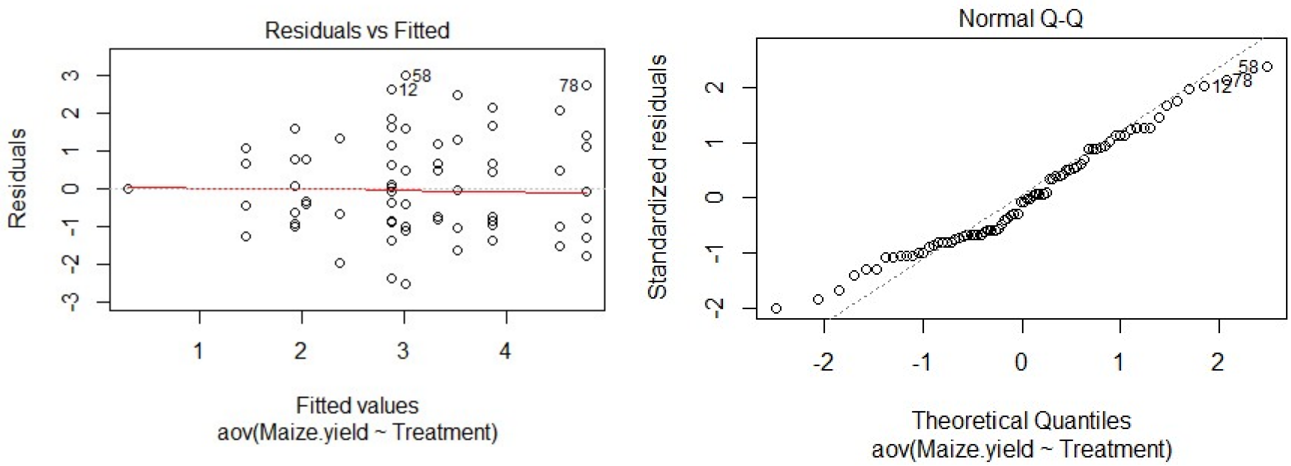

Figure 5. Plot of residuals & fitted values and QQ norm for the fixed effect model

4.5.2 Generalized Linear model (GLM)

We fit the GLM to data from Type III experiment and the results obtained are presented in the table 7 below.

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	3.8600	0.4303	8.970 4.92e-13 ***	
Treatmentcontrol	-1.9350	0.6455	-2.998	0.00383 **
TreatmentControl	-3.5600	1.4273	-2.494	0.01513 *
TreatmentCrot + Fert.	-0.8457	0.6706	-1.261	0.21172
TreatmentCrotalaria	-2.4100	0.8051	-2.993	0.00388 **
TreatmentFetilizer	-0.9800	0.6086	-1.610	0.11211
TreatmentLeuc + fert.	-0.3314	0.6706	-0.494	0.62280
TreatmentManure + fertilizer	0.9275	0.6455	1.437	0.15548
TreatmentManure alone	-1.4850	0.8051	-1.845	0.06960 .
TreatmentMucuna	-1.8267	0.8958	-2.039	0.04545 *
TreatmentMucuna + fert.	-0.9886	0.6706	-1.474	0.14521
TreatmentTithonia + fertilizer	-0.5267	0.7027	-0.749	0.45625
TreatmentTithonia alone	0.6650	0.8051	0.826	0.41178
—				
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
(Dispersion parameter for gaussian family taken to be 1.851907)				
Null deviance: 196.66	on 78	degrees of freedom		
Residual deviance: 122.23	on 66	degrees of freedom		
AIC: 286.67				
Number of Fisher Scoring iterations: 2				

Table 7. Results from generalized linear model (GLM) assuming the Gaussian family

The table 7 above represents the results from generalized linear model (GLM) assuming the Gaussian family. We notice that some treatments have negative effects while others have positive effects. Among those are significantly explaining the response variable at 0.05 level are shown by '**,* .' at the end while the rest are non-significant.

Looking at the deviance, we have null deviance at 196.66 on 78 degrees of freedom for the null model with intercept only while the residual deviance is 122.23 on 66 degrees of freedom hence we conclude the model fitted is the good fit.

The fisher scoring algorithm we only needed two iterations to perform the fit.

4.5.3 Mixed Model

We fit the GLM to data from Type III experiment and the results obtained are presented in the table 8 below.

Coefficients:				
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	3.8600	0.4303	8.970 4.92e-13 ***	
Treatmentcontrol	-1.9350	0.6455	-2.998	0.00383 **
TreatmentControl	-3.5600	1.4273	-2.494	0.01513 *
TreatmentCrot + Fert.	-0.8457	0.6706	-1.261	0.21172
TreatmentCrotalaria	-2.4100	0.8051	-2.993	0.00388 **
TreatmentFetilizer	-0.9800	0.6086	-1.610	0.11211
TreatmentLeuc + fert.	-0.3314	0.6706	-0.494	0.62280
TreatmentManure + fertilizer	0.9275	0.6455	1.437	0.15548
TreatmentManure alone	-1.4850	0.8051	-1.845	0.06960 .
TreatmentMucuna	-1.8267	0.8958	-2.039	0.04545 *
TreatmentMucuna + fert.	-0.9886	0.6706	-1.474	0.14521
TreatmentTithonia + fertilizer	-0.5267	0.7027	-0.749	0.45625
TreatmentTithonia alone	0.6650	0.8051	0.826	0.41178
—				
Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error: 1.361 on 66 degrees of freedom				
Multiple R-squared:	0.3785,	Adjusted R-squared:	0.2655	
F-statistic:	3.35 on 12 and 66 DF,	p-value: 0.0007696		

Table 8. Results for Mixed model

From the table 8 above, we conclude the mixed model is significance since $0.0007696 < 0.05$. Also, some of the treatments are individually significant at level of significance at 0.05 as shown in the table above by '**,*,.'.

4.5.4 Model comparisons

In this subsection we compare the three models fitted above to data from Type III experiment by using Akaike's Information Criterion.

Models	AIC
Fixed effects model	286.13
Generalized Linear model	286.67
Mixed model	280.01

Table 9. Model Comparison Results

From the table 9 above, we can conclude that the mixed model is the best fit for the data from the Type III experiment.

5 CONCLUSIONS AND RECOMMENDATION

5.1 Introduction

The main objective of this research was to determine the variability or treatment effects in type II and III experiments involving improved yields in Murangiri, Kenya. This objective was split into specific objectives so that to help us adequately achieve the main objective. These specific objectives were as follows:

i. To determine the best model that explains the variability of yields in type II and III experiments involving improved yields in Murangiri, Kenya.

ii. To determine the variability of yields in type II and III experiments involving improved yields in Murangiri, Kenya.

5.1.1 Interpretations of results

In this study we investigated the effects of different treatments ranging from organic, inorganic and mixture of the two on the maize yields in Murangiri in Tharaka Nithi constituency in Kenya. We applied three statistical models to the data obtained from Type II and Type III experiments namely ; fixed effects model, Generalize linear Model (GLM) and mixed model. We focused on interpretations and computation of model parameters and also we investigated which model best fits the two datasets from the two experiments.

Our study found that the treatments in general had the effects on the Maize yields in the two experiments as shown by all models fitted since the *p-values* of both mixed and fixed effect model are less than level of significance 0.05 while for GLM by using the deviance we show that the fitted model with treatments are significant on both cases.

On the best model, we used the model comparisons Akaike's Information Criterion (AIC) to determine model best fit the two data from the Type II and Type III experiments respectively. The study found that the mixed model was the best among the three models considered under this study as it was having the smallest values of AIC 169.071 and 280.01 for Type II and Type III experiments respectively as indicated in table 5 and 9 respectively.

5.1.2 Recommendations

Since the improvement of maize yields is critical to the eradication of starvation and hunger among the people in the world, modeling the effects of different treatments on maize yields is in line with the 2030 vision of eliminating food shortages across the country. Further studies should be done on maize yields research using different treatments and different flexible statistical methodologies. Although this research is motivated by maize yields studies, the basic concepts and methods developed here have much broader applications in other crop production systems.

References

[1] Cengiz Acar, Habibe Acar, and Lokman Altun. The diversity of ground cover species in rocky, roadside and forest habitats in trabzon (north-eastern turkey). *BIOLOGIA-BRATISLAVA-*, 59(4):477–492, 2004.

[2] OE Aina, SO Amoo, Liziwe Lizbeth Mugivhisa, Joshua Oluwole Olowoyo, et al. Effect of organic and inorganic sources of nutrients on the bioactive compounds and antioxidant activity of tomato. *Applied Ecology and Environmental Research*, 17(2):3681–3694, 2019.

[3] AA Ajao and AN Moteetee. Tithonia diversifolia (hemsl) a. gray.(asteraceae: Heliantheae), an invasive plant of significant ethnopharmacological importance: A review. *South African Journal of Botany*, 113:396–403, 2017.

[4] Julia Bailey-Serres, Jane E Parker, Elizabeth A Ainsworth, Giles ED Oldroyd, and Julian I Schroeder. Genetic strategies for improving crop yields. *Nature*, 575(7781):109–118, 2019.

[5] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

[6] Andrew Bell, Malcolm Fairbrother, and Kelvyn Jones. Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53(2):1051–1074, 2019.

[7] Lucia Helena Garófalo Chaves, Danila Lima Araujo, Hugo Orlando Carvallo Guerra, and Walter Esfrain Pereira. Yield of sunflower affected by npk fertilizer and available soil water content. In *2014 Montreal, Quebec Canada July 13–July 16, 2014*, page 1. American Society of Agricultural and Biological Engineers, 2014.

[8] Paul Clarke, Claire Crawford, Fiona Steele, and Anna F Vignoles. The choice between fixed and random effects models: some considerations for educational research. 2010.

[9] AB Dafaallah. Fundamentals of design and analysis of agricultural experiments (observation–experimentation–discussion). *Part one. First Edition. University of Gezira House for Printing and Publishing, Wad Madani, Sudan*, 246, 2017.

[10] Saleha Farjana, Md Ashraful Islam, and Tamanna Haque. Effects of organic and inorganic fertilizers, and mulching on growth and yield of cabbage (brassica oleracea var. capitata l.). *Journal of Horticulture and Postharvest Research*, 2(2-September 2019):95–104, 2019.

[11] VK Gupta, Rajender Parsad, and Baidya Nath Mandal. Significance of experimental designs in agricultural research. 2015.

[12] Alex Horsley, Llinos Jones, John White, and Michael Henry. Efficacy and complications of small-bore, wire-guided chest drains. *Chest*, 130(6):1857–1863, 2006.

[13] Kifayatullah Kakar, Tran Dang Xuan, Zubair Noori, Shafiqullah Aryan, and Gulbuddin Gulab. Effects of organic and inorganic fertilizer application on growth, yield, and grain quality of rice. *Agriculture*, 10(11):544, 2020.

[14] Dominique Lord, Seth D Guikema, and Srinivas Reddy Geedipally. Application of the conway–maxwell–poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*, 40(3):1123–1134, 2008.

[15] Peter McCullagh and JA Nelder. Generalized linear models ii, 1989.

[16] Emmanuel Ibukunoluwa Moyin-Jesu. Simple and blended organic fertilizers improve fertility of degraded nursery soils for production of kolanut (cola acuminate) seedlings in nigeria. *SOIL FERTILITY IMPROVEMENT AND INTEGRATED NUTRIENT MANAGEMENT–A GLOBAL PERSPECTIVE*, page 293, 2011.

[17] Jan Serroyen, Geert Molenberghs, Geert Verbeke, and Marie Davidian. Nonlinear models for longitudinal data. *The American Statistician*, 63(4):378–388, 2009.

[18] P Sharada and P Sujathamma. Effect of organic and inorganic fertilizers on the quantitative and qualitative parameters of rice (oriza sativa l.). *Current Agriculture Research journal*, 6(2):166, 2018.

[19] Bernard Vanlauwe, Job Kihara, Pauline Chivenge, Pieter Pypers, Ric Coe, and Johan Six. Agronomic use efficiency of n fertilizer in maize-based systems in sub-saharan africa within the context of integrated soil fertility management. *Plant and soil*, 339(1):35–50, 2011.

[20] Václav Voltr, Ladislav Menšík, Lukáš Hlisnikovskỳ, Martin Hruška, Eduard Pokornỳ, and Lubica Pospíšilová. The soil organic matter in connection with soil properties and soil inputs. *Agronomy*, 11(4):779, 2021.

[21] Muhammad Ahmed Waqas, Pete Smith, Xiaohan Wang, Muhammad Nadeem Ashraf, Mehmood Ali Noor, Martial Amou, Shengwei Shi, Yongchang Zhu, Jianling Li, Yunfan Wan, et al. The influence of nutrient management on soil organic carbon storage, crop production, and yield stability varies under different climates. *Journal of Cleaner Production*, 268:121922, 2020.

[22] Ismail Yasmeena, SA Mir, MA Bhat, Nazir Nageena, et al. An application of generalized linear model in survival analysis. *Asian Journal of Home Science*, 13(1):68–74, 2018.

APPENDIX A

R codes used in analysis.

```
###################Data from Type II experiments####
getwd()
RCBD<-read.csv("Type 3.csv",header=T)
RCBD
names(RCBD)
Treatment<-factor(RCBD$Treatment)
Treatment
Maize.yield<-RCBD$Maize.yield
Maize.yield
Farmers.name<-RCBD$Farmers.name
Farmers.name
fit1=aov(Maize.yield~Treatment)
stripchart(Maize.yield~Treatment,vertical=TRUE,data =RCBD,
main="Strip chart Type II DATA" )
qqnorm(Maize.yield,main = "Type II Data")
qqline(Maize.yield)
boxplot(Maize.yield~Treatment,data=RCBD,ylab="Maize yield",
main="Box plot Type II DATA")
shapiro.test(Maize.yield)
result=summary(fit1)
result
fit2=glm(Maize.yield~Treatment,
family = gaussian)##fitting GLM assuming Gaussian##
summary(fit2)
fit5=lm(Maize.yield~Treatment+(1/Farmers.name))###Mixed effect model###
summary(fit5)
plot(fit5,which=1)
plot(fit5,which=2)


###################Data from Type III experiments####
getwd()
RCBD1<-read.csv("Type 2.csv",header=T)
RCBD1
names(RCBD1)
Treatment<-factor(RCBD1$Treat)
Treatment
Maize.yield<-RCBD1$Mz_yield
Maize.yield
```

```
Farmers.name<-RCBD1$Farmer
fit3=aov(Maize.yield~Treatment)
result=summary(fit3)
result
plot(fit3,which=1)
plot(fit3,which=2)
stripchart(Maize.yield~Treatment,vertical=TRUE,
data =RCBD1,main="Strip chart Type III DATA" )
boxplot(Maize.yield~Treatment,data=RCBD1,
ylab="Maize yield",main="Box plot Type III DATA")
qqnorm(Maize.yield,main = "Type III data")
shapiro.test(Maize.yield)
qqline(Maize.yield)
fit4=glm(Maize.yield~Treatment,
family = gaussian)##fitting GLM assuming Gaussian##
summary(fit4)
fit6=lm(Maize.yield~Treatment+(1/Farmers.name))###Mixedeffect model###
summary(fit6)
```