International Journal of Applied Science and Mathematical Theory ISSN 2489-009X Vol. 4 No. 1 2018
www.iiardpub.org

Dummy Variable Regression Model and Two-Stage Nested Design of Agricultural Variables

Johnson, Q. E. & Essi, I. D.
Rivers State University, (RSU),
Nkpolu Oroworukwo,
Port Harcourt.
mrs.calvary@gmail.com, love4op@yahoo.com

Abstract

This research work is an attempt to obtain a dummy variable regression model and two-Stage nested design of agricultural variables in Nigeria. Univariate General Linear Model (GLM), Multiple Comparisons and Two-Stage Nested Design methods were used to compare the mean response for the different crops with respect to Area Planted and Production. These three methods are preferred in determining significance between or among factors. The dummy variable Regression model (GLM) result shows that cassava and yam production are significant at 5 percent, while others are not. In addition, two-Stage Nested Design ANOVA result also confirmed the result of dummy variable Regression model (GLM). This result can serves as a guide to individuals or corporate bodies that may wish to involve in area-planted/production of different crops in Nigeria.

Keyword: *Dummy Variable, Regression Model, Two-Stage Nested Design, Multiple Comparisons*

1. General Description

The examination of a baffling system requires the recognizing confirmation of target quality characteristics that depict the yield of the technique and of factors that may be related to those properties. Once a summary of potential components is recognized from subject-matter aptitude, the characteristics of the connection between those components and the target credits ought to be estimated. As an expert, norms of quantifiable speculation, straight factor based math, and examination control the headway of beneficial exploratory plans for factor settings. Once a subset of basic components has been withdrawn, resulting experimentation can choose the settings of those components that will update the target quality attributes.

Fortunately, modern software has exploited the propelled hypothesis. This product now encourages the advancement of good research work (or great plan) and makes strong examination more open to those with a negligible factual foundation. This product is more effective, finish, savvy, and less error-prone than creating a similar plan by hand with tables.
Also, it enables to make algorithmic plans (according to one of a couple of possible optimality criteria) that are as frequently as conceivable required to oblige objectives routinely experienced before long. Once a trial has been arranged and executed, the examination of the results should respect the doubts made in the midst of the layout method. For instance, this exploration work is fundamentally on Nested Design (or Hierarchical Models), which is use to perform completely settled (various levelled) investigation of fluctuation and to evaluate difference segments for every reaction variable where reactions demonstrate inconstancy that is in part the consequence of the impact and is somewhat arbitrary blunder. It is utilized to test theory that the mean estimations of the persistent variable are the same in various gatherings, when each gathering is isolated into subgroups,

subgroups must be discretionary (display II). Some use of settled outline in a few fields will be finished.

In specific examinations the levels of one factor (e.g. Factor B) are comparative yet not indistinguishable for various levels of another factor (e.g. Factor A). Such a course of action is known as a settled or various levelled outline, with the levels of factor B settled under the levels of factor A.

In some two-factor explores the level of one factor, say B, isn't "crossed" or "cross grouped" with the other factor, say A, yet is "Settled" with it. The levels of B are distinctive for various levels of A.

In these cases we are constrained into what is known as a settled format. We say we have a settled format when less than all levels of one factor happen inside each level of the other factor.

In any case, the statements of problem of this research work are the problems of finding/identification of target quality attributes and estimating variance components for each response variable (or discovering/distinguishing proof of target quality properties and assessing change segments for every reaction variable). Where reactions indicate inconstancy that is halfway the after effect of the impact and is incompletely arbitrary blunder, when the information of the distinctive fields are built up.

Decide when to apply Nested Design (or Hierarchical Models) in various field; such Social Science, Biological, Agricultural, Engineering, Physical Science and monetary. Additionally test theory that the mean estimations of the consistent variable are the same in various groups, when each group is partitioned into subgroups. The main objective of this research work is to provide a dummy variable regression analysis and two-Stage Nested Design of Agricultural variables in Nigeria. The objectives are; (1) Determine significant difference of the factors for the agriculture variables considered, using dummy variable regression model. (2) Use two-stage nested design to test: (a) difference between “crop” with respect to area planted and production; then (b) the variability of the factors A within factors B. (3) To find a significant variability among the factors A within factors B in each field region, then a significant difference between Factors B would be study if there is an effect, using multiple comparisons. Note: Factor A is Area planted and production; and Factor B is crop planted.
A nested design is suggested for concentrate the impact of wellsprings of inconstancy that show themselves after some time (or for studying the effect of sources of variability that manifest themselves over time). This exploration work means to consider the usage of settled plan in green portion and it will fill in as a basic material for pros and expert (or serves as an important material for researchers and statistician). Furthermore, fill in as a manual for individuals or collaborate bodies that may wish to include in planted/creation of different harvest items in Nigeria.

2. Nested Design or Hierarchical Models

A studying of this subject is basic subsequently to any cutting edge scholar, additionally unique fields of concentrate, for example, fund, medication, science, building and numerous different fields; demonstrating information successfully is imperative for some, basic leadership exercises, and it is without a doubt utilized various circumstances to test logical theories. There are three important (basic) designs in the analysis of variance:-a.) Completely randomized design, b.) Randomized complete block design and c.) Latin square. However, some others based on these three are factorial design and split-plot design, and *nested design*. Nested designs, also known as hierarchical designs, are covered in Sokal and Rohlf (1995), Sall and Lehman (1996), Zar (1984, 1995, 1999), and Ott and Longnecker (2001). According to Sokal and Rohlf (1995), nested designs usually contain *random effects* and are called Model II ANOVA. Nested designs are utilized when there are samples within samples. In horticulture, for instance, an investigator should need to think about the transpiration rates of five cross breeds of specific types of plant. For every hybrid, six plants are developed in three pots, two plants for every pot. At the end of the growth time frame, transpiration is estimated on four leaves of each plant. We in this way have leaves settled inside plants settled inside pots settled inside half breeds.

According to Crawley (2004), there are three main parts to this practical: Nested Designs, Designed Split-Plot Experiments and Mixed Effects Models. They are linked by two facts: (1) they involve categorical variables of two kinds (fixed effects and random effects); and (2) because their data frames all involve pseudo replication, they offer great scope for getting the analysis wrong.

Teddlie and Tashakkori, (2003), stated that mixed method research studies use qualitative and quantitative data collection and analysis techniques either in parallel or sequential phases." The benefit of joining different strategies is the assistance it gives towards better noting given research inquiries, and making more vigorous determinations. Drawing better and more dependable conclusions is conceivable when the consolidated utilization of the picked strategies strengthens and finishes every individual technique, and in this way diminishes the shortcomings and insufficiencies of each (Teddlie and Tashakkori, 2003).

According to Denzin (2009, 1978), in any examination the energy of the exploration configuration can basically be extended through triangulation, that is, the utilization of methodological blends. Denzin (1978) recognizes four fundamental techniques for triangulation. The first is triangulation of information, that is, the utilization of different information sources. The second is triangulation of members, i.e. including different specialists. The third is hypothetical triangulation, that is, the elucidation of results along different points of view and hypothetical foundations. The fourth is methodological triangulation, i.e., the utilization of different procedures in the investigation of a given research zone.

Lieberman's (2005) offered a venture which joined a relevant examination with factual investigation battling that, past the diverse purposes of enthusiasm of each approach, there is a sensible synergistic estimation of the settled research plan. He contended that "statistical analyses can guide case selection for in-depth research, provide direction for more focused case studies and comparisons, and be used to provide additional tests of hypotheses generated from small-N research. Small-N analyses can be used to assess the plausibility of observed statistical relationships between variables, to generate theoretical insights from outlier and other cases, and to develop better measurement strategies" (Lieberman, 2005). Lieberman's study conveys the advantages of unmistakable complementarities and a move from small-N analysis (SNA) towards Nested analysis. He portrays an arrangement of techniques for increasing most extreme explanatory use when consolidating SNA and LNA within a single framework (summarized in LNA is defined by Lieberman (2005) as "a mode of analysis in which the primary causal inferences are derived from statistical analyses which ultimately lead to quantitative estimates of the robustness of a theoretical model and SNA as a mode of analysis in which causal inferences about the primary unit under investigation are derived from qualitative comparisons of cases and/or process tracing of causal chains within cases across time, and in which the relationship between theory and facts is captured largely in narrative form" (Harrits, 2011).

Lieberman (2005) states that nested analysis formally begins with a quantitative analysis, or preliminary LNA, he stated that a prerequisite for carrying out a nested analysis is availability of a quantitative dataset, with a sufficient number of observations for statistical analysis. The preliminary LNA provides information that should ultimately complement the findings of the SNA, and that will guide the execution of the SNA" (Lieberman, 2005). The preliminary LNA ensures an express thought of the universe of cases for which the theory should apply. Additionally the objective of this preliminary LNA is to investigate however many fitting testable hypotheses as could be allowed. This LNA study obviously may have a few structures relying upon the availability of data (dichotomous, probabilistic or deterministic relationships).

Lieberman (2005) in the meantime separates LNA and SNA from the generally utilized strategies for quantitative and qualitative analysis. He contends that SNA may likewise incorporate quantitative analysis. This paper is basically on Nested Design (or Hierarchical Models), which is use to perform completely settled (progressive) investigation of change and to evaluate difference segments for every reaction variable. Where reactions demonstrate changeability that is halfway the after effect of the impact and is somewhat arbitrary mistake. It is utilized to test theory that the mean estimations of the consistent variable are the same in various groups, when each group is separated into subgroups, subgroups must be subjective (model II).

3. Materials and Methods

Data would be analysed using statistical package particularly designed for analysis. The statistical software's are Micro-Excel, Minitab 17 and IBM SPSS (Version 21).

3.1 General Linear Model (GLM) or Dummy variables Regression model

In matrix terms, the general linear regression model is:

$$\mathrm{Y} = \mathrm{X}\beta + \in_i \tag{3.1}$$

where

Y is a vector of responses,

β is a vector of parameters,

X is the design matrix of constants and e is a vector of independent normal random variables Kutner, *et al.*, (1985).

We estimate Equation (3.1), using the dummy variables regression model of the form;

$$y = \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ . \\ . \\ . \\ y_{ijk} \end{pmatrix}, \quad x = \begin{pmatrix} 1 & 1 & 0 & 0 & . & . & . & 0 \\ 1 & 1 & 0 & 0 & . & . & . & 0 \\ 1 & 1 & 0 & 0 & . & . & . & 0 \\ . & . & . & . & & & & . \\ . & . & . & . & & & & . \\ . & . & . & . & & & & . \\ 1 & 0 & . & . & . & & & 1 \\ 1 & 0 & . & . & . & & & 1 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{pmatrix}$$

Then

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{pmatrix} = (x'x)^{-1}(x'y) \qquad (3.2)$$

Design matrix: General Linear Model uses a regression approach to fit the model that is specify in Equation (3.1). The Minitab statistical software was used to create a design matrix, from the factors and covariates, and the model in Equation (3.1) was obtained. The columns of this matrix are the predictors for the regression.

The design matrix has n rows, where n = number of observations and several blocks of columns, corresponding to the terms in the model. The first block is for the constant and contains just one column, a column of all ones. The block for a covariate also contains just one column, the covariate column itself. The block of columns for a factor contains r columns, where r = degrees of freedom for the factor, and they are coded as shown in Table 3.1 below.

Since Region is a factor with 6 levels. Then it has 3 degrees of freedom and its block contains 2 columns (Area and production), call them constants, Cassava Area (CA), Rice Area (RA), Yam Area (YA), . . . , Beans/Cowpea (BC.P)

Table 3.1: Design Matrix coded for GLM

level of A	Constant	CA	RA	YA
1	1	1	0	0
2	1	1	0	0
3	1	1	0	0
4	1	1	0	0
5	1	1	0	0
6	1	1	0	0
1	1	0	1	0
2	1	0	1	0
3	1	0	1	0
4	1	0	1	0
5	1	0	1	0
6	1	0	1	0
1	1	0	0	1
2	1	0	0	1
3	1	0	0	1
4	1	0	0	1
5	1	0	0	1
6	1	0	0	1

etc.

3.2 The two-Stage Nested Design Model Specification

When factor B is nested in levels of factor A, the levels of the nested factor don't have exactly the same meaning under each level of the main factor, in this case factor A. In a nested design, the levels of factor (B) are not identical to each other at different levels of factor (A), although they might have the same labels.

- **Model specification:**

The linear statistical model for the two-stage nested design is:

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \in_{(ij)k} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases} \qquad (3.3)$$

where there are a levels of Factor A, b levels of Factors B nested under each level of A, and n replicates. The subscript i indexes "A" (often called the "major factor"), (i)j indexes "B" within "A" (B is often called the "minor factor") and (ij)k indexes replication.

The total corrected sum of squares as

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{...}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\left(\bar{Y}_{i..}-\bar{Y}_{...}\right)+\left(\bar{Y}_{ij.}-\bar{Y}_{i..}\right)+\left(\bar{Y}_{ijk}-\bar{Y}_{ij.}\right)\right)^2 \qquad (3.4)$$

Expanding the right-hand side of Equation (3.2)

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{...}\right)^2 = bn\sum_{i=1}^{a}\left(\bar{Y}_{i..}-\bar{Y}_{...}\right)^2 + n\sum_{i=1}^{a}\sum_{i=1}^{b}\left(\bar{Y}_{ij.}-\bar{Y}_{i..}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\bar{Y}_{ijk}-\bar{Y}_{ij.}\right)^2 \qquad (3.5)$$

Equation (3.4) indicates that the total sum of squares can be partitioned into a sum of squares due to factor A, a sum of square due to factor B under the levels of Factor A, and a sum of squares due to errors. Symbolically, we can write Equation (3.4) as

$$SS_T = SS_A + SS_{B(A)} + SS_{\in} \qquad (3.6)$$

There are abn -1 degrees of freedom for SS_T , a -1 degrees of freedom for SS_A , a(b-1) degrees of freedom for $SS_{\in}$. Note that $abn - 1 = (a-1) + a(b-1) + ab(n-1)$. If errors are $NID(0, \sigma^2)$, we may divide each sum of squares on the right of Equation (3.5) by its degrees of freedom to obtain independently distributed mean squares such that the ratio of any two mean squares is distributed as F.

The appropriate statistics fort testing the effects of factor A and B depend on whether A and B are fixed or random. If factors A- fixed and Factor B fixed, we assumed that $\sum \tau_i^2 = 0$ and $\sum \beta_{j(i)} = 0$ (i = 1,2,…,a). That is, the A treatment effects sum to zero, and the B treatment effects sum to zero within each level of A. Alternatively, if A and B are random, the we assume that τ_i is NID(0, σ_τ^2) and $\beta_{j(i)}$ is NID(0, σ_β^2).

Table 3.2: Expected Mean Squares in the Two-Stage Nested Design

$E(MS)$	A Fixed B Fixed	A Fixed B Random	A Random B Random
$E(MS_A)$	$\sigma^2 + \frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\tau^2$
$E\left(MS_{B(A)}\right)$	$\sigma^2 + \frac{n\sum\sum\beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$E(MS_\in)$	σ^2	σ^2	σ^2

Table 3.2 indicates that Expected Mean Squares in the Two-Stage Nested Design while the ANOVA for Two-Stage Nested Design is as follow:-
Computing formulas for sums of squares may be obtained by expanding the quantities in Equation (3.4) and simplifying. They are

$$SS_A = \sum_{i=1}^{a}\frac{Y_{i..}^2}{bn} - \frac{Y_{...}^2}{abn} \tag{3.6}$$

$$SS_{B(A)} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{n} - \sum_{i=1}^{a}\frac{Y_{i..}^2}{bn} \tag{3.7}$$

$$SS_\in = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}Y_{ijk}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{n} \tag{3.8}$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}Y_{ijk}^2 - \frac{Y_{...}^2}{abn} \tag{3.9}$$

Equation 3.6 for $SS_{B(A)}$ can be written as

$$SS_{B(A)} = \sum_{i=1}^{a}\left[\sum_{i=1}^{b}\frac{Y_{ij.}^2}{n} - \frac{Y_{i..}^2}{bn}\right] \tag{3.10}$$

Table 3.3: Analysis of Variance Table for the Two-Stage Nested Design

Sources of variation	Sum of squares	Degrees of Freedom	Mean Square
A	$\sum_{i=1}^{a}\frac{Y_{i..}^2}{bn} - \frac{Y_{...}^2}{abn}$	a -1	MS_A
B within A	$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{n} - \sum_{i=1}^{a}\frac{Y_{i..}^2}{bn}$	a(b – 1)	$MS_{B(A)}$
Error	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}Y_{ijk}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{n}$	ab(n – 1)	$MS_\in$
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}Y_{ijk}^2 - \frac{Y_{...}^2}{abn}$	abn – 1	

However, the Analysis of Variance Table for the Two-Stage Nested Design Hypothesis Testing considered in this research is A Fixed (Area/product) and B Fixed (crops) Factor.

4. Data and Results

The data considered Nigeria estimates of Agricultural production (Thousand Metric Tons) and Area Planted (Thousand Hectares) by Crop and Region, 2016. The section is divided into two part, (1) Preliminary Analysis of the response base on dummy variable regression techniques to determine the effect of the crops production and area planted [General Linear Model (GLM)]. (2) the two-Stage Nested Design to determine the appropriate statistics for testing the effects of factors Crop and Region depend on whether Crop and Region are fixed or random.

4.1 Dummy Variables Regression Analysis

General Linear Model (GLM) in Section 3, Equation (3.1) was used to determine each effect by region on crops production and area planted (See Appendix D for the design matrix coded and Appendix E for Minitab 17 results output). Table 4.1 is a summary of the GLM analysis.

Table 4.1: GLM analysis result

Variables	Estimates (p-values)
Constant	778 (0.470)
Cassava Area	240 (0.874)
Rice Area	75 (0.961)
Yam Area	126 (0.934)
Palm oil Area	-239 (0.875)
Beans/Cowpea Area	49 (0.974)
Cassava Production	10894 (0.000)*
Rice Production	1345 (0.379)
Yam Production	8605 (0.000)*
Palm oil Production	-367 (0.810)
Beans/Cowpea Production	Highly correlated

Footnote: * significant at 5 percent

From Table 4.1, the dummy variable regression model (GLM) result shows that cassava and yam production are significant at 5 percent, while others are not. It indicated that cassava and yam production yield's more in all regions in Nigeria.

4.2 Two Stage Nested Design Analysis

The row, column and interaction total were computed in Table 4.2; then two-stage nested design descript in section 3.4 chapter three was computed, using Micro-Excel, Minitab 17 and SPSS 21. The blocks [Area/Planted (Thousand Hectares) and Production (Thousand Metric Tons)] a = 2; Crops b = 5; region n = 6.

Computing Equation (3.9), we have

$$SS_T = \left(1537.54^2 + 1402.11^2 + \ldots + 700.62^2 + 235^2\right) - \frac{(171070.561)^2}{2\times5\times6}$$

$$=1742406017.00 - 487752280.00$$

$$=1254653737.00$$

Also, using Equation (3.6), we computed SS_A

$$SS_A = \frac{1}{5\times 6}\left(24859.45^2 + 146211.11^2\right) - \frac{(171070.56)^2}{2\times 5\times 6}$$

$$= 733189389.48 - 487752280.00$$

$$= 245437108.80$$

Using Equation (3.7), we computed $SS_{(B)A}$

$$SS_{(B)A} = \frac{1}{6}\left(6112.84^2 + 5117.76^2 + \ldots + 4670.37^2\right) - \frac{1}{5\times 6}\left(24859.4^2 + 133472.21^2\right)$$

$$= 1398742198.00 - 733189389.48$$

$$= 665552813.75$$

Using Equation (3.8), we computed SS_e

$$SS_e = \left(1537.45^2 + 1402.11^2 + \ldots + 235^2\right) - \frac{1}{6}\left(6112.84^2 + 5117.76^2 + \ldots + 4670.37^2\right)$$

$$= 1742406017 - 1398742198.00$$

$$= 343663819.00$$

We summarized in Table 4.3, using Table 3.2

Table 4.3: Analysis of Variance for the Two-Stage Nested Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square	F_{Cal}	P-value
SS_A	245437108.80	1	245437108.80	$\sigma^2 + 4\sigma_B^2 + 4\sum \tau_i^2$	2.95	0.000
$SS_{(B)A}$	665552813.75	8	83194101.71	$\sigma^2 + 4\sigma_B^2$	12.10	0.000
Error	343663819.00	50	6873276.38	σ^2		
Total	1254653737.00	59				

- Diagnostic Checking

Residual analysis is used for diagnostic checking. For the two-stage nested design, the residuals are

$$\epsilon_{ijk} = y_{ijk} - \hat{y}_{ijk}$$

The fitted value is

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_{j(i)}$$

and the usual restriction on the model parameters

$$\sum \tau_i = 0 \text{ and } \sum \hat{\beta}_{j(i)} = 0 \ , \quad i = 1, 2, \ldots, a$$

Then,

$$\hat{\mu} = \bar{y}_{\ldots} , \quad \hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{\ldots} \text{ and } \hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$$

Consequently,

$$y_{ijk} = \bar{y}_{...} + \left(\bar{y}_{i..} - \bar{y}_{...}\right) + \left(\bar{y}_{ij.} - \bar{y}_{i..}\right) = \bar{y}_{ij.}$$

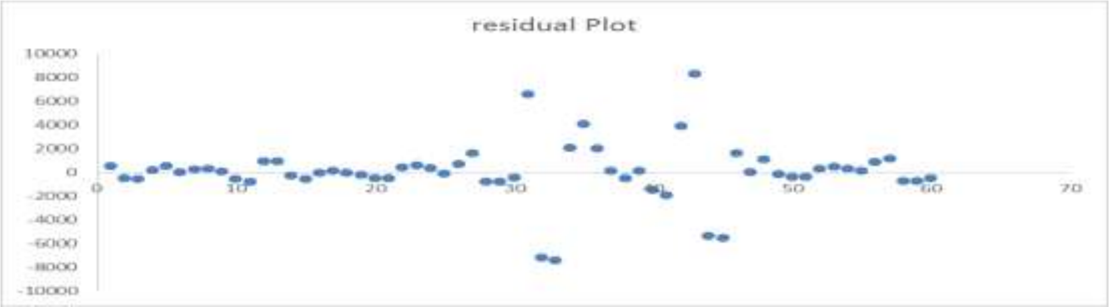
Therefore, the residual from the two-stage nested design are

$$\in_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

where

$\bar{y}_{ij.}$ are the individual batch averages

See Appendix D for the observations, fitted values and residuals for the production data. Figure 4.1 shows the residual plot, using Micro-excel.



Figures 4.1: Residual Plot of the Two-Stage Nested Design

Computing, using some statistical software packages to perform the analysis for a nested design (Minitab 17 and SPSS 21 in Appendix). The numerical results are in agreement with the manual calculations reported or Micro-Excel above. Minitab also reports the expected mean squares and presents the output from the Two-Stage Nested Design ANOVA procedure in Minitab and SPSS.

4.3 Discussion

The appropriate statistics for testing the effects of factors A and B depend on whether A nor B are random or fixed, using the hypothesis stated in Section 3.2 (Hypothesis A), which is summarized in Table 4.3 (Analysis of Variance for the Two-Stage Nested Design). The factor A (Area Planted and Production) and factor B within A (Crop) are fixed.

Two replications are obtained for each region. The goal is to compare the crop production of the different region. The factors are fixed because you are interested in comparing the particular regions. The factors are nested because each crop for each region is unique. The analysis done using GLM, multiple comparisons (to compare the mean response for the different crop) and Two-Stage Nested Design are discussed as follow:-

Minitab 17 and SPSS 21 displays a factor level table, an ANOVA table, multiple comparison confidence intervals for pairwise differences between companies, and the corresponding multiple comparison hypothesis tests. The ANOVA F-tests indicate that there is significant evidence for crop effects (see Table 4.3, 4.5 and Minitab ANOVA results in Appendix B). The calculated p-value is less than critical value 0.05 for crop with respect to area planted and production [since factor A is F_{table} 35.07 with p-values 0.000 and factor A is F_{table} 12.07 with p-values 0.000].

Examine the multiple comparison confidence intervals in Table 4.6 and 4.7 in Appendix C. There are five sets: 1) for the crop 1 mean subtracted from the crop 2, 3, 4 and 5 means; 2) for the crop 2 mean subtracted from the crop 1, 3, 4 and 5 means etc. The first interval, for the crop 1 mean minus the crop 2 mean, do not contain zero is in the confidence interval.

Thus, there is a significant between crop 1 and 2 at 0.05 for differences in means. However, there is evidence that all other pairs of means are different except crop 3, because the confidence intervals for the differences in means do not contain zero. An advantage of confidence intervals is that you can see the magnitude of the differences between the means.
The dummy variable Regression model (GLM) result shows that cassava and yam production are significant at 5 percent, while others are not. In addition, two-Stage Nested Design ANOVA result also confirmed the result of dummy variable Regression model (GLM) in Table 4.7 (Appendix C). Grouping cassava and yam as one subset group while rice, palm oil and beans/cowpea as another subset group.

5. Conclusion

Despite this study is agricultural basis research, the two-Stage Nested Design configuration is valuable in numerous logical and modern examinations. In Nested Design experimental settings, it is not unusual to find that some factors require large experimental units whereas other factors require small experimental units**.** In principle, we must carefully consider how the experiment must be conducted and incorporate all restrictions on randomization into the analysis. The fundamental point of the examination is applying dummy variable regression model and two-stage nested design of agricultural variable to determine significant between or among factors. The univariate General Linear Model (GLM), Multiple Comparisons and Two-Stage Nested Design methods where used to compare the mean response for the different crops with respect to Area Planted and Production.

This research identified significant difference among crop at 5 percent [(cassava and yam production) and (rice, palm oil and beans/cowpea)], using dummy variable Regression model (GLM) and Two-Stage Nested Design.
Then, multiple comparisons of the agricultural variables indicated that cassava and yam production yield’s more in all regions in Nigeria.
We recommend this result serve as a guide to individuals or co-operate body that may wish to involve in area- planted/production of different crops in Nigeria.

Contribution: This research was able to identify significant differences among crop, using dummy variable Regression model (GLM) and Two-Stage Nested Design.

6. References

Armitage, P. (2003). Fisher, Bradford Hill, and randomization. *International Journal of Epidemiology*, *32*(6), 925-928.

Bourdieu, P. (1973). The three forms of theoretical knowledge. *Information (International Social Science Council)*, *12*(1), 53-80.

Crawley, M. J. (2012). *The R book*. John Wiley & Sons.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.

Denzin, N. K. (1973). *The research act: A theoretical introduction to sociological methods*. Transaction publishers.

Denzin, N. K. (1978). The research act: A theoretical orientation to sociological methods. *McGraw-Hill, New York.——2012,‘Triangulation*, *2*, 80-88.

Eisenhart, C. (1947). Inverse sine transformation of proportions. *Selected techniques of statistical analysis for scientific and industrial research and production and management engineering*, 395-416.

Gelman, Andrew (2005). "Analysis of variance? Why it is more important than ever". *The Annals of Statistics* **33**: 1–53.

Gelman, Andrew (2008). "Variance, analysis of". *The new Palgrave dictionary of economics*.

Lehmann, E. L., and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review*, *99*(3), 435-452

Montgomery, D. C. (1999). Experimental design for product and process design and development. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *48*(2), 159-177.

Ott, R. Lyman, and Michael Longnecker. 2001. *An introduction to statistical methods and data analysis*.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 797-810

Sall, F., and Lehman, A. (1996). JMP IN, SAS Institute. *Inc. Ducksberry Press, Belmont, California*.

Sokal, R. R., & Rohlf, F. J. (1981). The principles and practice of statistic in biological research. *WH Creman, San Francisco*, 262-265.

Sokal, R. R., and Rohlf, F. J. 1995. *Biometry: the principles and practice of statistics in biological research*, 68-70.

Sommer, H. G. (2011). More than method?: A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, *5*(2), 150-166.

Tashakkori, A., and Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research*. Sage.

Zar, J. H. (1984). Two-factor analysis of variance. *Biostatistical analysis*, 206-235.

Zar, J. H. (1999). *Biostatistical analysis*. Pearson Education India.

Appendix A

4.3.1 Two Stage Nested Design, using Micro-Excel

Table 4.2: Area Planted (Thousand Hectares) and Production (Thousand Metric Tons) by Crop and Region, 2012-2014

			Area **Planted**					Production			
	Cassava	Rice	Yam	Palm oil	Beans/Cow pea	Cassava	Rice	Yam	Palm oil	Beans/Cow pea	$y_{..k}$
NC	1537.45	1042.11	1803.7	305.19	703.04	18232.05	2216.55	17676.34	232.87	896.66	44645.96
NE	462.62	1144.02	624.74	33.05	1496.77	4426.24	1580.59	3980.3	11.82	1636.89	15397.04
NW	432.94	919.29	308.12	17.38	2411.03	4241.76	2200.6	3827.4	7.23	1897.51	16263.26
SE	1178.2	247.38	844.08	898.24	18.43	13728.63	598.91	10985.47	693.62	3.49	29196.45
SS	1516.38	33.21	1008.52	1098.81	0.79	15729.3	158.04	9378.32	823.85	0.82	29748.04
SW	985.25	1731.75	837.65	882.02	337.292	13672.64	5984.209	10453.38	700.62	235.0032	35819.811
y_{ij}	6112.84	5117.76	5426.81	3234.69	4967.352	70030.62	12738.899	56301.21	2470.01	4670.3732	**171070.561**
$y_{i..}$		24859.4					146211.11				

NC: North Central, NE: North East, NW: North West., SE: South East, SS: South-South, SW: South West

Appendix B

Two Stage Nested Design, using Minitab 17

Using GLM and Multiple Comparisons with Nested Design (see coded procedure in Appendix D)

Minitab 17 software output result

General Linear Model: production versus Area/Planted, crop

```
Factor                Type   Levels  Values
Area/Planted          fixed     2    1, 2
Crop (Area/Planted)   fixed    10    1, 2, 3, 4, 5, 1, 2, 3, 4, 5

Analysis of Variance for production, using Sequential SS for Tests
Source                DF    Seq SS       Adj SS      Seq MS      F      P
Area/Planted          1     245437103    245437103   245437103   35.71  0.000
Crop (Area/Planted)   8     665552814    665552814   83194102    12.10  0.000
Error                 50    343663819    343663819   6873276
Total                 59    1254653736

S = 2621.69   R-Sq = 72.61%   R-Sq(adj) = 67.68%
Tukey 95.0% Simultaneous Confidence Intervals
Response Variable production
All Pairwise Comparisons among Levels of crop(Area/Planted)
Area/Planted = 1
crop = 1  subtracted from:

Area/Planted  crop   Lower      Center      Upper
1             2      -5175      -165.8      4843
1             3      -5123      -114.3      4895
1             4      -5489      -479.7       4529
1             5      -5200      -190.9       4818
2             1      5644       10653.0     15662
2             2      -3905      1104.3      6113
2             3      3356       8364.7      13374
2             4      -5616      -607.1      4402
2             5      -5249      -240.4      4769

Area/Planted  crop    ------+---------+---------+---------+
1              2           (----*----)
1              3           (----*----)
1             4           (----*----)
1             5           (----*----)
2             1                        (----*----)
2             2           (----*----)
2             3                        (----*----)
2             4          (----*----)
2             5           (----*----)
                      ------+---------+---------+---------+
                         -10000        0     10000     20000
Area/Planted = 1
crop = 2  subtracted from:
```

```
Area/Planted  crop  Lower   Center  Upper
1             3     -4958     51.5   5061
1             4     -5323   -313.8   4695
1             5     -5034    -25.1   4984
2             1      5810  10818.8  15828
2             2     -3739   1270.2   6279
2             3      3522   8530.6  13540
2             4     -5450   -441.3   4568
2             5     -5084    -74.6   4934

Area/Planted  crop  ------+---------+---------+---------+
1             3                (----*----)
1             4                (----*----)
1             5                (----*----)
2             1                           (----*----)
2             2                 (----*----)
2             3                         (----*----)
2             4                (----*----)
2             5                (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000
Area/Planted = 1
crop = 3  subtracted from:
Area/Planted  crop  Lower   Center  Upper
1             4     -5374   -365.4   4644
1             5     -5086    -76.6   4932
2             1      5758  10767.3  15776
2             2     -3790   1218.7   6228
2             3      3470   8479.1  13488
2             4     -5502   -492.8   4516
2             5     -5135   -126.1   4883

Area/Planted  crop  ------+---------+---------+---------+
1             4                (----*----)
1             5                (----*----)
2             1                           (----*----)
2             2                 (----*----)
2             3                        (----*----)
2             4               (-----*----)
2             5                (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000

Area/Planted = 1
crop = 4  subtracted from:
Area/Planted  crop  Lower   Center  Upper
1             5     -4720    288.8   5298
2             1      6124  11132.7  16142
2             2     -3425   1584.0   6593
2             3      3835   8844.4  13853
2             4     -5136   -127.4   4882
2             5     -4770    239.3   5248

Area/Planted  crop  ------+---------+---------+---------+
1             5                (----*----)
2             1                           (----*----)
2             2                  (----*----)
2             3                         (----*----)
2             4                (----*----)
2             5                (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000

Area/Planted = 1
crop = 5  subtracted from:
Area/Planted  crop  Lower   Center  Upper
2             1      5835  10843.9  15853
```

```
2             2     -3714   1295.3   6304
2             3      3547   8555.6  13565
2             4     -5425   -416.2   4593
2             5     -5059    -49.5   4960

Area/Planted  crop  ------+---------+---------+---------+
2             1                            (----*----)
2             2                 (----*----)
2             3                         (----*----)
2             4                (----*----)
2             5                (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000

Area/Planted = 2
crop = 1  subtracted from:
Area/Planted  crop   Lower  Center  Upper
2             2     -14558   -9549  -4540
2             3      -7297   -2288   2721
2             4     -16269  -11260  -6251
2             5     -15902  -10893  -5884

Area/Planted  crop  ------+---------+---------+---------+
2             2      (----*----)
2             3              (----*----)
2             4     (----*----)
2             5     (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000

Area/Planted = 2
crop = 2  subtracted from:

Area/Planted  crop  Lower  Center  Upper  ------+---------+---------+---------+
2             3      2251    7260  12269                    (----*----)
2             4     -6720   -1711   3298           (----*----)
2             5     -6354   -1345   3664            (----*----)
                                          ------+---------+---------+---------+
                                           -10000         0     10000     20000

Area/Planted = 2
crop = 3  subtracted from:
Area/Planted  crop   Lower  Center  Upper
2             4     -13981   -8972  -3963
2             5     -13614   -8605  -3596

Area/Planted  crop  ------+---------+---------+---------+
2             4       (----*----)
2             5       (----*----)
                    ------+---------+---------+---------+
                     -10000         0     10000     20000

Area/Planted = 2
crop = 4  subtracted from:
Area/Planted  crop  Lower  Center  Upper  ------+---------+---------+---------+
2             5     -4642   366.7   5376             (----*----)
                                          ------+---------+---------+---------+
                                           -10000         0     10000     20000

Tukey Simultaneous Tests
Response Variable production
All Pairwise Comparisons among Levels of crop(Area/Planted)
Area/Planted = 1
crop = 1  subtracted from:

                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
1             2         -165.8        1514  -0.1096    1.0000
```

```
1             3         -114.3        1514  -0.0755    1.0000
1             4         -479.7        1514  -0.3169    1.0000
1             5         -190.9        1514  -0.1261    1.0000
2             1        10653.0        1514   7.0380    0.0000
2             2         1104.3        1514   0.7296    0.9992
2             3         8364.7        1514   5.5263    0.0001
2             4         -607.1        1514  -0.4011    1.0000
2             5         -240.4        1514  -0.1588    1.0000

Area/Planted = 1
crop = 2  subtracted from:
                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
1             3           51.5        1514   0.0340    1.0000
1             4         -313.8        1514  -0.2073    1.0000
1             5          -25.1        1514  -0.0166    1.0000
2             1        10818.8        1514   7.1476    0.0000
2             2         1270.2        1514   0.8392    0.9975
2             3         8530.6        1514   5.6358    0.0001
2             4         -441.3        1514  -0.2915    1.0000
2             5          -74.6        1514  -0.0493    1.0000

Area/Planted = 1
crop = 3  subtracted from:
                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
1             4         -365.4        1514  -0.2414    1.0000
1             5          -76.6        1514  -0.0506    1.0000
2             1        10767.3        1514   7.1135    0.0000
2             2         1218.7        1514   0.8051    0.9982
2             3         8479.1        1514   5.6018    0.0001
2             4         -492.8        1514  -0.3256    1.0000
2             5         -126.1        1514  -0.0833    1.0000

Area/Planted = 1
crop = 4  subtracted from:
                    Difference       SE of            Adjusted
Area/Planted  crop    of Means  Difference   T-Value   P-Value
1             5          288.8        1514   0.19078    1.0000
2             1        11132.7        1514   7.35491    0.0000
2             2         1584.0        1514   1.04651    0.9876
2             3         8844.4        1514   5.84316    0.0000
2             4         -127.4        1514  -0.08420    1.0000
2             5          239.3        1514   0.15808    1.0000

Area/Planted = 1
crop = 5  subtracted from:
                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
2             1        10843.9        1514   7.1641    0.0000
2             2         1295.3        1514   0.8557    0.9971
2             3         8555.6        1514   5.6524    0.0001
2             4         -416.2        1514  -0.2750    1.0000
2             5          -49.5        1514  -0.0327    1.0000

Area/Planted = 2
crop = 1  subtracted from:
                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
2             2          -9549        1514   -6.308    0.0000
2             3          -2288        1514   -1.512    0.8811
2             4         -11260        1514   -7.439    0.0000
2             5         -10893        1514   -7.197    0.0000

Area/Planted = 2
crop = 2  subtracted from:
                    Difference       SE of           Adjusted
Area/Planted  crop    of Means  Difference  T-Value   P-Value
```

```
2               3          7260         1514    4.797     0.0006
2               4         -1711         1514   -1.131     0.9789
2               5         -1345         1514   -0.888     0.9962

Area/Planted = 2
crop = 3  subtracted from:
                     Difference        SE of             Adjusted
Area/Planted   crop    of Means   Difference   T-Value    P-Value
2               4         -8972         1514   -5.927     0.0000
2               5         -8605         1514   -5.685     0.0001


Area/Planted = 2
crop = 4  subtracted from:
                     Difference        SE of             Adjusted
Area/Planted   crop    of Means   Difference   T-Value    P-Value
2               5         366.7         1514   0.2423      1.000
```

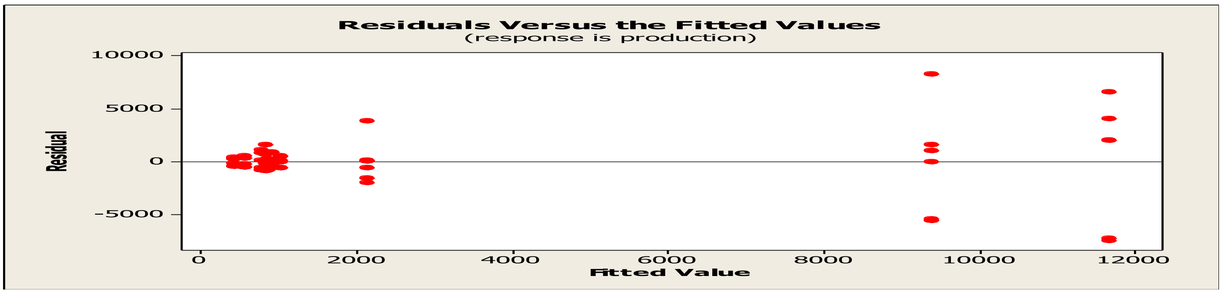


Figure 4.2: Residuals vs Fits for production

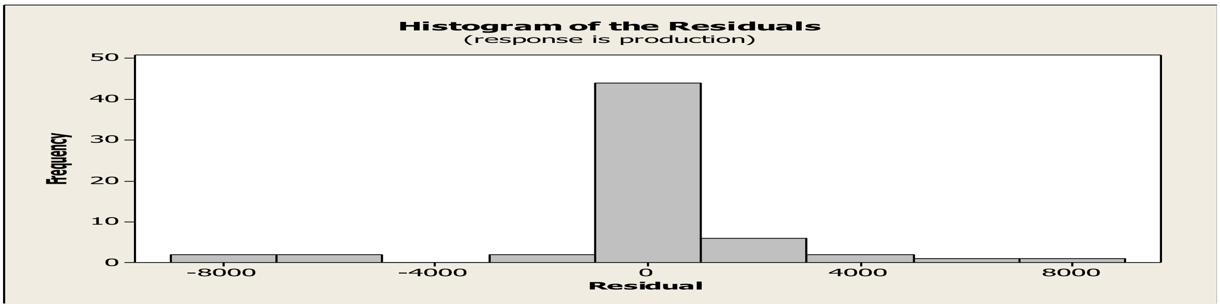


Figure 4.3: Residual Histogram for production

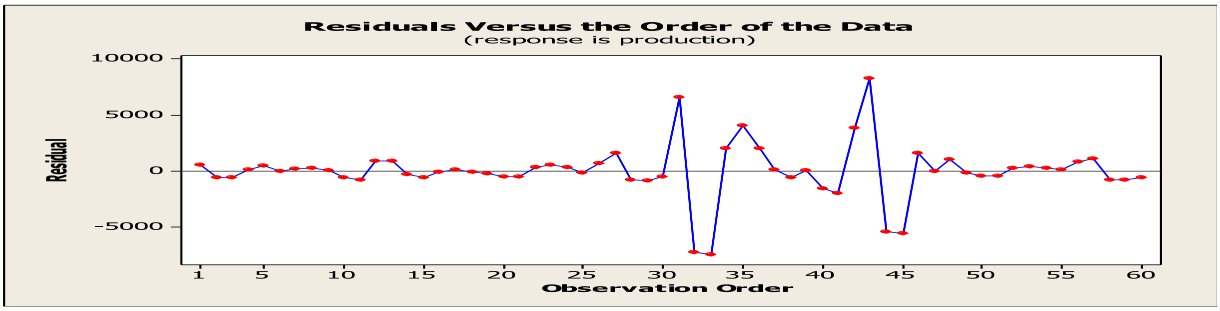


Figure 4.4: Residuals vs Order for production

Appendix C

Two Stage Nested Design, using SPSS 21

Using GLM and Multiple Comparisons with Nested Design (see coded procedure in Appendix D). **SPSS 21 software output result**

Table 4.4: Area Planted and Crop Summarized

Between-Subjects Factors

		N
Area.Planted	1	30
	2	30
Crop	1	12
	2	12
	3	12
	4	12
	5	12

Table 4.5: Two-Stage Nested Design ANOVA using SPSS 21

Tests of Between-Subjects Effects

Dependent Variable:production

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Area.Planted	2.454E8	1	2.454E8	35.709	.000
Crop	3.500E8	8	8.749E7	12.729	.000
Error	3.437E8	50	6873276.380		
Total	1.255E9	59			

a. R Squared = .726 (Adjusted R Squared = .677)

Post Hoc Tests

Table 4.6: Multiple Comparisons of the different crop

Multiple Comparisons

Dependent Variable: Production

						95% Confidence Interval	
	(I) crop	(J) crop	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Tukey HSD	1	2	4857.2337*	1.07030E3	.000	1828.4858	7885.9816
		3	1201.2867	1.07030E3	.794	-1827.4613	4230.0346
		4	5869.8967*	1.07030E3	.000	2841.1487	8898.6446
		5	5542.1446*	1.07030E3	.000	2513.3966	8570.8925
	2	1	-4857.2337*	1.07030E3	.000	-7885.9816	-1828.4858
		3	-3655.9470*	1.07030E3	.011	-6684.6949	-627.1991
		4	1012.6630	1.07030E3	.877	-2016.0849	4041.4109
		5	684.9109	1.07030E3	.968	-2343.8370	3713.6588
	3	1	-1201.2867	1.07030E3	.794	-4230.0346	1827.4613
		2	3655.9470*	1.07030E3	.011	627.1991	6684.6949
		4	4668.6100*	1.07030E3	.001	1639.8621	7697.3579
		5	4340.8579*	1.07030E3	.002	1312.1100	7369.6058
	4	1	-5869.8967*	1.07030E3	.000	-8898.6446	-2841.1487

		2	-1012.6630	1.07030E3	.877	-4041.4109	2016.0849
		3	-4668.6100*	1.07030E3	.001	-7697.3579	-1639.8621
		5	-327.7521	1.07030E3	.998	-3356.5000	2700.9958
	5	1	-5542.1446*	1.07030E3	.000	-8570.8925	-2513.3966
		2	-684.9109	1.07030E3	.968	-3713.6588	2343.8370
		3	-4340.8579*	1.07030E3	.002	-7369.6058	-1312.1100
		4	327.7521	1.07030E3	.998	-2700.9958	3356.5000
LSD	1	2	4857.2337*	1.07030E3	.000	2707.4691	7006.9982
		3	1201.2867	1.07030E3	.267	-948.4779	3351.0512
		4	5869.8967*	1.07030E3	.000	3720.1321	8019.6612
		5	5542.1446*	1.07030E3	.000	3392.3800	7691.9091
	2	1	-4857.2337*	1.07030E3	.000	-7006.9982	-2707.4691
		3	-3655.9470*	1.07030E3	.001	-5805.7116	-1506.1825
		4	1012.6630	1.07030E3	.349	-1137.1016	3162.4275
		5	684.9109	1.07030E3	.525	-1464.8537	2834.6754
	3	1	-1201.2867	1.07030E3	.267	-3351.0512	948.4779
		2	3655.9470*	1.07030E3	.001	1506.1825	5805.7116
		4	4668.6100*	1.07030E3	.000	2518.8455	6818.3745
		5	4340.8579*	1.07030E3	.000	2191.0934	6490.6224
	4	1	-5869.8967*	1.07030E3	.000	-8019.6612	-3720.1321
		2	-1012.6630	1.07030E3	.349	-3162.4275	1137.1016
		3	-4668.6100*	1.07030E3	.000	-6818.3745	-2518.8455
		5	-327.7521	1.07030E3	.761	-2477.5166	1822.0124
	5	1	-5542.1446*	1.07030E3	.000	-7691.9091	-3392.3800
		2	-684.9109	1.07030E3	.525	-2834.6754	1464.8537
		3	-4340.8579*	1.07030E3	.000	-6490.6224	-2191.0934
		4	327.7521	1.07030E3	.761	-1822.0124	2477.5166

Based on observed means.

The error term is Mean Square(Error) = 6873276.380.

*. The mean difference is significant at the 0.05 level.

Table 4.7: Multiple Comparisons of the different crop (Homogeneous Subsets)

Production

	Crop	N	Subset	
			1	2
Tukey HSD[a]	4	12	4.7539E2	
	5	12	8.0314E2	
	2	12	1.4881E3	
	3	12		5.1440E3
	1	12		6.3453E3
	Sig.		.877	.794

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 6873276.380.

a. Uses Harmonic Mean Sample Size = 12.000.

Note: 1= Cassava, 2= Rice, 3=Yam, 4=Palm Oil, 5=Beans/cowpea

Appendix D

Response	constant	CA	RA	YA	PO.A	BC.A	CP	RP	YP	PO.P	BC.P
1537.45	1	1	0	0	0	0	0	0	0	0	0
462.62	1	1	0	0	0	0	0	0	0	0	0
432.94	1	1	0	0	0	0	0	0	0	0	0
1178.2	1	1	0	0	0	0	0	0	0	0	0
1516.38	1	1	0	0	0	0	0	0	0	0	0
985.25	1	1	0	0	0	0	0	0	0	0	0
1042.11	1	0	1	0	0	0	0	0	0	0	0
1144.02	1	0	1	0	0	0	0	0	0	0	0
919.29	1	0	1	0	0	0	0	0	0	0	0
247.38	1	0	1	0	0	0	0	0	0	0	0
33.21	1	0	1	0	0	0	0	0	0	0	0
1731.75	1	0	1	0	0	0	0	0	0	0	0
1803.7	1	0	0	1	0	0	0	0	0	0	0
624.74	1	0	0	1	0	0	0	0	0	0	0
308.12	1	0	0	1	0	0	0	0	0	0	0
844.08	1	0	0	1	0	0	0	0	0	0	0
1008.52	1	0	0	1	0	0	0	0	0	0	0
837.65	1	0	0	1	0	0	0	0	0	0	0
305.19	1	0	0	0	1	0	0	0	0	0	0
33.05	1	0	0	0	1	0	0	0	0	0	0
17.38	1	0	0	0	1	0	0	0	0	0	0
898.24	1	0	0	0	1	0	0	0	0	0	0
1098.81	1	0	0	0	1	0	0	0	0	0	0
882.02	1	0	0	0	1	0	0	0	0	0	0
703.04	1	0	0	0	0	1	0	0	0	0	0
1496.77	1	0	0	0	0	1	0	0	0	0	0
2411.03	1	0	0	0	0	1	0	0	0	0	0
18.43	1	0	0	0	0	1	0	0	0	0	0
0.79	1	0	0	0	0	1	0	0	0	0	0
337.292	1	0	0	0	0	1	0	0	0	0	0
18232.1	1	0	0	0	0	0	1	0	0	0	0
4426.24	1	0	0	0	0	0	1	0	0	0	0
4241.76	1	0	0	0	0	0	1	0	0	0	0
13728.6	1	0	0	0	0	0	1	0	0	0	0
15729.3	1	0	0	0	0	0	1	0	0	0	0
13672.6	1	0	0	0	0	0	1	0	0	0	0
2216.55	1	0	0	0	0	0	0	1	0	0	0
1580.59	1	0	0	0	0	0	0	1	0	0	0
2200.6	1	0	0	0	0	0	0	1	0	0	0
598.91	1	0	0	0	0	0	0	1	0	0	0
158.04	1	0	0	0	0	0	0	1	0	0	0
5984.21	1	0	0	0	0	0	0	1	0	0	0
17676.3	1	0	0	0	0	0	0	0	1	0	0
3980.3	1	0	0	0	0	0	0	0	1	0	0
3827.4	1	0	0	0	0	0	0	0	1	0	0
10985.5	1	0	0	0	0	0	0	0	1	0	0
9378.32	1	0	0	0	0	0	0	0	1	0	0

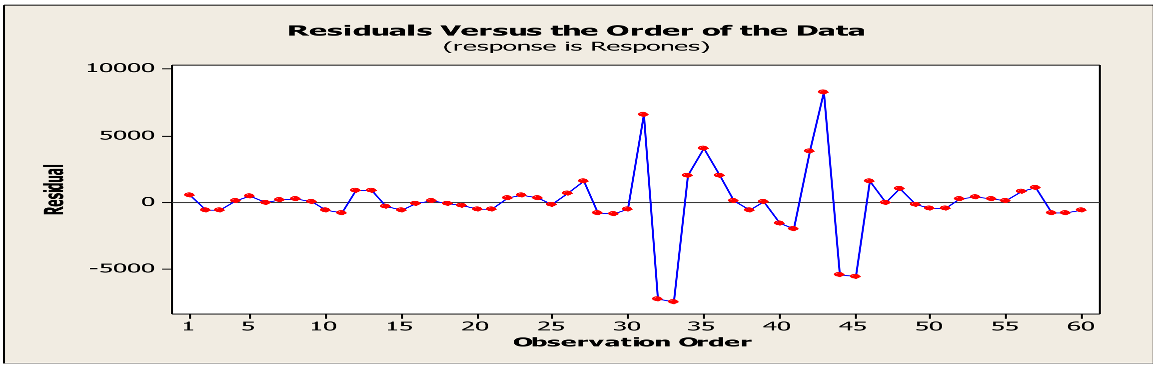
10453.4	1	0	0	0	0	0	0	0	1	0	0
232.87	1	0	0	0	0	0	0	0	0	1	0
11.82	1	0	0	0	0	0	0	0	0	1	0
7.23	1	0	0	0	0	0	0	0	0	1	0
693.62	1	0	0	0	0	0	0	0	0	1	0
823.85	1	0	0	0	0	0	0	0	0	1	0
700.62	1	0	0	0	0	0	0	0	0	1	0
896.66	1	0	0	0	0	0	0	0	0	0	1
1636.89	1	0	0	0	0	0	0	0	0	0	1
1897.51	1	0	0	0	0	0	0	0	0	0	1
3.49	1	0	0	0	0	0	0	0	0	0	1
0.82	1	0	0	0	0	0	0	0	0	0	1
235.003	1	0	0	0	0	0	0	0	0	0	1

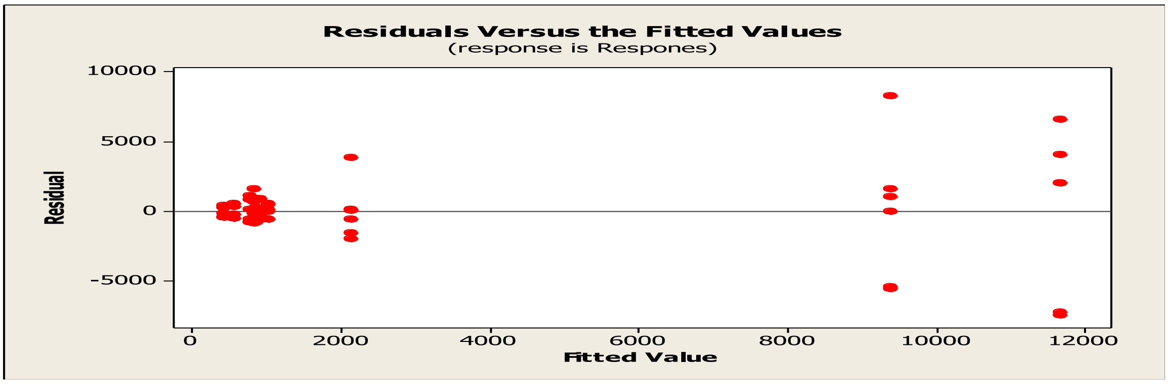
Appendix E

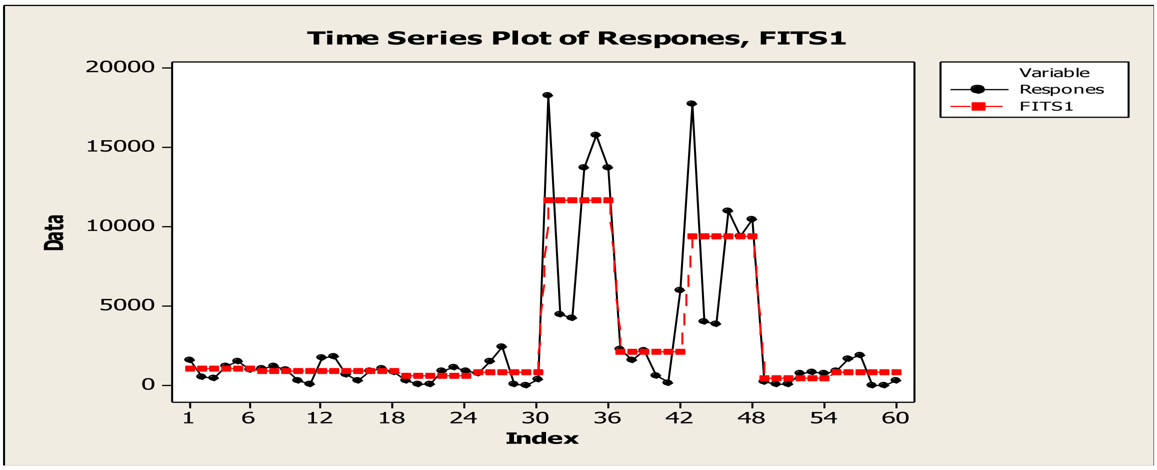
Regression Analysis: Respones versus constant, CA, ...

```
* BC.P is highly correlated with other X variables
* BC.P has been removed from the equation.

The regression equation is
Response = 778 + 240 CA + 75 RA + 126 YA - 239 PO.A + 49 BC.A + 10893 CP
         + 1345 RP + 8605 YP - 367 PO.P

Predictor   Coef  SE Coef      T      P
Constant     778     1070   0.73  0.470
CA           240     1514   0.16  0.874
RA            75     1514   0.05  0.961
YA           126     1514   0.08  0.934
PO.A        -239     1514  -0.16  0.875
BC.A          49     1514   0.03  0.974
CP         10893     1514   7.20  0.000
RP          1345     1514   0.89  0.379
YP          8605     1514   5.69  0.000
PO.P        -367     1514  -0.24  0.810

S = 2621.69   R-Sq = 72.6%   R-Sq(adj) = 67.7%

Analysis of Variance
Source          DF          SS         MS      F      P
Regression       9   910989917  101221102  14.73  0.000
Residual Error  50   343663819    6873276
Total           59  1254653736

Source  DF     Seq SS
CA       1   22383850
RA       1   32723873
YA       1   40341945
PO.A     1   68902454
BC.A     1   81842963
```

CP	1	346602787
RP	1	8837436
YP	1	308951145
PO.P	1	403467







Appendix D

Coded, Fitted values and Residuals

Area Planted	crop	Observed Values (Production) y_{ijk}	Fitted Micro-Excel $\hat{y}_{ijk} = \bar{y}_{ij.}$	FITTED Minitab 17 $\hat{y}_{ijk} = \bar{y}_{ij.}$	Predicted Value SPSS $\hat{y}_{ijk} = \bar{y}_{ij.}$	Residual $\in_{ijk} = y_{ijk} - \bar{y}_{ij.}$
1	1	1537	1018.807	1018.8	1018.807	518.6433
1	1	462.6	1018.807	1018.8	1018.807	-556.187
1	1	432.9	1018.807	1018.8	1018.807	-585.867
1	1	1178	1018.807	1018.8	1018.807	159.3933
1	1	1516	1018.807	1018.8	1018.807	497.5733
1	1	985.3	1018.807	1018.8	1018.807	-33.5567
1	2	1042	852.9595	853	852.9595	189.1505
1	2	1144	852.9595	853	852.9595	291.0605
1	2	919.3	852.9595	853	852.9595	66.33053
1	2	247.4	852.9595	853	852.9595	-605.579
1	2	33.21	852.9595	853	852.9595	-819.749
1	2	1732	852.9595	853	852.9595	878.7873
1	3	1804	904.4683	904.5	904.4683	899.2317
1	3	624.7	904.4683	904.5	904.4683	-279.728
1	3	308.1	904.4683	904.5	904.4683	-596.348
1	3	844.1	904.4683	904.5	904.4683	-60.3883
1	3	1009	904.4683	904.5	904.4683	104.0517
1	3	837.7	904.4683	904.5	904.4683	-66.8183
1	4	305.2	539.115	539.1	539.115	-233.925
1	4	33.05	539.115	539.1	539.115	-506.065
1	4	17.38	539.115	539.1	539.115	-521.735
1	4	898.2	539.115	539.1	539.115	359.125
1	4	1099	539.115	539.1	539.115	559.695
1	4	882	539.115	539.1	539.115	342.905
1	5	703	827.892	827.9	827.892	-124.852

1	5	1497	827.892	827.9	827.892	668.878
1	5	2411	827.892	827.9	827.892	1583.138
1	5	18.43	827.892	827.9	827.892	-809.462
1	5	0.79	827.892	827.9	827.892	-827.102
1	5	337.3	827.892	827.9	827.892	-490.6
2	1	18232	11671.77	11671.8	11671.77	6560.28
2	1	4426	11671.77	11671.8	11671.77	-7245.53
2	1	4242	11671.77	11671.8	11671.77	-7430.01
2	1	13729	11671.77	11671.8	11671.77	2056.86
2	1	15729	11671.77	11671.8	11671.77	4057.53
2	1	13673	11671.77	11671.8	11671.77	2000.87
2	2	2217	2123.15	2123.1	2123.15	93.40017
2	2	1581	2123.15	2123.1	2123.15	-542.56
2	2	2201	2123.15	2123.1	2123.15	77.45017
2	2	598.9	2123.15	2123.1	2123.15	-1524.24
2	2	158	2123.15	2123.1	2123.15	-1965.11
2	2	5984	2123.15	2123.1	2123.15	3861.059
2	3	17676	9383.535	9383.5	9383.535	8292.805
2	3	3980	9383.535	9383.5	9383.535	-5403.24
2	3	3827	9383.535	9383.5	9383.535	-5556.14
2	3	10985	9383.535	9383.5	9383.535	1601.935
2	3	9378	9383.535	9383.5	9383.535	-5.215
2	3	10453	9383.535	9383.5	9383.535	1069.845
2	4	232.9	411.6683	411.7	411.6683	-178.798
2	4	11.82	411.6683	411.7	411.6683	-399.848
2	4	7.23	411.6683	411.7	411.6683	-404.438
2	4	693.6	411.6683	411.7	411.6683	281.9517
2	4	823.9	411.6683	411.7	411.6683	412.1817
2	4	700.6	411.6683	411.7	411.6683	288.9517

2	5	896.7	778.3955	778.4	778.3955	118.2645
2	5	1637	778.3955	778.4	778.3955	858.4945
2	5	1898	778.3955	778.4	778.3955	1119.114
2	5	3.49	778.3955	778.4	778.3955	-774.906
2	5	0.82	778.3955	778.4	778.3955	-777.576
2	5	235	778.3955	778.4	778.3955	-543.392